

COLLECTED
PAPERS

GARETH EVANS

CLARENDON PRESS · OXFORD

1985

d20/27+
91/46109/85

ik
Oxford University Press, Walton Street, Oxford OX2 6DP
Oxford New York Toronto
Delhi Bombay Calcutta Madras Karachi
Kuala Lumpur Singapore Hong Kong Tokyo
Nairobi Dar es Salaam Cape Town
Melbourne Auckland
and associated companies in
Beirut Berlin Ibadan Nicosia

Oxford is a trade mark of Oxford University Press

Published in the United States
by Oxford University Press, New York

© Antonia Phillips 1985

All rights reserved. No part of this publication may be reproduced,
stored in a retrieval system, or transmitted, in any form or by any means,
electronic, mechanical, photocopying, recording, or otherwise, without
the prior permission of Oxford University Press

This book is sold subject to the condition that it shall not, by way
of trade or otherwise, be lent, re-sold, hired out or otherwise circulated
without the publisher's prior consent in any form of binding or cover
other than that in which it is published and without a similar condition
including this condition being imposed on the subsequent purchaser

British Library Cataloguing in Publication Data

Evans, Gareth, 1946-
Collected papers.

1. Analysis (Philosophy)

I. Title

149'.94 B808.5

ISBN 0-19-824737-0

phc
370

485a



Set by Joshua Associates Limited, Oxford
Printed in Great Britain
at the University Press, Oxford
by David Stanford
Printer to the University

Foreword

Of the thirteen papers by Gareth Evans included here all but two have been published before, either in journals or in collections; several of them appeared posthumously, although submitted for publication by the author. The most natural arrangement for these papers seemed to be a chronological one, not least because the suggestion that each paper equally reflects Evans's latest views is avoided. It is probable that he would have made some revisions, had he been preparing the collection himself. None but the most minor editorial alterations has been made to the published texts; of the two previously unpublished papers, only one, 'Does Tense Logic Rest Upon a Mistake?', was submitted for publication by Evans. The other, 'Molyneux's Question', is a first draft, and would not have been published by Evans as it stands. Even so, I have left it as he left it, merely correcting misquotations and the occasional infelicity of prose. Two appendices have been added. The first reprints a brief reply to a paper by Jerry Fodor, while the other provides an afterthought to 'Semantic Structure and Logical Form' communicated in a letter to David Wiggins. Had it been possible to sift through letters and notebooks in a systematic way, doubtless other such afterthoughts would have come to light.

A number of people have generously offered advice and assistance in preparing this collection, for which I am grateful: John McDowell, Colin McGinn, Galen Strawson, David Wiggins and Bruce Vermazen. My thanks are also due to the staff of Oxford University Press, who have been both helpful and patient.

ANTONIA PHILLIPS

Contents

1	The Causal Theory of Names	1
2	Identity and Predication	25
3	Semantic Structure and Logical Form	49
4	Pronouns, Quantifiers, and Relative Clauses (I)	76
5	Pronouns, Quantifiers, and Relative Clauses (II)	153
6	Can There Be Vague Objects?	176
7	Reference and Contingency	178
8	Pronouns	214
9	Things Without the Mind—A Commentary upon Chapter Two of Strawson's <i>Individuals</i>	249
10	Understanding Demonstratives	291
11	Semantic Theory and Tacit Knowledge	322
12	Does Tense Logic Rest upon a Mistake?	343
13	Molyneux's Question	364
	Appendix I: Commentary on Jerry A. Fodor's 'Methodological Solipsism'	400
	Appendix II: A Further Reflection on Semantic Structure and Logical Form	405
	Index	409

The Causal Theory of Names

1. In a paper which provides the starting-point of this enquiry Saul Kripke opposes what he calls the Description Theory of Names and makes a counter-proposal of what I shall call the Causal Theory.¹ To be clear about what is at stake and what should be the outcome in the debate he initiated seems to me important for our understanding of talk and thought about the world in general as well as for our understanding of the functioning of proper names. I am anxious therefore that we identify the profound bases and likely generalizations of the opposing positions and do not content ourselves with counter-examples.

I should say that Kripke deliberately held back from presenting his ideas as a theory. I shall have to tighten them up, and I may suggest perhaps unintended directions of generalization; therefore his paper should be checked before the Causal Theory I consider is attributed to him.

There are two related but distinguishable questions concerning proper names. The first is about what the name denotes upon a particular occasion of its use when this is understood as being partly determinative of what the speaker strictly and literally said. I shall use the faintly barbarous coinage: *what the speaker denotes* (upon an occasion) for this notion. The second is about *what the name denotes*; we want to know what conditions have to be satisfied by an expression and an item for the first to be the, or a, name of the second. There is an entirely parallel pair of questions concerning general terms. In both cases it is ambiguity

From *ASSV* 47 (1973), pp. 187-208. Reprinted by permission of the Aristotelian Society.

¹ S. A. Kripke, 'Naming and Necessity', in D. Davidson and G. Harman (eds). *Semantics of Natural Language* (Dordrecht: Reidel, 1972), pp. 253-355 (+ Appendix).

which prevents an easy answer of the first in terms of the second; to denote x it is not sufficient merely to utter something which is x 's name.

Consequently there are two Description Theories, not distinguished by Kripke.² The Description Theory of speaker's denotation holds that a name 'NN' denotes x upon a particular occasion of its use by a speaker S just in case x is uniquely that which satisfies all or most of the descriptions ϕ such that S would assent to 'NN is ϕ ' (or 'That NN is ϕ '). Crudely: the cluster of information S has associated with the name determines its denotation upon a particular occasion by *fit*. If the speaker has no individuating information he will denote nothing.

The Description Theory of what a name denotes holds that, associated with each name as used by a group of speakers who believe and intend that they are using the name with the same denotation, is a description or set of descriptions cullable from their beliefs which an item has to satisfy to be the bearer of the name. This description is used to explain the role of the name in existential, identity, and opaque contexts. The theory is by no means committed to the thesis that every user of the name must be in possession of the description; just as Kripke is not committed to holding that every user of the expression 'one metre' knows about the metre rod in Paris by saying that its reference is fixed by the description 'Length of stick S in Paris'. Indeed if the description is arrived at in the manner of Strawson³—averaging out the different beliefs of different speakers—it is most unlikely that the description will figure in every user's name-associated cluster.

The direct attack in Kripke's paper passes this latter theory by; most conspicuously the charge that the Description Theory ignores the social character of naming. I shall not discuss it explicitly either, though it will surface from time to time and the extent to which it is right should be clear by the end of the paper.

² This can be seen in the way the list of theses defining the Description Theory alternate between those mentioning a speaker and those that don't, culminating in the uneasy idea of an idiolect of one. The Description Theorists of course do not themselves distinguish them clearly either, and many espouse both.

³ P. F. Strawson, *Individuals* (London: Methuen, 1959), p. 191.

Kripke's direct attacks are unquestionably against the first Description Theory. He argues:

(a) An ordinary man in the street can denote the physicist Feynman by using the name 'Feynman' and say something true or false of him even though there is no description uniquely true of the physicist which he can fashion. (The conditions aren't necessary.)

(b) A person who associated with the name 'Gödel' merely the description 'prover of the incompleteness of Arithmetic' would none the less be denoting Gödel and saying something false of him in uttering 'Gödel proved the incompleteness of Arithmetic' even if an unknown Viennese by the name of Schmidt had in fact constructed the proof which Gödel had subsequently broadcast as his own. (If it is agreed that the speaker does not denote Schmidt the conditions aren't sufficient; if it is also agreed that he denotes Gödel, again they are not necessary.)

The strong thesis (that the Description Theorist's conditions are sufficient) is outrageous. What the speaker denotes in the sense we are concerned with is connected with saying in that strict sense which logicians so rightly prize, and the theory's deliverance of strict truth conditions are quite unacceptable. They would have the consequence, for example, that if I was previously innocent of knowledge or belief regarding Mr Y , and X is wrongly introduced to me as Mr Y , then I must speak the truth in uttering 'Mr Y is here' since X satisfies the overwhelming majority of descriptions I would associate with the name and X is there. I have grave doubts as to whether anyone has ever seriously held this thesis.

It is the weaker thesis—that some descriptive identification is necessary for a speaker to denote something—that it is important to understand. Strictly, Kripke's examples do not show it to be false since he nowhere provides a convincing reason for not taking into account speakers' possession of descriptions like 'man bearing such-and-such a name'; but I too think it is false. It can be seen as the fusion of two thoughts. First: that in order to be saying something by uttering an expression one must utter the sentence with

certain intentions; this is felt to require, in the case of sentences containing names, that one be aiming at something with one's use of the name. Secondly—and this is where the underpinning from a certain Philosophy of Mind becomes apparent—to have an intention or belief concerning some item (which one is not in a position to demonstratively identify) one must be in possession of a description uniquely true of it. Both strands deserve at least momentary scrutiny.

We are prone to pass too quickly from the observation that neither parrots nor the wind *say* things to the conclusion that to say that p requires that one must intend to say that p and therefore, so to speak, be able to identify p independently of one's sentence. But the most we are entitled to conclude is that to say something one must intend to say something by uttering one's sentence (one normally will intend to say what it says). The application of the stricter requirement would lead us to relegate too much of our discourse to the status of mere mouthing. We constantly use general terms of whose satisfaction conditions we have but the dimmest idea. 'Microbiologist', 'chlorine' (the stuff in swimming-pools), 'nicotine' (the stuff in cigarettes); these, and countless other words, we cannot define nor offer remarks which would distinguish their meaning from that of closely related words. It is wrong to say that we say nothing by uttering sentences containing these expressions, even if we recoil from the strong thesis, from saying that what we do say is determined by those hazy ideas and half-identifications we would offer if pressed.

The Philosophy of Mind is curiously popular but rarely made perfectly explicit.⁴ It is held by anyone who holds that S believes that a is F if and only if

$$\exists \phi [S \text{ believes } \exists x(\phi x \ \& \ (\forall y)(\phi y \rightarrow x = y) \ \& \ Fx) \ \& \ \phi a \ \& \ (\forall y)(\phi y \rightarrow y = a)]$$

⁴ See, e.g., J. R. Searle, *Speech Acts* (Cambridge: Cambridge University Press, 1969), p. 87; E. Gellner, 'Ethics and Logic', *Proceedings of the Aristotelian Society* 55 (1954-5), pp. 157-78; B. Russell, *Problems of Philosophy* (Oxford: Oxford University Press, paperback 1976), p. 29. E. Sosa criticizes it in 'Quantifiers Belief and Sellars', in J. W. Davis, D. J. Hockney, and W. K. Wilson (eds), *Philosophical Logic* (Dordrecht: Reidel, 1969), p. 69.

Obvious alterations would accommodate the other psychological attitudes. The range of the property quantifier must be restricted to exclude such properties as 'being identical with a ', otherwise the criterion is trivial.⁵ The situation in which a thinking, planning or wanting human has some item which is the object of his thought, plan or desire is represented as a species of essentially the same situation as that which holds when there is no object and the thought, plan or desire is, as we might say, purely general. There are thoughts, such as the thought that there are eleven-fingered men, for whose expression general terms of the language suffice. The idea is that when the psychological state involves an object, a general term believed to be uniquely instantiated and in fact uniquely instantiated by the item which is the object of the state will figure in its specification. This idea may be coupled with a concession that there are certain privileged objects to which one may be more directly related; indeed such a concession appears to be needed if the theory is to be able to allow what appears an evident possibility: object-directed thoughts in a perfectly symmetrical or cyclical universe.

This idea about the nature of object-directed psychological attitudes obviously owes much to the feeling that there must be something we can say about what is believed or wanted even when there is no appropriate object actually to be found in the world. But it can also be seen as deriving support from a Principle of Charity: so attribute objects to beliefs that true belief is maximized. (I do not think this is an acceptable principle; the acceptable principle enjoins minimizing the attribution of *inexplicable* error and therefore cannot be operated without a theory of the causation of belief for the creatures under investigation.)

We cannot deal comprehensively with this Philosophy of Mind here. My objections to it are essentially those of Wittgenstein. For an item to be the object of some psychological attitude of yours may be simply for you to be placed in a context which relates you to that thing. What makes it one rather than the other of a pair of identical twins that you are in love with? Certainly not some specification blueprinted

⁵ I owe this observation to G. Harman.

in your mind; it may be no more than this: it was one of them and not the other that you have met. The theorist may gesture to the description 'the one I have met' but can give no explanation for the impossibility of its being outweighed by other descriptions which may have been acquired as a result of error and which may in fact happen to fit the other, unmet, twin. If God had looked into your mind, he would not have seen there with whom you were in love, and of whom you were thinking.

With that I propose to begin considering the Causal Theory.

2. The Causal Theory as stated by Kripke goes something like this. A speaker, using a name 'NN' on a particular occasion will denote some item x if there is a causal chain of *reference-preserving links* leading back from his use on that occasion ultimately to the item x itself being involved in a name-acquiring transaction such as an explicit dubbing or the more gradual process whereby nicknames stick. I mention the notion of a reference-preserving link to incorporate a condition that Kripke lays down; a speaker S 's transmission of a name 'NN' to a speaker S' constitutes a reference-preserving link only if S intends to be using the name with the same denotation as he from whom he in his turn learned the name.

Let us begin by considering the theory in answer to our question about speaker's denotation (i.e., at the level of the individual speaker). In particular, let us consider the thesis that it is *sufficient* for someone to denote x on a particular occasion with the name that this use of the name on that occasion be a causal consequence of his exposure to other speakers using the expression to denote x .

An example which might favourably dispose one towards the theory is this. A group of people are having a conversation in a pub, about a certain Louis of whom S has never heard before. S becomes interested and asks: 'What did Louis do then?' There seems to be no question but that S denotes a particular man and asks about him. Or on some subsequent occasion S may use the name to offer some new thought to one of the participants: 'Louis was quite right to do that.' Again he clearly denotes whoever was the subject of conversation in the pub. This is difficult to reconcile with

the Description Theory since the scraps of information which he picked up during the conversation might involve some distortion and fit someone else much better. Of course he has the description 'the man they were talking about' but the theory has no explanation for the impossibility of its being outweighed.

The Causal Theory can secure the right answer in such a case but I think deeper reflection will reveal that it too involves a refusal to recognize the insight about contextual determination I mentioned earlier. For the theory has the following consequence: that at any future time, no matter how remote or forgotten the conversation, no matter how alien the subject-matter and confused the speaker, S will denote one particular Frenchman—perhaps Louis XIII—so long as there is a causal connection between his use at that time and the long-distant conversation.

It is important in testing your intuitions against the theory that you imagine the predicate changed—so that he says something like 'Louis was a basketball player' which was not heard in the conversation and which arises as a result of some confusion. This is to prevent the operation of what I call the 'mouthpiece syndrome' by which we attach sense and reference to a man's remarks only because we hear someone else speaking through him; as we might with a messenger, carrying a message about matters of which he was entirely ignorant.

Now there is no knock-down argument to show this consequence unacceptable; with pliant enough intuitions you can swallow anything in philosophy. But notice how little *point* there is in saying that he denotes one French king rather than any other, or any other person named by the name. There is now nothing that the speaker is prepared to say or do which relates him differentially to that one king. This is why it is so outrageous to say that he believes that Louis XIII is a basketball player. The notion of saying has simply been severed from all the connections that made it of interest. Certainly we did not think we were letting ourselves in for this when we took the point about the conversation in the pub. What has gone wrong?⁶

⁶ Kripke expresses doubts about the sufficiency of the conditions for this sort of reason, see op. cit., p. 303.

The Causal Theory again ignores the importance of surrounding context, and regards the capacity to denote something as a magic trick which has somehow been passed on, and once passed on cannot be lost. We should rather say: in virtue of the context in which the man found himself the man's dispositions were bent towards one particular man—Louis XIII—whose states and doings alone he would count as serving to verify remarks made in that context using the name. And of course that context can persist, for the conversation can itself be adverted to subsequently. But it can also disappear so that the speaker is simply not sensitive to the outcome of any investigations regarding the truth of what he is said to have said. And at this point saying becomes detached, and uninteresting.

(It is worth observing how ambivalent Kripke is on the relation between denoting and believing; when the connection favours him he uses it; we are reminded for example that the ordinary man has a false belief about Gödel and not a true belief about Schmidt. But it is obvious that the results of the 'who are they believing about?' criterion are bound to come dramatically apart from the results of the 'who is the original bearer of the name?' criterion, if for no other reason than that the former must be constructed to give results in cases where there is no name and where the latter cannot apply. When this happens we are sternly reminded that 'X refers' and 'X says' are being used in *technical* senses.⁷ But there are limits. One could regard the aim of this paper to restore the connection which must exist between strict truth conditions and the beliefs and interests of the users of the sentences if the technical notion of strict truth conditions is to be of interest to us.)

Reflection upon the conversation in the pub appeared to provide one reason for being favourably disposed towards the Causal Theory. There is another connected reason we ought to examine briefly. It might appear that the Causal Theory provides the basis for a general non-intentional answer to the Problem of Ambiguity. The problem is clear enough: What conditions have to be satisfied for a speaker to have said that *p* when he utters a sentence which may

⁷ *Ibid.*, p. 348 fn.

appropriately be used to say that *q* and that *r* and that *s* in addition? Two obvious alternative answers are:

- (a) the extent to which it is reasonable for his audience to conclude that he was saying that *p*
- (b) his intending to say that *p*

and neither is without its difficulties. We can therefore imagine someone is hoping for a natural extension of the Causal Theory to general terms which would enable him to explain for example how a child who did not have determinative intentions because of the technical nature of the subject-matter may still say something determinate using a sentence which is in fact ambiguous.

I touch upon this to ensure that we are keeping the range of relevant considerations to be brought to bear upon the debate as wide as it must be. But I think little general advantage can accrue to the Causal Theory from thus broadening the considerations. The reason is that it simply fails to have the generality of the other two theories; it has no obvious application, for example, to syntactic ambiguity or to ambiguity produced by attempts to refer with non-unique descriptions, or pronouns. It seems inconceivable that the general theory of disambiguation required for such cases would be inadequate to deal with the phenomenon of shared names and would require *ad hoc* supplementation from the Causal Theory.

I want to stress how, precisely because the Causal Theory ignores the way context can be determinative of what gets *said*, it has quite unacceptable consequences. Suppose, for example, on a TV quiz programme I am asked to name a capital city and I say 'Kingston is the capital of Jamaica'; I should want to say that I had said something strictly and literally true even though it turns out that the man from whom I had picked up this scrap of information was actually referring to Kingston upon Thames and making a racist observation.

It may begin to appear that what gets said is going to be determined by what name is used, what items bear the name, and general principles of contextual disambiguation. The

causal origin of the speaker's familiarity with the name, save in certain specialized 'mouthpiece cases', does not seem to have a critical role to play.

This impression may be strengthened by the observation that a causal connection between my use of the name and use by others (whether or not leading back ultimately to the item itself) is simply not necessary for me to use the name to say something. Amongst the Wagera Indians, for example, 'newly born children receive the names of deceased members of their family according to strict rules . . . the first born takes on the name of the paternal grandfather, the second that of the father's eldest brother, the third that of the maternal grandfather'.⁸ In these and other situations (names for streets in US cities etc.) a knowledgeable speaker may excogitate a name and use it to denote some item which bears it without any causal connection whatever with the use by others of that name.

These points might be conceded by Kripke while maintaining the general position that the denotation of a name in a community is still to be found by tracing a causal chain of reference preserving links back to some item. It is to this theory that I now turn.

3. Suppose a parallel theory were offered to explain the sense of general terms (not just terms for natural kinds). One would reply as follows:

'There aren't two fundamentally different mechanisms involved in a word's having a meaning: one bringing it about that a word acquires a meaning, and the other—a causal mechanism—which operates to ensure that its meaning is preserved. The former processes are operative all the time; whatever explains how a word gets its meaning also explains how it preserves it, if preserved it is. Indeed such a theory could not account for the phenomenon of a word's changing its meaning. It is perfectly possible for this to happen without anyone's intending to initiate a new practice with the word; the causal chain would then lead back too far.'

Change of meaning would be decisive against such a theory

⁸ E. Delhaise, 'Les Wagera', *Monogr. Ethnogr.* (1909).

of the meaning of general terms. Change of denotation is similarly decisive against the Causal Theory of Names. Not only are changes of denotation imaginable, but it appears that they actually occur. We learn from Isaac Taylor's *Names and their History* (1898):

In the case of 'Madagascar' a hearsay report of Malay or Arab sailors misunderstood by Marco Polo . . . has had the effect of transferring a corrupt form of the name of a portion of the African mainland to the great African Island.

A simple imaginary case would be this: 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.

It is clear that the Causal Theory unamended is not adequate. It looks as though, once again, the intentions of the speakers to use the name to refer to something must be allowed to count in determination of what it denotes.

But it is not enough to say that and leave matters there. We must at least sketch a theory which will enable 'Madagascar' to be the name of the island yet which will not have the consequence that 'Gödel' would become a name of Schmidt in the situation envisaged by Kripke, nor 'Goliath' a name of the Philistine killed by David. (Biblical scholars now suggest that David did not kill Goliath, and that the attribution of the slaying to Elhannan the Bethlehemite in 2 Sam. 21: 19 is correct. David is thought to have killed a Philistine but not Goliath.)⁹ For although this has never been explicitly argued I would agree that even if the 'information' connected with the name in possession of an entire community was merely that 'Goliath was the Philistine David slew' this would still not mean that 'Goliath' referred in that community to that man, and therefore that the sentence expressed a truth. And if we simultaneously thought that the name *would* denote the Philistine slain by Elhannan then both the necessity and sufficiency of the conditions suggested by the Description

⁹ H. W. Robinson, *The History of Israel* (London: Duckworth, 1941). p. 187.

Theory of the denotation of a name are rejected. This is the case Kripke should have argued but didn't.

4. Before going on to sketch such a theory in the second part of this paper let me survey the position arrived at and use it to make a summary statement of the position I wish to adopt.

We can see the undifferentiated Description Theory as the expression of two thoughts.

(a) The denotation of a name is determined by what speakers intend to refer to by using the name.

(b) The object a speaker intends to refer to by his use of a name is that which satisfies or fits the majority of descriptions which make up the cluster of information which the speaker has associated with the name.

We have seen great difficulties with (a) when this is interpreted as a thesis at the micro-level. But consideration of the phenomenon of a name's getting a denotation, or changing it, suggests that there being a community of speakers using the name with such-and-such as the intended referent is likely to be a crucial constituent in these processes. With names as with other expressions in the language, what they signify depends upon what we use them to signify; a truth whose recognition is compatible with denying the collapse of saying into meaning at the level of the individual speaker.

It is in (b) that the real weakness lies: the bad old Philosophy of Mind which we momentarily uncovered. Not so much in the idea that the intended referent is determined in a more or less complicated way by the associated information, but the specific form the determination was supposed to take: *fit*. There is something absurd in supposing that the intended referent of some perfectly ordinary use of a name by a speaker could be some item utterly isolated (causally) from the user's community and culture simply in virtue of the fact that it fits better than anything else the cluster of descriptions he associates with the name. I would agree with Kripke in thinking that the absurdity resides in the absence

of the causal relation between the item concerned and the speaker. But it seems to me that he has mislocated the causal relation; the important causal relation lies between that item's states and doings and the speaker's body of information—not between the item's being dubbed with a name and the speaker's contemporary use of it.

Philosophers have come increasingly to realize that major concepts in epistemology and the philosophy of mind have causality embedded within them. Seeing and knowing are both good examples.

The absurdity in supposing that the denotation of our contemporary use of the name 'Aristotle' could be some unknown (n.b.) item whose doings are causally isolated from our body of information is strictly parallel to the absurdity in supposing that one might be seeing something one has no causal contact with solely upon the ground that there is a splendid match between object and visual impression.

There probably is some *degree of fit* requirement in the case of seeing which means that after some amount of distortion or fancy we can no longer maintain that the causally operative item was still being seen. And I think it is likely that there is a parallel requirement for referring. We learn, for example, from E. K. Chambers's *Arthur of Britain* that Arthur had a son Anir 'whom legend has perhaps confused with his burial place'. If Kripke's notion of reference fixing is such that those who said Anir was a burial place of Arthur might be denoting a person it seems that it has little to commend it, and is certainly not justified by the criticism he makes against the Description Theory. But the existence or nature of this 'degree of fit' requirement will not be something I shall be concerned with here.

We must allow, then, that the denotation of a name in the community will depend in a complicated way upon what those who use the term intend to refer to, but we will so understand 'intended referent' that typically a *necessary* (but not sufficient) condition for *x*'s being the intended referent of *S*'s use of a name is that *x* should be the source of causal origin of the body of information that *S* has associated with the name.

II

5. The aim I have set myself, then, is modest; it is not to present a complete theory of the denotation of names. Without presenting a general theory to solve the problem of ambiguity I cannot present a theory of speaker's denotation, although I will make remarks which prejudice that issue. I propose merely to sketch an account of what makes an expression into a name for something that will allow names to change their denotations.

The enterprise is more modest yet for I propose to help myself to an undefined notion of speaker's reference by borrowing from the theory of communication. But a word of explanation.

A speaker may have succeeded in *getting it across* or in *communicating* that p even though he uses a sentence which may not appropriately be used to say that p . Presumably this success consists in his audience's having formed a belief about him. This need not be the belief that the speaker intended to say in the strict sense that p , since the speaker may succeed in getting something across despite using a sentence which he is known to know cannot appropriately be used to say that p . The speaker will have referred to a , in the sense that I am helping myself to, only if he has succeeded in getting it across that Fa (for some substitution F). Further stringent conditions are required. Clearly this notion is quite different from the notion of denotation which I have been using, tied as denotation is to saying in the strict sense. One may refer to x by using a description that x does not satisfy; one may not thus denote x .

Now a speaker may know or believe that there is such-and-such an item in the world and intend to refer to it. And this is where the suggestion made earlier must be brought to bear, for *that* item is not (in general) the satisfier of the body of information the possession by the speaker of which makes it true that he knows of the existence of the item; it is rather that item which is causally responsible for the speaker's possession of that body of information, or dominantly responsible if there is more than one. (The point is of course not specific to this intention, or to intention as

opposed to other psychological attitudes.) Let us then, very briefly, explore these two ideas: source and dominance.

Usually our knowledge or belief about particular items is derived from information-gathering transactions, involving a causal interaction with some item or other, conducted ourselves or is derived, maybe through a long chain, from the transactions of others. Perception of the item is the main but by no means the only way an item can impress itself on us; for example, a man can be the source of things we discover by rifling through his suitcase or by reading his works.

A causal relation is of course not sufficient; but we may borrow from the theory of knowledge and say something like this. X is the source of the belief S expresses by uttering ' Fa ' if there was an episode which caused S 's belief in which X and S were causally related in a type of situation apt for producing knowledge that something F -s ($\exists x(Fx)$)—a type of situation in which the belief that something F -s would be caused by something's F -ing. That it is a way of producing knowledge does not mean that it cannot go wrong; that is why X , by smoking French cigarettes, can be the source of the belief S expresses by ' a smokes Greek cigarettes'.

Of course some of our information about the world is not so based; we may deduce that there is a tallest man in the world and deduce that he is over 6 feet tall. No man is the source of this information; a name introduced in relation to it might function very much as the unamended Description Theory suggested.

Legend and fancy can create new characters, or add bodies of sourceless material to other dossiers; restrictions on the causal relation would prevent the inventors of the legends turning out to be the sources of the beliefs their legends gave rise to. Someone other than the ϕ can be the source of the belief S expresses by ' a is the ϕ '; Kripke's Gödel, by claiming the proof, was the source of the belief people manifested by saying 'Gödel proved the incompleteness of Arithmetic', not Schmidt.

Misidentification can bring it about that the item which is the source of the information is different from the item about which the information is believed. I may form the belief about the wife of some colleague that she has nice legs

upon the basis of seeing someone else—but the girl I saw is the source.

Consequently a cluster or dossier of information can be dominantly of¹⁰ an item though it contains elements whose source is different. And we surely want to allow that persistent misidentification can bring it about that a cluster is dominantly of some item other than that it was dominantly of originally.

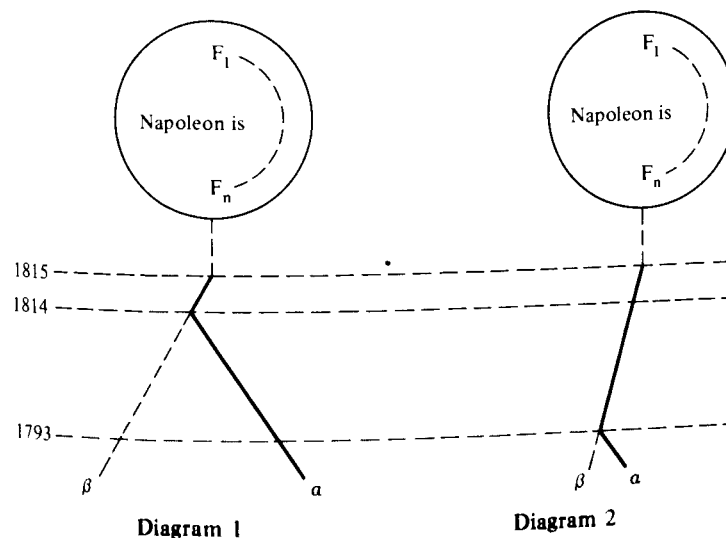
Suppose I get to know a man slightly. Suppose then a suitably primed identical twin takes over his position, and I get to know him fairly well, not noticing the switch. Immediately after the switch my dossier will still be dominantly of the original man, and I falsely believe, as I would acknowledge if it was pointed out, that *he* is in the room. Then I would pass through a period in which neither was dominant; I had not misidentified one as the other, an asymmetrical relation, but rather confused them. Finally the twin could take over the dominant position; I would not have false beliefs about who is in the room, but false beliefs about, for example, when I first met the man in the room. These differences seem to reside entirely in the differences in the believer's reactions to the various discoveries, and dominance is meant to capture those differences.

Dominance is not simply a function of *amount* of information (if that is even intelligible). In the case of persons, for example, each man's life presents a skeleton and the dominant source may be the man who contributed to covering most of it rather than the man who contributed most of the covering. Detail in a particular area can be outweighed by spread. Also the believer's reasons for being interested in the item at all will weigh.

Consider another example. If it turns out that an impersonator had taken over Napoleon's role from 1814 onwards (post-Elba), the cluster of the typical historian would still be dominantly of the man responsible for the earlier exploits

¹⁰ The term is D. Kaplan's, see 'Quantifying In', in D. Davidson and J. Hintikka (eds), *Words and Objections* (Dordrecht: Reidel, 1969): I think there are clear similarities between my notion of a dominant source and notions he is there sketching. However I want nothing to do with vividness. I borrow the term 'dossier' from H. P. Grice's paper 'Vacuous Names' in the same volume.

(α in Diagram 1) and we would say that they had false beliefs about who fought at Waterloo. If however the switch had occurred much earlier, it being an unknown Army Officer being impersonated, then their information would be dominantly of the later man (β in Diagram 2). They did not have false beliefs about who was the general at Waterloo, but rather false beliefs about that general's early career.



I think we can say that *in general* a speaker intends to refer to the item that is the dominant source of his associated body of information. It is important to see that this will not change from occasion to occasion depending upon subject-matter. Some have proposed¹¹ that if in case 1 the historian says 'Napoleon fought skilfully at Waterloo' it is the imposter β who is the intended referent, while if he had said in the next breath '... unlike his performance in the Senate' it would be α . This seems a mistake; not only was what the man said false, what he intended to say was false too, as he would be the first to agree; it wasn't Napoleon who fought skilfully at Waterloo.

¹¹ K. S. Donnellan, 'Proper Names and Identifying Descriptions', in Davidson and Harman (eds), *op. cit.*, p. 371.

With this background, then, we may offer the following tentative definition:

- 'NN' is a name of x if there is a community C
1. in which it is common knowledge that members of C have in their repertoire the procedure of using 'NN' to refer to x (with the intention of referring to x)
 2. the success in reference in any particular case being intended to rely on common knowledge between speaker and hearer that 'NN' has been used to refer to x by members of C and not upon common knowledge of the satisfaction by x of some predicate embedded in 'NN'.¹²

(In order to keep the definition simple no attempt is made to cover the sense in which an unused institutionally approved name is a name.)

This distinction (between use-because-(we know)-we-use-it and use upon other bases) is just what is needed to distinguish dead from live metaphors; it seems to me the only basis on which to distinguish referential functioning of names, which may grammatically be descriptions, from that of descriptions.¹³

The definition does not have the consequence that the description 'the man we call "NN"' is a name, for *its* success as a referential device does not rely upon common knowledge that *it* is or has been used to refer to x .

Intentions alone don't bring it about that a name gets a denotation; without the intentions being manifest there cannot be the common knowledge required for the practice.

Our conditions are more stringent than Kripke's since for him an expression becomes a name just so long as someone has dubbed something with it and thereby caused it to be in common usage. This seems little short of magical. Suppose one of a group of villagers dubbed a little girl on holiday in

¹² For the notion of 'common knowledge', see D. K. Lewis, *Convention* (Cambridge, Mass.: Harvard University Press, 1969) and the slightly different notion in S. R. Schiffer, *Meaning* (Oxford: Clarendon Press, 1972). For the notion of 'a procedure in the repertoire' see H. P. Grice 'Utterer's Meaning, Sentence Meaning, Word Meaning', *Foundations of Language* (1968). Clearly the whole enterprise owes much to Grice but no commitment is here made to any specific version of the theory of communication.

¹³ And if Schiffer is right much more as well—see *Meaning*, chap. V.

the vicinity 'Goldilocks' and the name caught on. However suppose that there were two identical twins the villagers totally fail to distinguish. I should deny that 'Goldilocks' is the name of either—even if by some miracle each villager used the name consistently but in no sense did they fall into two coherent sub-communities. (The name might denote the girl first dubbed if for some peculiar reason the villagers were deferential to the introducer of the name—of this more below.)

Consider the following case. An urn is discovered in the Dead Sea containing documents on which are found fascinating mathematical proofs. Inscribed at the bottom is the name 'Ibn Khan' which is quite naturally taken to be the name of the constructor of the proofs. Consequently it passes into common usage amongst mathematicians concerned with that branch of mathematics. 'Khan conjectured here that . . .' and the like. However suppose the name was the name of the scribe who had transcribed the proofs much later; a small '*id scripsit*' had been obliterated.

Here is a perfect case where there is a coherent community using the name with the mathematician as the intended referent and a consequence of the definition would be that 'Ibn Khan' would be one of his names. Also, 'Malachi' would have been the name of the author of the biblical work of the same name despite that its use was based upon a misapprehension ('Malachi' means my messenger).¹⁴

Speakers within such traditions use names under the misapprehension that their use is in conformity with the use of other speakers referring to the relevant item. The names would probably be withdrawn when that misapprehension is revealed, or start a rather different life as 'our' names for the items (cf. 'Deutero Isaiah' etc.). One might be impressed by this, and regard it as a reason for denying that those within these traditions spoke the literal truth in using the names. It is very easy to add a codicil to the definition which would have this effect.

Actually it is not a very good reason for denying that speakers within such traditions are speaking the literal

¹⁴ See O. Eissfeldt, *The Old Testament: An Introduction* (Oxford: Oxford University Press, 1965), p. 441.

truth.¹⁵ But I do not want to insist upon any decision on this point. This is because one can be concessive and allow the definition to be amended without giving up anything of importance. First: the definition with its codicil will still allow many names to change their denotation. Secondly: from the fact that, in our example, the community of mathematicians were not denoting the mathematician it obviously fails to follow that they were denoting the scribe and were engaged in strictly speaking massive falsehood of him.

Let me elaborate the first of these points.

There is a fairly standard way in which people get their names. If we use a name of a man we expect that it originated in the standard manner and this expectation may condition our use of it. But consider names for people which are obviously nicknames, or names for places or pieces of music. Since there is no standard way in which these names are bestowed subsequent users will not in general use the name under any view as to its origin, and therefore when there is a divergence between the item involved in the name's origin and the speakers' intended referent there will be no *mis*-apprehension, no latent motive for withdrawing the name, and thus no bar to the name's acquiring a new denotation even by the amended definition. So long as they have no reason to believe that the name has dragged any information with it, speakers will treat the revelation that the name had once been used to refer to something different with the same sort of indifference as that with which they greet the information that 'meat' once meant groceries in general.

We can easily tell the story in case 2 of our Napoleon diagram so that α was the original bearer of the name 'Napoleon' and it was transferred to the counterfeit because of the similarity of their appearances and therefore without the intention on anyone's part to initiate a new practice. Though this is not such a clear case I should probably say that historians have used the name 'Napoleon' to refer to β . They might perhaps abandon it, but that of course fails to show that they were all denoting α . Nor does the fact that someone in the know might come along and say 'Napoleon

¹⁵ John McDowell has persuaded me of this, as of much else. He detests my conclusions.

was a fish salesman and was never at Waterloo' show anything. The relevant question is: 'Does this contradict the assertion that was made when the historians said "Napoleon was at Waterloo"?' To give an affirmative answer to this question requires the prior determination that they have all along been denoting α .

We need one further and major complication. Although standardly we use expressions with the intention of conforming to the general use made of them by the community, sometimes we use them with the *overriding* intention to conform to the use made of them by some other person or persons. In that case I shall say that we use the expression *deferentially* (with respect to that other person or group of persons). This is true of some general terms too: 'viol', 'minuet' would be examples.

I should say, for example, that the man in the conversation in the pub used 'Louis' deferentially. This is not just a matter of his ignorance; he could, indeed, have an opinion as to who this Louis is (the man he met earlier perhaps) but still use the expression deferentially. There is an important gap between

intending to refer to the ϕ and believing that $a =$ the ϕ ;

intending to refer to a

for even when he has an opinion as to who they are talking about I should say that it was the man they were talking about, and not the man he met earlier, that he intended to refer to.

Archaeologists might find a tomb in the desert and claim falsely that it is the burial place of some little known character in the Bible. They could discover a great deal about the man in the tomb so that he and not the character in the Bible was the dominant source of their information. But, given the nature and point of their enterprise, the archaeologists are using the name deferentially to the authors of the Bible. I should say, then, that they denote that man, and say false things about him. Notice that in such a case there is some point to this characterization.

The case is in fact no different from any situation in which a name is used with the overriding intention of referring to

something satisfying such-and-such a description. Kripke gives the example of 'Jack the Ripper'. Again, after the arrest of a man *a* not in fact responsible for the crimes, *a* can be the dominant source of speakers' information but the intended referent could well be the murderer and not *a*. Again this will be productive of a whole lot of falsehood.

We do not use all names deferentially, least of all deferentially to the person from whom we picked them up. For example, the mathematicians did not use the name 'Ibn Khan' with the *overriding* intention of referring to whoever bore that name or was referred to by some other person or community.

We must thus be careful to distinguish two reasons for something that would count as 'withdrawing sentences containing the name'

- (a) the item's not bearing the name 'NN' ('Ibn Khan', 'Malachi')
- (b) the item's not being NN (the biblical archaeologists).

I shall end with an example that enables me to draw these threads together and summarize where my position differs from the Causal Theory.

A youth *A* leaves a small village in the Scottish highlands to seek his fortune having acquired the nickname 'Turnip' (the reason for choosing a nickname is I hope clear). Fifty or so years later a man *B* comes to the village and lives as a hermit over the hill. The three or four villagers surviving from the time of the youth's departure believe falsely that this is the long-departed villager returned. Consequently they use the name 'Turnip' among themselves and it gets into wider circulation among the younger villagers who have no idea how it originated. I am assuming that the older villagers, if the facts were pointed out, would say 'It isn't Turnip after all' rather than 'It appears after all that Turnip did not come from this village.' In that case I should say that they use the name to refer to *A*, and in fact, denoting him, say false things about him (even by uttering 'Here is Turnip coming to get his coffee again').

But they may die off, leaving a homogeneous community

using the name to refer to the man over the hill. I should say the way is clear to its becoming his name. The story is not much affected if the older villagers pass on some information whose source is *A* by saying such things as 'Turnip was quite a one for the girls', for the younger villagers' clusters would still be dominantly of the man over the hill. But it is an important feature of my account that the information that the older villagers gave the younger villagers could be so rich, coherent, and important to them that *A* could be the dominant source of their information, so that they too would acknowledge 'That man over the hill isn't Turnip after all.'

A final possibility would be if they used the name deferentially towards the older villagers, for some reason, with the consequence that no matter who was dominant they denote whoever the elders denote.

6. *Conclusion.* Espousers of both theories could reasonably claim to be vindicated by the position we have arrived at. We have secured for the Description Theorist much that he wanted. We have seen that for at least the most fundamental case of the use of names (non-deferentially used names) the idea that their denotation is fixed in a more or less complicated way by the associated bodies of information that one could cull from the users of the name turns out not to be so wide of the mark. But of course that the fix is by causal origin and not by fit crucially affects the impact this idea has upon the statement of the truth conditions of existential or opaque sentences containing names. The theorist can also point to the idea of dominance as securing what he was trying, admittedly crudely, to secure with his talk of the 'majority of' the descriptions, and to the 'degree of fit requirement' as blocking consequences he found objectionable.

The Causal Theorist can also look with satisfaction upon the result, incorporating as it does his insight about the importance of causality into a central position. Further, the logical doctrines he was concerned to establish, for example the non-contingency of identity statements made with the use of names, are not controverted. Information is individuated by source; if *a* is the source of a body of information

nothing else could have been. Consequently nothing else could have been *that a*.

The only theorists who gain no comfort are those who, ignoring Kripke's explicit remarks to the contrary,¹⁶ supposed that the Causal Theory could provide them with a totally *non-intentional* answer to the problem posed by names. But I am not distressed by their distress.

Our ideas also point forward; for it seems that they, or some close relative, must be used in explaining the functioning of at least some demonstratives. Such an expression as 'That mountaineer' in 'That mountaineer is coming to town tonight' may avert to a body of information presumed in common possession, perhaps through the newspapers, which fixes its denotation. No one can be *that* mountaineer unless he is the source of that information no matter how perfectly he fits it, and of course someone can be that mountaineer and fail to fit quite a bit of it. It is in such generality that defence of our ideas must lie.

But with these hints I must leave the subject.

¹⁶ Kripke, *op. cit.*, p. 302.

Identity and Predication

A translation is one thing, a theory of meaning another. A manual of translation aims to provide, for each sentence of the language under study, a way of arriving at a quoted sentence of another language which has the same meaning. A theory of meaning, on the other hand, entails, for each sentence of the language under study, a statement of what it means. A translator states no semantical truths at all, nor has he any need of the concepts of truth, denotation, and satisfaction. Semantical truths relate expressions to the world, and can be stated only by using, not mentioning, expressions of some language or other. Such truths arise inevitably in the construction of a theory of meaning, for statements of what a sentence means can be made only by using an equivalent sentence of the theory's language, and, if the theory is to be finite, must be deduced from statements assigning semantical properties to its parts.

The two activities proceed under quite different constraints. Certainly a manual of translation must be finite, deriving translations for sentences from correlations of the sentences' parts, but a translator may use any segmentation of sentences he finds convenient. The finiteness demand upon a theory of meaning, on the other hand, goes to the very heart of what it is a theory of, since a language for which there is no finite theory of meaning is an unlearnable language.¹ And the semanticist aims to uncover a structure

From *The Journal of Philosophy* 72, 13 (17 July 1975), pp. 343-63. Reprinted by permission. This is a revised version of a paper read at Princeton University and to the 1974 York Conference on Logic and Semantics. I would like to thank H. Margain, S. Schiffer, B. Stroud, and B. Taylor for their help, and to acknowledge an especial debt to conversation with J. H. McDowell and to the writings of D. Davidson.

¹ See D. Davidson, 'Theories of Meaning and Learnable Languages', in Y. Bar-Hillel (ed.), *Logic, Methodology, and Philosophy of Science* (Amsterdam: North-Holland, 1965). These points on the relation between translation and the theory of meaning are due to Davidson.

in the language that mirrors the competence speakers of the language have actually acquired. This does not mean that he aims to uncover a theory that he supposes his subjects *know*, in any acceptable sense of that word. It means merely this: if (but only if) speakers of the language can understand certain sentences they have not previously encountered, as a result of acquaintance with their parts, the semanticist must state how the meaning of these sentences is a function of the meanings of those parts. He must assign semantical properties to the parts and state the general significance of the construction in such a way that a statement of what those sentences mean is deductively entailed. There may be more than one way of doing this.

Even when a translator arrives at a translation of a sentence by way of its parts, his procedure is quite different. Since he is not using the expressions, his manipulations do not have to be deductive. He can, for example, deliver translations of an infinite number of Spanish sentences by the simple rule:

Trans ('Es necesario que' S) =
'It is necessary that' (Trans (S))

Nor is there any reason why a translator should aim to minimize ambiguity, but it is central to the semanticist's concern that he provide, where possible, a uniform account of the semantic contribution of single lexical items. When Tarski constructed a theory of meaning for quantificational languages he was obliged to show that it was not an accident that the very same words 'and' and 'not' occur both as sentence formers and as predicate formers. By providing a uniform account of their roles, he thereby explained the validity of countless inferences connecting sentences containing the expressions in their different positions.

Translation being one thing and the theory of meaning another, indeterminacy of translation is one thing and indeterminacy in the theory of meaning another. It is a quarrelsome man who would bicker with Quine over the indeterminacy of translation—the constraints upon that enterprise being so slight. But Quine clearly intended his arguments to show, and wrote as though his arguments

had shown, indeterminacy in the concepts of reference, denotation, extension—all concepts of the theory of meaning. And perhaps it is not difficult to see why it should be thought that indeterminacy in one is tantamount to indeterminacy in the other. For, after all, although not every translation manual provides the basis for a theory of meaning, every theory of meaning must provide the basis for a translation manual, since it must, in effect, pair sentences of the jungle language with equivalent sentences of the theory's language. And in so doing the semanticist must draw upon exactly the same data, of native assent, dissent, and unsolicited utterance, which are available to the translator. The possibility is thus raised that different theories of meaning could be constructed each of which entails semantical properties for the totality of jungle sentences adequate to the data but which make quite different assignments of semantical properties to those sentences' parts.

That there is some such indeterminacy, even in the construction of homophonic theories of meaning for English in English, seems to me to be a virtual certainty. However, in this paper I do not wish to consider the question in general, but to examine whether one particular argument of Quine's can be used to establish the rather gross degree of indeterminacy in theories of meaning which it has so widely been thought to establish.

I

The argument I have in mind runs like this.² Some expressions that divide their reference differently ('rabbit', 'rabbit stage', 'rabbit part') and some expressions that do not divide their reference at all ('rabbithood', 'the rabbit fusion', 'rabbitheth') have incontestably the same stimulus meaning when used in one-word sentences. An expression may sensibly be regarded as a predicate only if it interacts with 'the apparatus of individuation' (the identity predicate, plurals, numerals, the definite article), and the stimulus conditions that trigger assent to the sentences in which such interaction occurs

² See *Word and Object* (Cambridge, Mass.: MIT Press, 1960), pp. 51-4, 68-79; *Ontological Relativity* (New York: Columbia, 1969), pp. 1-6, 30-5.

provide the only empirical evidence that bears upon what extension the expression, as a predicate, should be regarded as possessing. But the identification of the apparatus of individuation in a foreign language is empirically quite undetermined; the expression that one theory regards as the identity predicate, for example, may, with suitable adjustments, be treated by another as an expression for some distinct equivalence relation—both theories assigning to whole sentences a significance completely in accordance with the behavioural data. Consequently, whether an expression is a predicate at all, and, if a predicate, what extension it has, are matters undetermined by all actual and possible observations. Time without end.

When this argument is intended to show an indeterminacy in the theory of meaning, it must rest upon the belief that the sole reason a semanticist can have for treating an expression as a predicate with a particular divided reference is to account for that expression's interaction with the (putative) apparatus of individuation. This, it seems to me, is to identify the empirical location of the scheme of predication by means of a quite secondary anchoring. Not only is the scheme anchored independently, but the anchoring provided by the individuating apparatus is secure only because it is, in its turn, fixed to the scheme's primary anchor.

The primary function of construing an expression, *G*, as a term dividing its reference over rabbits—or trees—is to explain how the truth conditions of certain elementary, but compound, sentences into which it enters are determined by their parts. The apparatus of individuation may be entirely absent from such sentences. To see the notion 'what *G* is a predicate of' in this way is to see it as constrained by a theory of sentence composition into which it fits and which alone gives it sense. If this is right, it will appear quite unsurprising that the application of the concept is undetermined by data exclusively derived from the assent conditions of laconic one-word utterances and of sentences putatively involving interaction with the apparatus of individuation. For by concentrating upon such data we thereby disregard precisely those compound sentences which give the notion its point.

To see predication in this way is to see the intelligibility of Davidson's response³ to the need to account for the way the truth conditions of adverbially modified sentences are determined by their parts—his suggestion that such sentences involve predicates of runnings and whistlings. It was not the interaction of verbs like 'whistle' with devices from an already agreed to apparatus of individuation in such adverbially modified sentences which prompted his suggestion, for no such interaction there takes place. Certainly, Davidson has pointed to other, explicitly nominalized, sentences in which related expressions like 'a whistling' do interact with the apparatus of individuation,⁴ but his case for discerning predication in the adverbially modified sentences did not, or at least should not, rest upon these. For the same reason, it makes perfectly good sense to deny that we have predicates of objects—voices—in sentences such as 'He has a deep voice', despite the perfectly healthy interaction that occurs elsewhere between the expression 'voice' and the apparatus of individuation ('Thirty voices rang out', etc.).

I have said that to discern predicational structure in certain sentences is to provide a specific explanation of how their truth conditions depend upon their parts. Our questions must now be: What kind of explanation is afforded? What features must sentences exhibit to make that explanation appropriate?

The rest of this paper has the following structure. In the next section I try to answer the questions concerning predication which I have just raised. My strategy is to begin by considering impoverished languages in whose theory of meaning predication has no role to play at all. I then see how the languages might be enriched to make it not only plausible but necessary to explain the way their sentences are put together by invoking the semantic structure of predication. By focusing on this development I hope to bring out the point of introducing objects and their properties into semantics, and by seeing that point I hope we may see why certain languages

³ See 'The Logical Form of Action Sentences', in N. Rescher (ed.), *The Logic of Decision and Action* (Pittsburgh, Pa.: University Press, 1967). Intelligibility, but not necessarily correctness.

⁴ See 'The Individuation of Events', in N. Rescher (ed.), *Essays in Honor of Carl G. Hempel* (Dordrecht: Reidel, 1970).

might demand explanation by mention of rabbits and their properties, and not of rabbit parts and their properties. In the final section I bring the general theory to bear more specifically on Quine's arguments, by trying to show that Quine's suggested alternative theories of meaning for a language, plausibly treated as speaking of rabbits, are not workable alternatives at all.

I have two final introductory points. First, when I speak of discerning reference to and predication of *G*s, I have in mind treating sentences as involving quantification over the *G*s (as in 'A *G* is *F*', 'Some *G* is *F*') as well as treating sentences as involving singular definite reference to a *G*, as in 'This *G* is *F*'. My general argument will not assume that singular reference is behaviourally detectable⁵—I shall concentrate mainly upon indefinite sentences gaining reference to a particular location by *conversio ad phantasmata*.⁶

Second, I want to acknowledge that in the body of the paper I concentrate upon rather simple, but obviously central, cases of discourse about observable spatio-temporal phenomena. In order to keep this paper of manageable proportions I say nothing about more theoretical or abstract discourse. I have established to my own satisfaction that such discourse can be fitted into the framework I develop in this paper, but no such assurance is going to prevent a sense of unease in the reader when his eye turns from the vaunting structure to the examples upon which it is founded. As it is, I will not chew to everyone's satisfaction all I have bitten off, so there is nothing to do but make this acknowledgement and proceed.

II

1. Let us begin by considering languages in whose theory of meaning predication plays no part.

There might be a language, or quasi-language, which consisted of a finite number of utterance types used to respond communicatively to the publicly observable presence of some

⁵ But see below, sect. III, 4.

⁶ See P. T. Geach, *Mental Acts* (London: Routledge & Kegan Paul; New York: Humanities, 1967), p. 65.

general condition of the environment. There might be an utterance type appropriately used whenever it is raining in the vicinity of the speaker at the time of utterance, another whenever a rabbit or rabbits are in the vicinity at the time of utterance, and so on. Here there is no question of assigning any expression an extension, or of identifying any construction as that of predication, for there are no constructions to be explained.

Suppose now that such expressions were not only stood on their own but could be compounded, so that we had expressions like 'Red Water', 'Warm Fog', 'White Rabbit', and so on. Assuming that these expressions are not learned one by one, but are genuinely structured, we have a simple case of the general semantical problem. How are we to account for the significance of these compound expressions?

The answer depends, I think, upon what significance the compound sentences actually have. By trying these utterances out on the natives we might find that the compound sentences require for their truth merely a (reasonably sized) spatial overlap (at the time of utterance) of the features of the environment required to prompt assent to the constituents standing on their own. Thus, 'Red Water?' is assented to provided there is a bit of water in the vicinity which is red, and 'White Rabbit?' is assented to provided a decent-sized bit of the rabbit stuff in the vicinity is white.⁷

If these are indeed the truth conditions of the compounds, I would argue that it is a mistake to assign to the constituents satisfaction conditions, and to explain the construction as predication—thereby giving the sentences the overall import of 'There is a red water here', 'There is a white rabbit here'.

To treat an expression as a predicate is to associate with it a certain condition, upon whose satisfaction by objects depends the truth or falsity of the sentences in which the expression occurs. If we are to treat an expression as a predicate, we must be in a position, as theorists, to indicate, at least roughly, what that satisfaction condition is. But being in a position to identify every one of the empirical conditions in

⁷ The conditions need not be quite so simple; the speakers might distinguish red wine from white wine suffused with red light. But let us stay with the very simplest case.

which some object or objects satisfy the predicate does not enable us to identify such a condition, for by itself that information provides us with no clue as to what an object that *fails* to satisfy the condition would be like. (The only judgements of the form 'X does not satisfy F' we can make upon such a basis are those which are derived from the knowledge that nothing (here) satisfies F.) The trouble is, of course, that we have no clue as to the *identity* of the objects satisfying the predicate in those situations in which we are assumed that there are some such.

In the case of the language we are considering, the use made of 'Red' does not enable us to say, even roughly, what an object that fails to satisfy the predicate we inappropriately take it to be would be like, which means that we have no idea of what an object that satisfies the predicate is like. Thus, consider. Some drops of red ink fall into a swimming-pool. 'Red Water?' Certainly. But which object is it, even roughly, whose satisfaction of the predicate 'Red' makes the remark true? Is it the whole pool, or just the water immediately diluting the ink, or one of the indefinitely many intermediate alternatives? The language provides us with no way of answering these questions, to which we must somehow find an answer if the construction is predication. The indeterminacy we find here in the satisfaction conditions of 'Red' is not the beginning of an insight but the consequence of a blunder.

To discern predication in these sentences, then, is to assign to 'Red' and the other expressions in the language richer semantic properties than any that are required to account for, and therefore any that can empirically be warranted by, the constructions that actually exist in the language. It is to claim that those expressions are disposed to behave in certain ways in phantom contexts, contexts that do not exist in the language. For we have fitted the expressions out with semantical properties that will yield truth conditions for a whole range of sentences that cannot be expressed in the language: in particular those whose truth conditions are given by our sentences:

There is a water here which is not red.

There is a water here which is not red and also a water here which is red.

It does not seem to me to require dedication to pernicious mentalism, or to a view of the theory of meaning as an exercise in proleptic neurophysiology, to suppose that we should refrain from attributing to expressions of the languages we study semantic properties for which there is absolutely no evidence, even when it is possible to make such attributions without attributing to whole sentences resulting semantic properties that are actually inconsistent with the available evidence.⁸

In deciding against predicational theories of such simple languages we are in step with Quine, for whom the excess semantic power stems from the absence of the individuating apparatus. However, let us proceed to consider a language without any such apparatus, yet whose compound sentences seem to demand explanation in terms of predication.

2. I shall consider a very elementary example.

Let us suppose that there are expressions G_1, G_2, \dots, G_n which when uttered as queries, are assented to when material objects of various kinds are presented to the informant. They are, as one-word sentences, stimulus-synonymous with 'A rabbit!' (and, of course, 'A rabbit part!'), 'A man!', 'A hut!', and so on. There are also, we may suppose, expressions F_1, F_2, \dots, F_n which, when queried, are assented to when the environment manifests the presence of certain general features that do not require the presence of any specific kind of object — expressions stimulus-synonymous with our 'White!', 'Warm!', 'Furry!', 'Bloodstained!', and so on.

These expressions not only stand alone, and with a sentential negation device, but enter into complex sentences coupling one of the F terms with one of the G terms. (We do not

⁸ How should we account for such compounds as we have been discussing in this section? Quine suggests construing them as mereological compounds (see 'Grades of Theoreticity', in L. Foster and J. W. Swanson (eds), *Experience and Theory* (Amherst, Mass.: University of Massachusetts Press, 1970), pp. 9 ff.). But there may be difficulties of principle in the way of providing a straightforward semantic account in a metalanguage that contains no structures of a similar kind. The only way then to proceed would be by invoking the relations of 'partial denotation' and 'partial satisfaction' (see, e.g., H. Field, 'Quine and the Correspondence Theory', *Philosophical Review* 83, 2 (April 1974), pp. 200-28). Such an approach concedes *nothing* to those who wish to see predication in these sentences.

find any couplings of two G words.) In addition, we find the particle that produces sentential negation also occurring in concatenation with the F terms to yield an 'internal' negation; (not- $F G$) is syntactically and behaviourally distinguishable from Not- $(F G)$.

When we look at the assent conditions, the following picture emerges. We find that it is not sufficient for assent to $(F G)$, e.g., 'White Rabbit', that F and G both be assented to nor that there be an overlap between the features associated with F and with G . For example, 'White Rabbit?' can be dissented from even though some of the exposed rabbit stuff is white, even, indeed, though a decent-sized (rabbit-sized) continuous portion is white, as when, for example, several brown rabbits are so organized that their white tails are contiguous.

What is required for many of these compounds is that the F feature be distributed in a characteristic way in relation to the boundaries of a *single* object whose presence prompts assent to the queried G terms. And when the simple overlap principle does not seem to work for the affirmative sentence (as it does, for example, for the compound 'Bloodstained Rabbit') we find the assent condition of the internally negated sentences ('not-Bloodstained Rabbit) again show a sensitivity to the boundaries of an object, for assent requires the *absence* of the associated feature from the entire exposed surface of that object.⁹

These facts are not mysterious or empirically inaccessible, and they immediately rule out any hope of construing the G term as a feature-placing term, like 'Rabbiteth', and accounting for the construction as a kind of mereological compounding. We cannot accommodate the sensitivity of these sentences to the distribution of features or evidence within the confines of a single rabbit if we assign to the

⁹ I concentrate upon monadic predicates, but this is only for simplicity. For evidently we might have a series of words R_1, \dots, R_n assented to, when queried with suitable pairs of pointings, in the presence of certain relational features. When coupled with two G words ('Rabbit Darker than Fox?') we find the same sensitivity to the boundaries of rabbit and fox, for it is not sufficient for assent that there be a pair of points in relation to which we can say 'Here rabbit, and there fox, and here darker than there'. And whenever such a simple condition is sufficient (as for 'Rabbit Touching Fox') the internally negated sentence shows sensitivity to the boundaries of objects.

G term a significance the whole point of which is that it abstracts from the difference between one rabbit and two.

We can, however, explain the sentences' observed truth conditions by discerning predication. We can explain the contribution of the F term to any particular compound *in a way consistent with its capacity to enter into couplings with other G terms* by supposing it to be associated with a particular condition requiring a certain distribution of a feature in relation to an object's boundaries. And we can explain the contribution of the G term by supposing it to be associated not merely with a recurrent feature (its criterion of application) but with a particular set of identity conditions—a particular divided reference. And we treat the mode of composition as predication yielding a sentence true if and only if at least one object (in the vicinity) of the G kind satisfies the condition associated with the F term.

By speaking of 'boundaries' and of 'identity conditions' we have introduced identity into our explanation of these compounds. But by what right do we speak of identity here? Is it not merely that assent to these compounds requires that the F feature come shaped in a way determined by the G term—that the white must come rabbit-shaped? But this is *not* all; in order to account for these compound sentences we do have to suppose that the surface of the rabbit stuff upon which the incidence of the white feature is relevant to the truth of the judgement 'White Rabbit', must genuinely be a boundary that separates a single rabbit from things that are distinct.

We are, as semanticists, obliged to state how 'White' occurs in 'not-White'; the two expressions are clearly not associated with independent conditions. The behavioural evidence warrants, we may assume, this general principle for generating the semantical contribution of 'not' ϕ from that of ϕ : an object satisfies 'not' ϕ iff the object does not satisfy ϕ . It follows from this account¹⁰ that contradiction will result if both predicates are applied to the same object. Thus: the

¹⁰ Of course, this *may* not be the only account (which has the great merit, as Tarski showed, of unifying internal and external uses of 'not'). But I know of nothing which indicates that this is anything other than the barest theoretical possibility.

distribution of whiteness throughout just that area is relevant to the judgements 'White Rabbit' and 'not-White Rabbit' precisely because either judgement affirmed upon an insufficiently extended survey is liable to be *contradicted* by the other judgement, warranted by the condition of, and made with respect to, the same rabbit. In other words, the line tracing the area of relevance delimits that area in relation to which one or the other, but not both, of a pair of contradictory predicates may be chosen. And that is what it is for a line to be a boundary, marking something off from other things.

We touch here upon the deep connection between predication and identity, and thus with the rest of the apparatus of individuation. Like all logical constants, the identity predicate can be identified only by the inferential connections between sentences containing it and other sentences of the language. No two-place predicate is the identity predicate unless the way speakers use sentences containing it reveals a disposition to withhold contradictory predicates from the things identified—no matter how involved those sentences may be in rituals that look like counting or pluralization.

The strongest form of the dependence of the apparatus of individuation upon the predicative structure in the rest of the language is espoused by one who believes with Quine that the identity predicate is *definable* in terms of that structure. But it is not necessary to have such a belief to accept this: that although we can recognize predication without identifying an identity predicate—for the structure of sentences may force such a recognition upon us—we could never recognize an identity sentence save by its inferential relations to such predicative sentences—in particular, its use in conformity with the principle of the non-identity of discernibles.

So of course Quine was right to find the identification of the identity predicate underdetermined by the data he considered: local samplings of the stimulus meaning of particular candidate expressions. For upon such a basis one could not show that an expression behaves in the required way in relation to contradictory predicates. But, conversely, if identity is tied to the rest of the language, *this* underdetermination is a relatively insignificant matter: we may

suppose that what objects a language distinguishes and talks about is a matter embedded much deeper than Quine's talk of jiggling with the translation of the individuating apparatus would lead us to believe.

These points, which will re-emerge when we come to discuss Quine's suggested alternative translations in detail (see sections III.3 and 4), explain why I believe Quine's anchoring of the predicative scheme is secondary.

3. We have considered how we might be able to explain the truth conditions of compound sentences of the form (FG) and $(\text{not-}FG)$, given their manifest sensitivity to the identity conditions of rabbits, by suggesting that the sentences involve predicates of rabbits. (It is interesting to observe that we might be forced to this conclusion even though there was no explicit G term of divided reference with whose significance we were concerned.)

There are other kinds of compound sentences which present similar problems and which require similar treatment.

We might find utterances coupling two or more of the F words with a G word, forming such expressions as $(F_1 F_2 G)$ (e.g., 'Bloodstained White Rabbit') for assent to which it is not sufficient that $(F_1 G)$ and $(F_2 G)$ both be assented to, nor necessary that there be a spatial overlap between the features associated with F_1 , F_2 , and G . (No part need be both white and bloody.) There must be some principle explaining in general how F_1 , F_2 , and G here go together, and the truth conditions exclude an explanation either in terms of sentential conjunction or in terms of mereological compounding. These sentences are again sensitive to the identity conditions of G s: what is required for the truth of $(F_1 F_2 G)$ is not merely that there be evidence such as to warrant the separate judgements $(F_1 G)$ and $(F_2 G)$, but that such evidence pertain to the *same* G .

Or, more generally, we find that there are identity conditions we can associate with the G term and satisfaction conditions we may associate with the F terms so that a compound of the form $(F_1 \dots F_n G)$ is true if and only if a single object of the G -kind satisfies F_1 and satisfies $F_2 \dots$ and satisfies F_n .

It is on the basis of this latter sort of evidence that we might justify the claim that an expression F , demonstrably assented to upon the basis of a particular sensory modality, say the sense of smell, was a predicate of a material object accessible to several senses, and not merely a predicate of a smell, if a predicate at all. (Evidence of the first kind, relating to the incidence or distribution of evidence in relation to the object's boundaries, would not help us here.) We would have to show that F occurred in compound sentences involving predicates applied upon the basis of the other modalities.

4. Our question was: 'What is the point of introducing objects and their properties into semantics?', and I hope that one answer is becoming clear. The semanticist has to explain the observed assent conditions of certain compound sentences by attributing to the parts of those sentences properties consistent with their occurrences in all other contexts and by characterizing the construction in such a way that those conditions may be deduced. Objects have a contribution to make to this enterprise when the way the assent conditions of the compounds depend upon the parts may be explained by mentioning the identity conditions of those objects; or, that is to say, when the systematic mastery¹¹ of those sentences requires mastery of the identity conditions of those objects.

This way of putting the point is liable to be misunderstood. It is not open to dispute that the truth or falsity of certain sentences turns upon the identity and distinctness of rabbits. But of course, turning upon the identity and distinctness of rabbits *just is* turning upon the holding or failure to hold of a certain equivalence relation among rabbit parts, or rabbit stages. The behavioural evidence merely shows that the ability to use these compound sentences requires the ability to sort situation pairs into two: those in which the same rabbit is involved in both from those in which the same rabbit is not involved in both. It is a leap of theory (precisely the theory of this paper) to represent this ability as the mastery of the identity conditions of rabbits in some strict sense, since it might rather be represented as the mastery of these alternative equivalence relations.

¹¹ I.e. mastery based upon familiarity with the parts.

The suspicion that we are cheating can be laid to rest in the way that is best with all such suspicions—by showing that we need not cheat. All we are entitled to is the use of the phrase 'mastery of the identity conditions of As ' in a weak sense, in which it means the ability to make that distinction in the world which is made by any sentence stimulus-synonymous with 'That is the same A as that'. But this is all we need. It is the fact that the compound sentences we have been considering require mastery of the identity conditions of rabbits in this weak sense that generates the semantical problem that discerning reference to rabbits solves. And the observation that there are such sentences is not trivial. For, from the fact that there are sentences whose systematic mastery requires, in this weak sense, mastery of the identity conditions of rabbits, it simply does not follow that there are sentences whose systematic mastery requires, in the same weak sense, mastery of the identity conditions of rabbit parts or of rabbit stages.

With that clarification, the position we have arrived at is this. To say that an expression has a particular divided reference makes sense only in the context of the explanation of compound sentences. To decide that a term divides its reference over rabbits is to decide that the sentences in which it occurs involve predication of rabbits. And to decide that a set of sentences involves predication of rabbits is to identify the way those sentences' assent conditions are generated from their parts as depending upon the identity conditions of rabbits, and so systematic mastery of those sentences requires mastery of the identity conditions of rabbits.

I propose the following constraint: it is unacceptable to discern predication of As in sentences of a given kind unless those sentences are such that their systematic mastery requires mastery of the identity conditions of As .

This constraint follows from the previous considerations when these are taken together with the general principle that prohibits the attribution of semantic properties for which there is no evidence. As we have seen, to assign to an expression, G , a particular way of dividing its reference is to fit it out for interaction with other expressions of the language in such a way that the truth or falsity of the resulting sentences

turns upon the distribution of evidence in relation to that division of reference. If there are such sentences, then their mastery will be such as to require mastery of that division of reference. Therefore, if we have assigned to an expression a divided reference in defiance of the constraint, this can only be because there are no such sentences. And if there are no such sentences, the claim that *G* divides its reference in just that way must be entirely without foundation. It is the wholly empty claim that *G* would behave in a certain way in phantom contexts—contexts that do not occur in the language.

I do not think there is much room for argument over the claim that the sentences of a language naïvely treated as speaking of rabbits will manifest the kinds of dependence upon the identity conditions of rabbits which I have outlined, and will fail to manifest such dependence upon the identity conditions of rabbit parts or of rabbit stages. And I expect at least a slight inclination to agree that this striking asymmetry has *something* to do with our readiness to suppose that users of such languages were speaking of rabbits—*something* to do with the assignments of predicational structure issuing from our, no doubt naïve and inchoate, semantical theory. But it would be unreasonable to expect any great confidence to be placed in this link between predication and identity until the striking dependence upon the rabbits can be turned from an argument in favour of discerning predicates of rabbits into arguments against alternative semantical proposals. This is what I attempt in the next section.

III

1. Let us now turn to Quine's specific illustrative alternatives. For ease of exposition, let us suppose that the language we are concerned with is that elementary fragment of English, 'Elementary English' naïvely taken as speaking of rabbits and other enduring material objects.

I will begin by discussing those semantical proposals on which 'rabbit' is not construed as a predicate at all. I will

deal with them briefly, since they seem to be obviously unworkable.

The language is naïvely taken as speaking of rabbits and their properties because it contains sentences like 'White rabbit' and 'Bloodstained white rabbit', which are sensitive in the required way to the identity conditions of rabbits. As far as I can see, this straight away eliminates the feature-replacing alternative, 'rabbiteth', since I fail to see how truth-conditions that are sensitive to the distinction between one rabbit and two can be generated on the basis of elements that are not.

Let us now consider the semantic theory on which 'rabbit' is treated as a singular term designating the universal, Rabbithood. In order to generate an appropriate truth-condition for the sentence 'A white rabbit!' or 'A rabbit (here) is white', the expression 'white' has to be regarded as a predicate satisfied by an object iff that object has a white instance, so that the whole sentence is rendered by the stimulus-synonymous: 'Rabbithood has a white instance (here)'. But how then are we to deal with the internally negated sentence, 'A non-white rabbit!', or 'A rabbit (here) is not white'? The truth conditions cannot be generated by predicate negation out of the assignments made to 'white' and 'rabbit'; that way we can produce only the truth-conditions of 'Rabbithood has no white instance (here)', which, as the linguists would say, is sharply non-synonymous.

Nor is there any way to generate the truth conditions of 'A bloodstained white rabbit!', or 'A white rabbit (here) is bloodstained', out of the singular term, 'Rabbithood', and the unbreakable predicates '. . . has a white instance (here)' and '. . . has a bloodstained instance (here)'. We have no way of registering the requirement that one and the same rabbit be both white and bloodstained.

Exactly the same difficulties lie in store for the other theory on which 'rabbit' is treated as a singular term, this time denoting the rabbit fusion, once we have adjusted the satisfaction conditions of 'white'. But to what should they be adjusted? '. . . iff *x* has a white part' is too weak, given brown rabbits with white feet, and '. . . iff *x* has a white, rabbit-sized part' is too weak, given the earlier brown rabbits

with their contiguous white tails. The only condition I can think of is '. . . iff x has a part that is a white rabbit', which is ruled out by the disposition of 'white' to couple with terms other than 'rabbit'.

2. Taking 'rabbit' to divide its reference over rabbit parts requires amending the satisfaction conditions of the expressions with which it couples. For if we continued to say, for example, that white things satisfied 'white', then we could demonstrate that 'rabbit' cannot be true of rabbit parts by citing the dissent to the query 'White rabbit?' elicited by the presentation of a brown rabbit with a white foot. However, the modification required appears simple: 'white' must be satisfied only by things which are parts of a white rabbit. Not even the white foot of a brown rabbit is a part of a white rabbit.

We now face one immediate, and fundamental, difficulty. Having changed the extension of 'rabbit' and therefore the extent of a single 'rabbit', we have been able to contrive a sentence with the same overall stimulus meaning only by incorporating an indication of the required incidence of whiteness into the satisfaction condition of the predicate—hence, '. . . iff x is part of a white rabbit'. But that prevents 'white' from coupling with other terms; when speakers of Elementary English say 'White hut!', they do not mean that any hut, or hut-part, is part of a white rabbit.

The difficulty appears trivial; surely we can manage something more general. Well, '. . . iff x is part of a white thing' will not do, since there are several rabbit parts which are parts of a white thing (in this case, a white foot) when the white-footed brown rabbit is present, yet 'White rabbit?' is dissented from. The difficulty is this: in order to accommodate such a remark as 'A clean carburettor!', we must find a fitting ϕ in the satisfaction condition: x satisfies 'clean' iff x is part of a clean ϕ , where ϕ is both general and yet in this case true of just the carburettor—not of the engine or car going up, or the float chamber and throttle slide going down. How

can this be, when we also have to accommodate 'A clean engine!' and 'A clean float chamber!?'¹²

However, let us waive this difficulty; for another more interesting difficulty emerges. The change we had to make in the satisfaction conditions of 'white' is required quite generally; if the satisfaction conditions of F were given by use of the predicate W on the old scheme, they must be given by use of the predicate 'is a part of a W rabbit' on the new. We should not be surprised to find the following consequence: *different parts of the same rabbit are indistinguishable by the predicates of the language*. Consequently, by Quine's absolute and objective criterion¹³ of what two-place predicate to count as the identity predicate, the predicate the theory attempts to treat as 'is a part of the same rabbit as' has the substitutivity property of, and therefore, for Quine, is, the identity predicate. It turns out to be no more in accordance with Quine's principles than it is with mine to discern predication of rabbit parts in this discourse.

One does not have to accept the identification of indiscernibles as an absolutely universal maxim of translation¹⁴ to suppose that this indistinguishability points to a mistake, and the mistake it points to should be clear from our earlier discussion. We have associated with 'rabbit' criteria of identity that dispose it to behave in such a way that a sentence of the form 'A rabbit (here) is ϕ and a rabbit (here) is not- ϕ ' may be true given a certain distribution of evidence in a scene in which just one rabbit is present; yet this disposition

¹² This is a good place to illustrate the difference between the jobs of the translator and the theorist of meaning, and how that difference bears upon the argument. The translator can say: where $\text{Trans}(G) = \text{'is a part of a } \overline{W}$, for all predicates F , $\text{Trans}(FG) = \text{'is a part of a } \overline{\text{Trans}(F)\overline{W}}$, and can continue to affirm: $\text{Trans}(\text{'white'}) = \text{'white'}$. This is why the indeterminacy of translation is such a trivial affair, since, as this example shows, the 'pairing' of 'rabbit' with 'rabbit part' and 'white' with 'white' has no semantical significance whatever. Translation is to the theory of meaning as constructing cardboard houses with scissors and paste is to building real houses.

¹³ Quine's own words; see 'Reply to Professor Marcus', in *Ways of Paradox* (New York: Random House, 1966), p. 178.

¹⁴ Indiscernibility may be due to genuine, *ad hoc* impoverishment. The indiscernible objects will then be only accidentally related, and identifying indiscernibles will lead to construing a man as talking about items of which he has not had the opportunity to gain knowledge. Such considerations clearly do not apply in this case.

can never be exercised, since the language does not contain a suitable candidate for ϕ .

It might be thought that this difficulty arises only because the language we have been considering is very elementary. For if the language also contained some unstructured expressions whose satisfaction conditions were given on the old scheme by use of 'is partly red' and 'is partly green', then, when we permute, it might be thought that these expressions could be given conditions satisfied by some but not all parts of the same rabbit. Thus the sentence whose truth-conditions used to be given by 'A rabbit here is partly red' is now rendered by the use of 'A rabbit part here is red'—which has incontestably the same stimulus meaning.

This theory, however, will not work if the old theory did. For the sentence 'A rabbit here is partly red and partly green' would occasionally elicit assent, inexplicable upon the new theory, under which the sentence emerges as true iff there is a rabbit part present which is both red and green.

3. It is the occurrence, in Elementary English talk involving the word 'rabbit', of compound sentences of the form:

This rabbit was F_1 and will be F_2

A rabbit is F_1 and will be F_2 ,

which requires and supports the supposition that 'rabbit' is a term that divides its reference over persisting animals. For these compounds are manifestly sensitive to the distribution, within the life of a single rabbit, of the evidence associated with the F terms.

How do these sentences fare when 'rabbit' is taken as true of rabbit stages? If we eliminate the context-dependence of tense in the most familiar way,¹⁵ we arrive, in effect, at the following general clauses (where t_u is the time of utterance):

x satisfies 'is' \widehat{F} iff $\langle x, t_u \rangle$ satisfy F

x satisfies 'was' \widehat{F} iff $(\exists t') [(\text{Before } t_u, t') \text{ and } (\langle x, t' \rangle \text{ satisfy } F)]$,

¹⁵ I am going to assume that treating tenses as operators requires an ontology of persisting things.

with the base clauses such as the following:

$\langle x, t \rangle$ satisfy 'warm' iff there is something warm at t , of which x is a stage.

Any such treatment again has the consequence that whatever is affirmed of one stage of a rabbit is true of every other stage of the same rabbit.

It might appear that there is a way of avoiding this difficulty. We might try to deal with time reference by the following clauses:

x satisfies 'is' \widehat{F} iff $\text{At}(x, t_u)$ and x satisfies F

x satisfies 'was' \widehat{F} iff $(\exists t') [\text{Before}(t_u, t') \text{ and}$

$(\exists y) (\text{Co-membered}(x, y) \text{ and } \text{At}(y, t') \text{ and } y \text{ satisfies } F)]$.

Then we would have base clauses of the form:

x satisfies 'warm' iff x is a stage of a thing which is warm while that thing is warm

Now, it appears, we have predicates truly affirmed of some stages which are not true of every stage in the life of the same animal. For example, only stages after a warm stage in the life of a rabbit are stages that satisfy 'was warm'.

This scheme, however, does not appear to work. We are supposing that an object satisfies the tensed predicate 'was warm' iff it is a stage later in the life of some object than some stage which satisfies the simple predicate. But this does not get the truth conditions right. 'A rabbit was running' may be true even though there is no stage of a rabbit later than some running stage—the running stage might have been the last.

4. I have deliberately placed no weight upon arguments that rest upon the reasons a semanticist might have to treat sentences as containing singular terms, for I am impressed by the fact that my case can be made without them. However, it is worth remarking that the identification of an expression as a demonstrative singular term, and the assignment to the expression with which it couples of a particular divided reference, together serve to explain another way in which sentences are sensitive to identity.

We find a bewilderment response elicited from speakers of Elementary English by sentences of the form 'This rabbit is *F*' whenever they are issued in connection with a scene in which no rabbits are discernible or no *one* rabbit is distinguished from several that are. Quine acknowledges¹⁶ that one *might* use such behaviour to spot the singular terms, but 'only as a voluntarily added maxim for relieving indecision among otherwise equally eligible systems of analytical hypotheses'. But we should not be influenced by this assertion until we have some idea how the rather peculiar pattern of behaviour is to be explained on these alternative hypotheses.

The way these simple sentences manifest sensitivity to the identity conditions of, say, rabbits, presents characteristic difficulties for the alternative semantical proposals upon which 'rabbit' divides its reference differently—difficulties which have been obscured by two false assumptions. The first is the assumption that every act of demonstrative reference is accompanied by a pointing gesture. The second is expressed by Quine¹⁷ when he says: 'Point to a rabbit, and you have pointed to a stage of a rabbit, to an integral part of a rabbit.' Point to a rabbit and you have thereby pointed to *many* parts and stages. If 'this' is to be treated as a demonstrative singular term, and 'rabbit' is to divide its reference over rabbit parts, we must suppose some fixed principles for dividing *everything* into parts—as upon a butcher's wall chart. For dividing things into stages we need something more complicated—a fixed point of origin, say the birth of Christ, and a determinative length of time, say a day. But these desperate manoeuvrings serve only to re-emphasize how far we are being pushed into assignments of utterly unempirical dispositions. Midnight turns out to be a magic moment, semantically speaking; how odd that the language does not have the resources to voice the significance of its passing.

So 'this' cannot be a demonstrative singular term. How, then, are the sentences containing it to be accounted for?

¹⁶ 'Reply to Strawson', in D. Davidson and J. Hintikka (eds.) *Words and Objections* (Dordrecht: Reidel; New York: Humanities, 1969).

¹⁷ *Word and Object*, p. 53.

5. The last suggestion of Quine's I wish to discuss also concerns singular terms. Quine suggests¹⁸ that, from the point of view of the radical semanticist, there is no distinguishing the singular term 'Ortcutt' from a general term true of Orcutt's several parts and stages—no distinguishing, that is, save by reference to the elusive apparatus of individuation.

It is characteristic of names that they couple with other expressions of the language, and in so doing they exhibit two crucial and behaviourally detectable characteristics:

(1) We never find (F_1 Orcutt) and (F_2 Orcutt) assented to without the compound ($F_1 F_2$ Orcutt) being assented to.

(2) For those who are prepared to use the name, there is no distinction between internal and external negation; we never find Not-(F_1 Orcutt) receiving a different verdict from (not- F_1 Orcutt).

It is possible to explain these properties of the expression 'Ortcutt' while treating 'Ortcutt' as a general term true of several things, upon one assumption only: that whatever is true of one Orcutt is true of all. That is to say, expressions that are treated with reason as names may be treated as general terms true of several things only if those things are indiscernible by the predicates with which the expression couples. This seems to suggest that the indeterminacy here is also illusory.

IV

There is a pattern in the difficulties encountered by Quine's alternative semantic hypotheses. Ignoring the seriously inadequate feature-placing suggestion, 'rabbiteth', the novel theories cut the reference of 'rabbit' either coarser or finer than it is cut on the orthodox theory. The coarser theories appear not to work at all. The finer theories have a better chance of working, but involve attributing to the speakers of the language unwarranted dispositions. The difficulties for the coarser theories stem from the *presence* of sentences that manifest dependence upon the identity and distinctness of rabbits; the difficulties for the finer theories from the

¹⁸ *Ibid.*, p. 52.

absence of sentences manifesting a similar dependence upon the objects they discover.

I do not pretend to have shown that a viable semantic theory based upon one of Quine's suggestions cannot be constructed. Perhaps an ingenious person will show that the difficulties are less severe than they look, and thereby make something of Quine's *example* of the indeterminacy of semantics. But I have been examining an *argument* of Quine's for the indeterminacy of semantics:

[T]he whole notion of terms and their denotation is bound up with our own grammatical analysis of the sentences of our own language. It can be projected on the native language only as we settle what to count in the native language as analogues of our pronouns, identity, plurals, and related apparatus.¹⁹

and because of the

. . . broadly structural and contextual considerations that could guide us to native translations of the English cluster of interrelated devices of individuation . . . [t] here seem bound to be systematically very different choices [of the semantic treatment of 'gavagai'] all of which do justice to all dispositions to verbal behavior on the part of all concerned.²⁰

I have tried to show that this argument is not sound.

¹⁹ 'On the Reasons for the Indeterminacy of Translation', *Journal of Philosophy* 67, 6 (26 March 1967), pp. 181-2.

²⁰ *Ontological Relativity*, p. 34.

3

Semantic Structure and Logical Form

The validity of some inferences is said to be explained by reference to the meanings of the particular expressions occurring in them, while that of other inferences is due, rather, to the way in which the sentences are constructed out of their parts. The inference from 'John knows that snow is white' to 'Snow is white' is given as an example of the first type of inference; for it is said to be explained by providing an analysis of the semantical primitive 'knows'. The inferences from 'John ran breathlessly' to 'John ran' and from 'John is a large man' to 'John is a man' may, tentatively, be taken to be examples of the second type.

The distinction I have gestured towards is not without its intuitive appeal, and for many years philosophers have been trying to provide a basis for it in harmony with what they took to be its importance. The debate centred upon, and eventually ran aground upon, the problem of identifying a set of expressions as the logical constants. For if we are determined to say that the inference from ' P and Q ' to P is valid in virtue of structure, then the distinction between it and the detachment inference with 'knows' must reside in some difference between 'knows' and 'and'.

Donald Davidson has given the notion of structurally valid inference new life and importance by locating it within a highly suggestive theory of semantic theories. In his writings there appears to be support both for the conviction that there is an important difference between the two types of

From *Truth and Meaning: Essays in Semantics*, eds, G. Evans and J. H. McDowell (Oxford: Clarendon Press, 1976). Reprinted by permission of Oxford University Press. An earlier version of this paper was read at a weekend conference on 'Language and Meaning' at Cumberland Lodge, in November 1973. I would like to thank the following for their help and encouragement: D. Davidson, D. Isaacson, J. H. McDowell, and P. F. Strawson. Special thanks are due to B. Taylor for help with both technical and theoretical problems. [For an afterthought to this paper, see Appendix II, p. 405. Ed.]

inference, and also for the view that, when an inference has been shown to be structurally valid, a deeper explanation of its validity will have been provided. He writes concerning one inference:

By saying exactly what the role of [a certain recurrent element] is and what the roles of the other significant features of the sentence are, we will have a deep explanation of why one sentence entails the other; an explanation that draws upon a systematic account of how the meaning of each sentence is a function of its structure.¹

My interest in the notion of structurally valid inference was awakened by these writings of Davidson, and like him I am ambitious to make precise that distinction which underlies the intuitions with which we began in such a way as would support the claim that, by showing an inference to be structurally valid, we thereby provide a deep explanation of its validity.

In the first part of this paper I shall examine the account of structurally valid inference which Davidson's writings suggest and try to indicate why I think that we have not there reached a finally satisfactory account of the matter. In the second part of the paper I shall sketch another approach to the idea with which I began. Finally from the vantage point provided by the sketch I shall look briefly at some recent and not so recent proposals concerning semantic structure.

I

We can distinguish two kinds of definition or explication in semantics which I shall label 'immanent' and 'transcendent'.

One provides an immanent definition of some semantical term W if one does not define it absolutely but rather defines the notion ' e is W according to theory T '. One provides a transcendent definition when the definition contains no such relativity to a theory; when one says, rather, what a theory *ought* to treat as W .

Our pre-systematic theorizing about semantics has provided us with a set of terms not all of which may suitably be provided with transcendent definitions. Although some theories

¹ D. Davidson, 'Action and Reaction', *Inquiry* 13 (1970), p. 144.

will be right to treat an expression e as W (for such a term), and some wrong, this will be because some theories are right or wrong overall. In so far as empirical considerations bear upon the correctness of the claim that an expression is W , they bear globally upon the theory according to which it is W , and cannot be brought into any more direct relation with that feature of the internal constitution of the theory in virtue of which it may be said to be treating e as W .

An example of such a semantical term might be 'is a designator'. If there is a theory of meaning, satisfying the global constraints upon such theories, which treats some expression e as a designator, and another such theory which does not, there is, perhaps, no sense to the question 'Is e really a designator?' because, perhaps, no evidence can additionally be brought to bear upon it.

However, this is not always the case. Consider the notion ' e is (semantically) composite'. If the theories with which we are concerned are recursive definitions of truth, an immanent definition of this notion is readily obtainable:

e is semantically composite according to T iff

- (i) e is a semantical unit according to T and
- (ii) there is no base or recursive axiom assigning e a semantical property in T .

But we should not rest content with such an immanent definition, since there is a consideration which bears directly upon the correctness of a theory's decision as to whether to treat an expression as semantically composite. It is this. Are the speakers of the language for which T is a semantical theory capable of understanding new expressions constructed in the way e is constructed, upon the basis of their understanding of the (syntactic) parts of those expressions? This consideration provides materials for the construction of a transcendent definition of the notion ' e is semantically composite'.

Clearly, a question which we must constantly keep in mind is whether the notion of structurally valid inference admits of a transcendent definition; whether there are considerations which bear directly upon the correctness of decisions concerning structural validity of inferences.

I shall not attempt to answer this question now. My present concern is with the way in which Davidson has made the notion of structural validity precise. It seems evident that he has provided an immanent definition, which I shall now try to make explicit.

I shall assume familiarity with Davidson's conception of a theory of meaning as a theory of truth. It seems clear from his work that a semantic theory for any natural language would in his view have the following two tiers:

- (1) a theory of truth, conforming to Tarski's Convention T, for a suitably chosen, regimented, but still interpreted, language, which will probably be, in large part, a fragment of the natural language, but could be a constructed language;
- (2) a set of translation rules mapping any sentence of the natural language not provided with truth conditions at level (1) on to a sentence for which truth is defined directly.

Showing an inference to be structurally valid has a slightly different significance when the sentences concerned are in the fragment for which truth is directly defined and when they are not. I shall consider these in turn, starting with the fundamental case of sentences for which truth is directly defined.

Davidson writes: 'there is no giving the truth conditions of all sentences without showing that some sentences are logical consequences of others'.² This is explained as follows: 'A truth definition does not distinguish between analytic sentences and others except for sentences which owe their truth to the presence alone of the constants which give the theory its grip on structure.'³ In a later paper he expands this idea: 'but it will be evident from a theory of truth that certain sentences are true solely on the basis of the properties assigned to the logical constants. The logical constants may be identified as those iterative features of the language which require a recursive clause (not in the basis) in the definition of truth or satisfaction.' He continues, in an explicit acknowledgement of the immanence of the definition:

² D. Davidson, 'Semantics for Natural Languages' in *Linguaggi nella Società e nella Tecnica* (Edizioni di comunità, Milan, 1970, pp. 184-5).

³ D. Davidson, 'Truth and Meaning', *Synthese* 17 (1967), p. 318.

'Logical form, on this account, will of course be relative to the choice of a metalanguage (with its logic) and a theory of truth.'⁴

Let us say that the conditional 'If S_1 is true, . . . , and S_{n-1} is true, then S_n is true' is the *validating conditional* of the inference $S_1 \dots S_{n-1} \vdash S_n$. I think Davidson's idea is that an inference is structurally valid according to a theory T if and only if its validating conditional is a *semantic* consequence of the theory's recursive clauses.

Thus to take a perhaps excessively simple case, a theory of truth which has the clause

for any sentences S and S' , \widehat{S} 'and' $\widehat{S'}$ is true if and only if S is true and S' is true

treats 'and' as a logical constant and attributes to it the semantic property of forming truths when and only when conjoining truths. It is a (semantic) consequence of this characterization, independent of the characterization of elements given in the base clauses, that if ' P and Q ' is true, then P is true. The theory thus treats the inference ' P and Q ' $\vdash P$ as a structural inference.

However, it appears possible to treat a great many expressions which are not typically regarded as logical constants in recursive clauses of the truth definition. (I ignore the fact that, on this view, the standard semantic treatment of identity shows it not to be a logical constant.) In particular, such a treatment appears possible for an attributive adjective, such as 'large'. The full details are given in the appendix, but the leading idea would be to have a clause in the definition of satisfaction along the lines of:

for all (possibly complex) monadic predicates ϕ , a satisfies 'large' $\widehat{\phi}$ if and only if a is a large satisfier of ϕ .

Such treatment takes seriously the grammatical status of 'large' as a predicate modifier. We will of course need some deductive machinery allowing for substitution within the scope

⁴ D. Davidson, 'In Defense of Convention T', in H. Leblanc (ed.), *Truth, Syntax and Modality* (Amsterdam: North-Holland, 1973), p. 81. This way of identifying the logical constants is also to be found in M. A. E. Dummett, *Frege* (London: Duckworth, 1973), p. 22.

of the metalinguistic modifier 'large', if the semantical material introduced into the right-hand side is to be eliminated and a homophonic biconditional thus to be derived. But an obviously valid rule permitting merely the substitution predicates which are provably equivalent in the metalanguage will suffice.

Then, such a theory will have shown the inference from 'X is a large man' to 'X is a man' to be formally valid. And if the theory were to treat 'small' in a parallel fashion, with its recursive clause, the same will hold for the inference from 'X is a large man' to 'X is not a small man', since it would be a semantic consequence of the two axioms taken together with the axiom for 'not' that if an object satisfies 'large man' it also satisfies 'not a small man'.

I am not claiming that this is an admirable semantical proposal. The point of the example is to cast doubt upon the idea that we have here captured the basis of the distinction with which we started, even in its application to those cases (e.g. the truth-functional connectives and the quantifiers) for which the background theory is more orthodox. For we were surely encouraged to entertain hopes that a theory which showed an inference to be structurally valid would provide us with a kind of explanation of its validity which would *contrast* with those in which final appeal is made to the inferential properties of particular expressions. And the conviction that the recursive clause of 'large' provides us with no insight into why the detachment inference is valid is surely reinforced by the observation that an entirely parallel recursive treatment will suffice for such apparently heterogeneous modifiers as 'good' and 'breathlessly', and for 'fake', which does not even sustain the inference.

If you feel like saying that you have as yet no idea of what kind of semantic role 'large' is supposed to be playing in the recursive clause of the metalanguage, what kind of function from what kinds of elements to what kinds of elements it introduces, I must ask you to be patient in the knowledge that the drift of the paper is on your side. But you should also realize that you seem to be appealing to semantic notions richer than any Davidson appeals to in characterizing the notion of structurally valid inference.

I have interpreted Davidson's talk of a theory's *entailing* a validating conditional in terms of semantic consequence. But this is to render uninterpretable his own observation that what inferences count as structural will be relative to the *logic* of the metalanguage. I adopted this interpretation because to introduce mention of the deductive apparatus of the metalanguage simply invites the question: 'How much is relevant to determinations of structural validity?' In general, the inference pattern in the metalanguage which is necessary for providing the validating conditional with respect to a certain object-language pattern of inference is that very same pattern of inference. If the logic of the metalanguage contains the detachment inference for the metalinguistic modifier 'large' we can validate the detachment inference in the object language, if not, not.

Perhaps there is a limit upon metalinguistic logic which would save the notion of structurally valid inference, interpreted in terms of what consequences are provable from the axioms, from free-wheeling entirely. Count as logic for deducing the relevant validating conditionals just the logic the theory requires for deriving its T-sentence theorems.

Certainly this is far from arbitrary, but the results are meagre, since very little logic is necessary. It is known that an adequate theory of truth for a first-order language susceptible of a classical interpretation can be constructed in a metalanguage whose logic is intuitionistic, and so, on this account, not all classically valid inference patterns will be structural. Proof of the T-sentences requires, on the whole, substitution inferences; it is certain that, provided the biconditional is undefined, neither the detachment inference with 'and', nor that with 'large', will be required for such proofs.

Let us now turn to the other way in which an inference may, on Davidson's view, be shown to be formally valid; when the sentences concerned do not have their truth conditions defined directly by the theory of truth, but are mapped on to sentences which do. If $F(S)$ is the sentence in the fragment on to which S is mapped by the translation rules, then $S_1 \dots S_{n-1}$ formally entail S_n iff $F(S_1) \dots F(S_{n-1})$ formally entail $F(S_n)$.

Since this manoeuvre rests upon the idea of formally valid

inference we have discussed, it clearly inherits any difficulties we were able to point to in that idea. But there are additional questions and a slightly new perspective.

I do not think Davidson would want to claim that the methodology underlying this section of the semantic theory has been entirely worked out. What considerations should guide us in our choice of the background language which is to provide the kernel? Save for eschewing any commitment to first-order languages, little has been said about this. Does the chosen fragment even have to be a fragment of any natural language; and if not, what defence can be provided for, or significance attached to, the requirement that the semantics for a natural language be stateable as a homophonic truth theory? Is the relation between S and $F(S)$ to be incorporated into the idea of semantic theory as having something to do with how speakers understand a natural language, and if so, how?

However, our immediate task is to consider the impact the construction of this additional tier of our theory has upon the notion of formally valid inference. It appears at first that it must reinforce the impression of arbitrariness which we brought away from our consideration of the leading idea.

For consider again the sentence 'John is a large man'. If we decide that our fragment is to be first-order, and thus without predicate modifiers, it appears that we have the choice of mapping our sentence either on to the sentence

Large-for (John, $\{y: \text{Man}(y)\}$)

or on to the sentence

Large-for* (John, $\{y: \text{Man}(y)\}$) & John $\in \{y: \text{Man}(y)\}$.⁵

If we adopt the first we determine that the detachment inference is to be accounted for by the analysis of the primitive relation 'Large-for', while if we adopt the second, using the primitive relation 'Large-for*', different in being satisfiable by objects and sets of which those objects are not members, the inference will be formally valid in virtue of the

⁵ This latter is a suggestion found in S. C. Wheeler, 'Attributives and their Modifiers', *Nous* (1972), p. 310.

presence of the logical constant '&'. We may well wonder what is to stop someone adding ' $\dots \& p$ ' on to whatever translation his colleague provided for ' X knows that p ', claiming to show thereby that the detachment inference with 'know' is formally valid.

However, I think a deeper and more sympathetic reading of Davidson's ideas would reveal that at least *this* charge of arbitrariness is unfair. Davidson writes: ' \dots we must uncover enough structure to make it possible to state, for an arbitrary sentence, how its meaning depends upon its structure \dots '⁶ and the suggested continuation is ' \dots and no more'. In 'On Saying That' he writes: 'For the purposes of the present paper, however, we can cleave to the most austere interpretation of logical consequence and logical form, those that are forced upon us when we give a theory of truth.'⁷ We can discern the following plan. In order to provide for the sentence 'John is a white protestant American' a suitable sentence of, say, a first-order fragment with only a finite number of semantical primitives, we *have* to uncover a conjunctive structure. But the addition ' $\dots \& p$ ' adds nothing to the enterprise of providing a suitable translation for ' X knows that p ', nor, incidentally, does the addition of the conjunct ' $\dots \& \text{John} \in \{y: \text{Man}(y)\}$ '. The only way in which we can incorporate ' $\text{John stopped } \phi\text{-ing at } t$ ' into a fragment with a finite number of semantical primitives (given that it is to contain no expression syntactically parallel to 'stopped') is to map it on to a sentence which contains ϕ occurring in its basic predicative role. Perhaps the only way to do that is to uncover the structure ' $\text{John } \phi\text{-ed up to } t$ and after t John did not ϕ '. If and only if this is the situation, we shall have shown that the inference from ' $\text{John stopped } \phi\text{-ing at } t$ ' to ' $\text{John } \phi\text{-ed before } t$ ' is a formally valid one.

This certainly provides us with a far from arbitrary account of what counts as revealing form and what counts as philosophical analysis. But there are still some grounds for

⁶ D. Davidson, 'The Logical Form of Action Sentences', in N. Rescher (ed.), *The Logic of Decision and Action* (Pittsburgh, Pa.: University of Pittsburgh Press, 1967), p. 82.

⁷ D. Davidson, 'On Saying That', in D. Davidson and H. Hintikka (eds.), *Words and Objections* (Dordrecht: Reidel, 1969), pp. 160-1.

reservation as to whether we have provided a finally satisfactory foundation for that distinction with whose allure we began.

First: it is not wholly nugatory to point to the fact that upon this conception the inference from 'John is a large man' to 'John is a man' will almost certainly not count as formally valid.

The second ground I have already mentioned. It is clear that what inferences come out as formally valid will vary radically, depending upon which language is used as the canonical fragment, and this imports an arbitrariness we might feel is misplaced in a wholly explanatory account of the matter. Briefly, we miss the transcendence.

Finally, and more seriously: it appears that the constraint upon how much structure to uncover will not, without considerable supplementation, achieve the intended results.

Consider the two radically different kinds of attributive adjectives exemplified by 'large' and 'good'. 'Large'-type attributives sustain all the inferences 'good'-type attributives sustain, although the former in addition sustain coextensive predicate substitution. Consequently, it is difficult to see why the general type of translation provided for 'good' will not also serve for 'large'. Suppose, merely for illustration, that we discerned a reference to the attribute 'being a man' in the sentence 'X is a good man', thus

Good-in ($X, \lambda y [\text{Man}(y)]$).

This would certainly also be a type of structure that would work for 'large'. From the point of view of merely being able to generate the T-sentences, the incorporation of reference to a set in the translation provided for the latter attributive was as gratuitous as the incorporation of the additional conjunct.

This difficulty is quite general. Provided we have a scheme for the suitable translation of what are inferentially the weakest members of a certain grammatical category, we seem to be enjoined to run that scheme quite generally, treating the inferences which the inferentially stronger members sustain as being consequences of their analysis. It is obvious that the maxim 'Translate so as to maximize formally valid inferences'

simply sends us back to wreck upon the shore of the superfluous conjuncts.

It would be foolish to hold that no more refined criterion could be forthcoming, but I find it difficult to believe that it can be drafted without drawing upon the slightly richer conception of semantics which I shall develop in the next section. Before proceeding to that, I should like to draw the comments I have been making together by putting forward a more general observation.

Although Davidson's main conception of formally valid inference drew its inspiration from wholly novel considerations, it rests squarely within that tradition which distinguishes formally valid inferences as those which depend upon the presence of certain favoured words, the logical constants. The novelty lay in two points. First: in the idea that those words would receive a special treatment in any semantic theory through being recursive elements. Second, and, of course, connected, was the stress placed upon a profoundly important *necessary* condition for the inference from $\Sigma(e)$ to $\Sigma'(e)$ to be formally valid. This was the condition that e should be shown to occur as a semantical unit in both $\Sigma(e)$ and $\Sigma'(e)$. Davidson was at pains to point out that, although a syntactic transformation can easily be made to be sensitive to the occurrences of e in $\Sigma(e)$ and $\Sigma'(e)$, nevertheless unless e can be shown to occur there semantically, by some theory of truth, there will be no hope of showing that the transformation is valid, i.e. generally truth-preserving. For example, there is no difficulty whatever in writing a syntactic rule of inference which captures the inference pattern exemplified by:

X is a tall man
 Y is taller than X
 Y is a man

Therefore Y is a tall man

But in the absence of a semantic theory which discerns 'tall' in 'taller than' (presumably by the construction of a recursive clause for the element '. . . er than'), we will have no hope

of proving that each of the infinite number of inferences licensed by the syntactic rule is valid.

But, important though this second idea is, it does only provide us with a necessary condition. For even when a theory of truth has shown e to occur in $\Sigma(e)$, there is the additional question: 'Is the mode of containment $\Sigma(\)$ by itself of a kind that warrants certain inferences, or are they rather due to certain special features of the particular constituents exemplifying that mode of containment?'

In addressing ourselves to that question, we have been following Davidson, and taking the way a theory of truth validates the inference from 'P and Q' to P as the model of the way in which a semantic theory shows an inference to be structurally valid. But following that model with 'large' provides us with the very reverse of that explanation of which our original intuition seemed to encourage the expectation. The possibility that suggests itself is that this model is wrong. Perhaps there is a notion of structurally valid inference, for which we have been groping, which does not comprehend inferences which rely upon the logical constants, and which thus cannot be married with the traditional notion of logical validity without disappearing. If there is such a different conception, whatever its merits and whether or not anyone else has been groping for it, we can only benefit by having it distinguished from its more traditional rival. It is this that I attempt in the next section.

II

Let us return to the intuition about structurally valid inferences with which we began. Surely the natural way of defending the claim that the inference from 'John knows that p ' to p is not structurally valid would not mention the logical constants at all, but would rather run as follows: the inference cannot be a matter of structure, since the sentence 'John believes that p ' has precisely the same semantic structure as, is composed in exactly the same way out of the same types of semantic elements as, the sentence 'John knows that p ', and yet it does not sustain the inference.

We have here the idea of a contrast between inferences whose

validity depends merely upon the *kind* of semantic elements out of which a sentence is constructed, and its manner of construction, on the one hand, and inferences whose validity depends upon the special variation of a particular semantic element is playing upon the theme all expressions of its kind must play, on the other. Can we make anything of this idea?

The central task, if we are to make anything of it, will be to provide a way of telling when two expressions are of the same semantic category, and our hopes of doing this might appear rather dim. For we seem to be confronted with the following dilemma. If we are allowed to consider all the valid inferences involving some expression in determining the kind of semantic contribution it makes, there is nothing to stop us inventing a category of *factive attitudinatives*, unified in being like 'know' and sustaining the detachment inference. If we are not allowed to take all inferences into account, we will have to decide upon a partition amongst inferences as a preliminary to an enterprise whose object was the construction of just such a partition.

What makes a parallel problem in the taxonomy of natural kinds tractable is the conception of kinds as being differentiated by underlying structures from which the characteristics we use in classification may be regarded as flowing. To solve our problem for semantic kinds we need to find room for a parallel conception of something from which an expression's inferential properties may be regarded as flowing.

Just such a conception is provided for by what I shall call an *interpretational semantics*. A semantic theory of that type specifies, for each kind of semantic expression, an entity—a set, a truth value, a function from sets to truth values, or whatever—which may appropriately be assigned to members of that kind upon an arbitrary interpretation of the language. We can regard the specification of the kind of assignment as a specification of the underlying real essence which a word has in common with many other words, and of which the validity of certain inferences involving it is a consequence. These will be the structurally valid inferences; inferences which are truth-preserving no matter how we permute assignments within the limits laid down as appropriate for members of that category.

Thus, to take a central but simple case, we justify the intuition that there is a unitary semantic category of n -place predicates, and provide a general characterization of their role, by stating that upon any admissible interpretation of the language each n -place predicate must be assigned a set of n -tuples of the domain, and by providing an adequate definition of truth upon an arbitrary interpretation using assignments of that type.

A slightly less obvious case is the following. We can justify the belief that there is a semantic category to which 'large', 'tall', 'expensive', and 'heavy' all belong (the extensional attributive adjectives), by showing that there is an assignment which may suitably be made to each (a function from sets to subsets of those sets, possibly satisfying certain additional conditions), and which provides the basis for an adequate definition of truth upon an interpretation. It is again clear that certain inferential features will flow from this underlying real essence. We see why the detachment inference is valid, why substitution of coextensive predicates within the scope of such adjectives is valid, and why commuting the adjectives will not in general be valid.

The metaphor of underlying essence suggests that we are attempting to *explain* why the expression has certain inferential properties, but is there any more justification in speaking of explanation here than there was when validation of the inference followed merely from the presence of the parallel inference in the metalanguage? Are we not merely restating, in set-theoretical terms, the various inferences we are concerned to explain?

I think this charge can be resisted. A certain, syntactically identified, pattern of inferential behaviour is a feature which can and does flow from many different underlying constitutions. For example, both standard adverbs—like 'breathlessly'—and intensional attributives—like 'good' in 'good (as a) king'—sustain a parallel detachment inference; yet it is clear that the 'restatement' we are supposed to have provided is inappropriate for both kinds of expressions. Neither can be seen as involving functions from sets to subsets of those sets. Davidson can be seen as providing a quite different explanation for the detachment inference in the case of

adverbs (making assignments of sets of events to adverbs, and sets of $n + 1$ -tuples, of events and n -tuples of objects, to verbs). No one yet knows how to provide an adequate explanation of the detachment inference for 'good'.

The requirement that one provide an account of an expression's contribution by specifying an appropriate kind of assignment is not an easy one to meet, and certainly puts a quite different complexion upon the taxonomic problem. Some groupings of expressions upon the basis of shared inferential behaviour will be simply impossible, for no coherent kind of assignment to the members of the heterogeneous class which results will be available. While we are merely concerned with grouping expressions together in a way which imposes the best organization upon their inferential behaviour, we are bound to feel the force of a certain arbitrariness since it appears certain that an equally good scheme which organizes the data in another way can always be put together. Locating the taxonomic problem in the context of an interpretational semantics puts an end to this free-wheeling. What we expect, then, is the provision of the most determinate and yet economical statement of the kind of semantic contribution made by any expression of a given type, thereby making structurally valid as many inferences as possible. One hopes that this will not involve one restriction upon the assignment for every inference validated: we aim at the sort of illumination that can come from an economical axiomatization of the behaviour of groups of expressions. Then we can say: '*This* is the kind of expression *e* is, and that is why these inferences are valid.'

But this may simply be to raise another, deeper problem. I have spoken of making structurally valid as many inferences as possible. Does this not raise the old difficulty? For will we not be obliged to subdivide the class of attitudinatives, and discover a category of factive attitudinatives, the assignment to which differs from the assignment to their non-factive brothers solely in requiring the contained sentence to be true?

Now that we are working within the context of an interpretational semantics, I think we can quiet this worry. We should regard our construction of, and assignments to,

categories in the following spirit: if two expressions behave in the same way but are in different categories, this is a lost generalization. If we regard the enterprise in this light, we will construct a new category out of an older and more comprehensive category only when we can make an assignment to members of the new category which provides a *different* explanation for the behaviour which members of the new category had in common with the old, the provision of which explanation would show that the apparent unity in the behaviour of members of the old category was deceptive, concealing deep differences of functioning. Only in this case will the discovery of a new category not lose us a significant generalization. By imposing this requirement we make the notion of structural validity transcendent.

Thus, to recognize a category of factive attitudinatives would require the provision of a different explanation, for its members, of the inferential properties they share with words like 'believe'; an explanation which would unify these properties with the detachment inference. It is by no means out of the question that just such an explanation ought to be provided. Zeno Vendler has provided arguments which cast doubt upon the apparent similarity of 'knows' and 'believes',⁸ while the causal theory of knowledge provides us with a rationale for treating the contained sentence as designating a state of affairs with which the knower is stated to have come into epistemological contact. If we did so treat the contained sentence, this would be to treat 'know' like 'enjoy' in one of its uses, as introducing a function from states of affairs and persons to truth values. This is to provide an explanation of the detachment inference. And provided the identity conditions of states of affairs are sufficiently fine-grained, the explanation unifies that inference with the expression's other inferential properties, for example, its opacity.

If we are to preserve the promise in the notion of semantic structure, we have to steer between two courses. We must resist saying, with Montague, that no detachment inference with a predicate modifier is structurally valid, on the ground

⁸ 'On What One Knows', in Z. Vendler, *Res Cogitans* (Ithaca, NY: Cornell University Press, 1972), p. 89.

that the expression 'fake' must be regarded as belonging in that category.⁹ But equally we must not allow the freedom to subdivide and form new categories to lead to that proliferation of categories which threatens our interest in the notion. We can steer this middle course only if we require that an appropriate assignment be provided for each category, and illuminatingly different assignments be provided for different categories.

The construction of an interpretational semantics for a natural language will doubtless follow a rather different course from that taken in the construction of such theories for artificial languages, not merely in the heterogeneity of entities assigned. In the first place, there is no reason for the different interpretations to have different domains. Secondly, instead of a single unsorted domain, it will be convenient to have a domain divided into fundamental sorts of objects: places, times, material objects, animate objects, events . . . (This would enable us to describe the admissible assignment to an action verb, e.g., as 'a set of pairs of animate objects and times'.) Fundamental arguments can be expected on the question of how many, and which, sorts of objects we need, and on which we ought to take as primitive and which we can define. Is the sort of object, *Events*, fundamental or can it be defined? Do we need *Facts*, or *States of affairs*, and if so can they be defined?

Arguments can also be expected upon how we are to extend the programme to intensional areas of a natural language—how we are to represent meaning-sensitive functions. Perhaps by associating with each semantical primitive an intension and recursively defining intensions for complex expressions. But perhaps not.

All these are interesting, indeed fundamental, questions, but from the lofty viewpoint I have adopted in this paper, matters of detail. What is important is that we should have a general picture of the enterprise, and that it should be recognizable as a representation of at least part of what philosophers have been aiming at in their exploration of the semantic structure of sentences of natural languages. They

⁹ R. Montague, 'English as a Formal Language', in *Linguaggi nella Società e nella Tecnica*, pp. 212-14.

are seen to have been asking questions like: 'What kind of element is this? Does it introduce a class?' (cf. Davidson's remark 'Intentional actions are not a class of actions'),¹⁰ 'Can vague predicates introduce classes?', 'If what this expression introduces is a function, what are its inputs and outputs?'

It seemed that questions of this kind were not necessarily answered by the construction of a truth theory (except in a mildly perverse way, when the background language is taken to be first-order). More interestingly, the answers have little to do with the logical constants.

How do standard inferences involving the logical constants fare? In any interpretational semantics for English the propositional connectives, 'not', 'and', 'or', and so on, would surely fall into a common semantic category, to which, on the classical conception, there would be assigned functions from truth-values to truth-values. In the definition of satisfaction upon an arbitrary interpretation, I , we would have a clause for any sentence involving any n -ary connective θ (possibly joining open sentences), and for any finite sequence s :

$$\text{Sats}_I(s, \theta(S_1 \dots S_n)) \leftrightarrow \text{Assig}_I(\theta) [\langle \text{Val}_I(s, S_1) \dots \text{Val}_I(s, S_n) \rangle] = T$$

where

$$\text{Val}_I(s, S_1) = T \leftrightarrow \text{Sats}_I(s, S_1).$$

Consequently, with the exception of inferences involving substitution of sentences with the same truth value, none of the standard inferences involving the sentential connectives is structurally valid. Briefly, the sentences ' P and Q ' and ' P or Q ' have the same semantic structure; the former's entailing P is due to the special variation the word 'and' plays upon a theme it has in common with 'or'.¹¹

In view of this, it is not surprising that we found the manoeuvrings necessary for validating the inference from ' P and Q ' to P an unilluminating model, and it is clear why, when we followed it with 'large', the distinction we were

¹⁰ 'The Logical Form of Action Sentences', p. 94.

¹¹ 'One might surely have expected that if any pair of non-synonymous expressions exhibit non-difference of type or category, "or" and "and" would be one such pair and "all" and "some" another.' P. F. Strawson, 'Categories', in O. P. Wood and G. Pitcher (eds), *Ryle* (London: Macmillan, 1970), p. 184.

trying to capture between structurally valid inferences and those requiring proper axioms petered out into nothing. To put the point succinctly, there is no deep explanation of why ' P and Q ' entails P .¹²

Quantifiers are more complicated but they too can be seen as falling into a single semantic category. There is work by Lindstrom and Mostowski on generalized quantifiers in which the characterization of the category is attempted.¹³

It is now time to deal with something I have left unclear up to this point. I spoke of the structural inferences as being those whose validity was due to the types of semantic element involved and the significance of their construction. What is meant by the semantic significance of a grammatical construction?

An interpretational semantics must not only provide appropriate assignments to expressions; it must also provide an adequate definition of truth upon an arbitrary interpretation making use of those assignments.¹⁴ Each clause in that definition can be regarded as making explicit the significance of one kind of grammatical construction, and certain inferential consequences will follow from the significance thus characterized.

To understand this idea consider the following simple example. Suppose we have a language like English in that attributive adjectives can be stacked, yet unlike English in that from 'John is a strong tall man' it follows both that John is a strong man and also that John is a tall man, and the former sentence is entailed by these taken together. One way of capturing these facts is by making (to such attributive adjectives) the assignments we have earlier suggested (functions from sets to subsets), and writing the following clause in the definition of satisfaction upon an interpretation:

Where $M, M' \dots$ range over object-language attributive

¹² It is worth observing that intuitionistic sentential connectives cannot be regarded as representing truth functions in any finite many-valued logic. See K. Gödel, 'Zum intuitionistischen Aussagenkalkül', *Ergebnisse eines mathematischen Kolloquiums* 4 (1933), pp. 34-8.

¹³ Per Lindstrom, 'First Order Predicate Logic with Generalized Quantifiers', *Theoria* 32 (1966); A. Mostowski, 'On a Generalization of Quantifiers', *Fundamenta Mathematicae* 64 (1957).

¹⁴ What constitutes adequacy will be considered later.

modifiers, ϕ over object-language monadic (possibly complex) predicates, and 'Var_{*i*}' designates the *i*th object-language variable, for any infinite sequence of elements *s*,

$$\text{Sats}_I(s, M(M'(\phi)) \widehat{\text{Var}}_n) \leftrightarrow \text{Sats}_I(s, M(\phi) \widehat{\text{Var}}_n) \text{ and } \text{Sats}_I(s, M'(\phi) \widehat{\text{Var}}_n).$$

A consequence of this clause will be that upon the designated interpretation,¹⁵ 'John is a large strong man' is true iff John is a large man and John is a strong man.

If what has gone before is correct, it would be a mistake to try to render explicit the semantical potential lodged in the construction of stacking modifiers by saying that our original sentence has the same semantic structure as some sentence containing the connective 'and'. For such a sentence achieves what is admittedly the same net effect by means of different semantic elements and different constructions.

Just as one can never dispense with rules of inferences by enriching one's axioms (though a limited trade-off is possible) so, no matter how interlocking the assignments to interacting expressions are made, one can never obviate the need for *some* semantic potential to be lodged in the grammatical constructions—even if it is merely that concatenation signifies functional application.¹⁶

The need for the distinction between inferences whose validity is due to the semantic significance of grammatical constructions and those which depend upon the logical words would certainly evaporate if we could see our way to accepting Quine's suggestion¹⁷ that the logical constants should be regarded as 'absorbed into the constructions' they signalize. But, as Quine recognizes, we lack a clear reason for distinguishing them in this way.

It is not part of the thesis of this paper that no satisfactorily transcendent account of the logical constants can be produced, nor, consequently, that there is no well-defined notion of logical form, logical consequence, and logical

¹⁵ For an explanation of the notion of the designated interpretation see below.

¹⁶ As it does in the system put forward in D. K. Lewis, 'General Semantics', in D. Davidson and G. Harman (eds), *Semantics of Natural Languages* (Dordrecht: Reidel, 1972), p. 169.

¹⁷ W. V. Quine, *Philosophy of Logic* (Englewood Cliffs: Prentice-Hall, 1970), pp. 28–30.

validity. It is merely that there is another, deeper notion, of semantic structure and of structural validity, from which those notions should be distinguished.

Does it follow from what has gone before that we are only interested in relative truth—truth upon an interpretation (truth_{*I*})—and not simple truth? Certainly not: no theory of meaning which fails to state the actual truth conditions of sentences of a language is worthy of the name. But since an interpretational semantics will comprise a definition of 'true_{*I*}' there is a very natural way of modulating to truth by specifying, for each semantical primitive of the language, what assignment it is to receive on the designated interpretation, *I*^{*}, and thus enabling us to derive for each sentence of the language its truth_{*I*^{*}} (= truth) conditions.¹⁸ In other words, we require not merely that we are told what *kind* of function 'large' is, but *which* function of that kind it is. Thus we will expect the specification of *I*^{*} to include such clauses as

$$\text{Assig}_{I^*}(\text{'large'}) = \{ \langle \alpha, \beta \rangle : a \in \beta \leftrightarrow \text{large } \hat{y} [y \in \alpha] a \};$$

i.e. on the designated interpretation 'large' is to be assigned that function from sets to sets such that an object is a member of the output set iff it is a large member of the input set.

When the whole weight of the notion of structurally valid inference was being placed upon them, we found clauses like these—clauses in which a term is used to state its own semantic contribution—threateningly trivial. But now that the structurally valid inferences are fixed without their help, they can play their part, for it is an essential one.

Such a modulation to absolute truth is not merely desirable, it is necessary. For it is only by testing consequences of this kind generated by an interpretational semantic theory that we can impose any criterion of adequacy upon such theories. The truth_{*I*} conditions of a sentence will necessarily be stated in rich semantical and set-theoretical notation, and in that condition they are immune from empirical control. We must bring an interpretational semantics down to earth by insisting that it should meet just the constraints that Convention T imposes; that it generate for each sentence

¹⁸ See Montague, *op. cit.*, p. 211.

of the language under consideration a true sentence of the form

$\lceil S \text{ is true}_{I^*} \text{ iff } p \rceil$

where p is a translation of the sentence S names.¹⁹ This means *at least* that p should be no richer in conceptual content or ontology, and thus should contain no semantic terms not obviously present in the original.

To summarize, then, an *interpretational semantics for a language L* is a theory consisting of three main parts:

- (i) a definition of an *admissible interpretation of L* , which proceeds, in part, by stating, for each grammatical category of semantical primitives of L , what constitutes an appropriate assignment to a member of that category;
- (ii) a definition of *truth upon an arbitrary interpretation of L* ;
- (iii) a definition of the *intended or designated interpretation of L* , which identifies, for each semantical primitive of L , which of all the appropriate assignments it is to receive on the designated interpretation.

It is important to observe that we can now allow the deductive apparatus of the metalanguage in which we state the interpretational semantics to comprise any *complete* proof theory for expressions of that language, without the danger of trivializing the notion of structurally valid inferences.

Now let us say that a phrase-structure grammar, G , is adequate for an interpretational semantics of L iff it generates a structural description of every sentence of L such that:

- (a) There is an effective procedure for determining, upon the basis of this description of any sentence S , which clauses

¹⁹ Given what I have said about clauses in the definition of truth_I stating, with the use of logical constants, the significance of grammatical construction not containing such expressions, I cannot require homophonic truth_I -biconditionals. How important a difference this is between myself and Davidson depends upon the view one takes of the significance of requiring a homophonic truth theory merely for the regimented fragment, and allowing sentences of the natural language to be mapped on to sentences of the fragment which are very different.

of the definition of truth_I are to be used in deriving the truth_I conditions of S , and in what order. (This is needed to deal with scope.)

(b) Where a *penultimate non-terminal node* is a node directly dominating only a terminal node (i.e. a lexical item of L): at each *penultimate non-terminal node* of each tree generated by G is a symbol for a category an appropriate assignment to the members of which is given by the definition of an admissible interpretation of L .

Then we may say: two sentences have the same semantical structure relative to an interpretational semantics iff the grammar adequate to that semantics assigns them structural descriptions which agree up to all non-terminal nodes.

The effect of this definition is that two sentences have the same semantical structure iff they are composed out of elements of the same semantic categories in the same way—where ‘composed in the same way’ means that they will trace exactly parallel courses through the definition of truth_I in the generation of their truth_I conditions.

And finally let us say that an inference from $S_1 \dots S_{n-1}$ to S_n is structurally valid iff S_n is true upon every admissible interpretation on which $S_1 \dots S_n$ are true; or more strictly, for every interpretation, every sequence which satisfies $S_1 \dots S_{n-1}$ upon that interpretation also satisfies S_n upon that interpretation.

III

A logically perfect language would have one-to-one correspondence between its semantic and syntactic categories. I see no reason to suppose that natural languages are logically perfect, at any level.²⁰ There can be a breakdown of the one-to-one correspondence in either direction. We may find it necessary to subdivide a syntactically unitary category, so that, for example, even if ‘It is certain that . . .’ and ‘It is not the case that . . .’ were everywhere substitutable *salva congruitate*, they could be members of different semantic categories. And equally we may find it convenient to make assignments of the same kind to expressions of different

²⁰ Here, it seems to me, I depart from D. K. Lewis, op. cit.

syntactic categories. Thus we might find it convenient to assign to an adverb and also an adjective a set of events (indeed, if they stand in some morphological relationship, we may even require that they be assigned the same set of events). We do not have to gerrymander the grammar to get the two expressions into the same syntactic position; we have room for the idea of one and the same semantic element's being realized by expressions of different syntactic categories, no one of which is basic in a transformational sense.

In addition, logical constants will appear only when they appear to appear. We leave the clauses in the theory of truthy to capture the semantic significance of the constructions.

These features taken together mean that the structures we regard as exhibiting semantic structure can be relatively close to object-language syntactic structures; close, that is, relative to some recent proposals concerning semantic structure. A good example is provided by Davidson's proposal that the semantic structure of 'John ran breathlessly' is that of 'There is an event e such that John ran e and e was breathless'.

Now there is no doubt that proposals of this kind have tended to mystify people. I quote James Cargile as a representative example:

First: we might think that this sentence ['Shem kicked Shaun'] consisted of two names and a two-place relation . . . But this is wrong! Wrong? Yes, wrong! Second: the sentence really is of a three-place relational form, with two names and an existential quantifier. An existential quantifier? Where is it?²¹

It is not difficult to make sense of proposals such as Davidson's from the standpoint of the theory we have been developing, if we regard the sentence which is claimed to exhibit the semantic structure of a natural-language sentence as doing two related things. First, it indicates to what semantic category an expression is to be regarded as belonging, by representing it by an expression in the new sentence which belongs to a grammatical category of a canonical syntax, the admissible assignments to which are taken to be well

²¹ J. Cargile, 'Davidson's Notion of Logical Form', *Inquiry* 13 (1970), pp. 137-8.

understood. Secondly, it makes explicit with logical constants the semantic significance of the grammatical constructions employed in the original sentence.

Thus, instead of saying that 'breathlessly' should be assigned a set of events, we say that it is *really* a predicate, and instead of saying that to 'runs' should be assigned a set of pairs of agents and events, we say that *really* it is a two-place predicate.

This is a perfectly natural way to proceed; indeed, there has been a long tradition of representing intuitions about semantic functioning as intuitions about syntactic position (in some, admittedly rather mysterious, level of syntax). For example, philosophers have said that 'John feels a pain' has the form 'John feels painfully' (the 'adverbial analysis'), that 'carelessly' is a sentential adverb, that 'before' is not a sentential connective, that 'exists' is not a predicate, that definite descriptions are not terms, and so on.²² It was natural to present semantic proposals in this way since it represented, before the development of formal semantic theories, one way of registering one's intuition that two expressions did or did not belong to the same semantic category.

Thus I think we can understand proposals like Davidson's as involving much more than claims to mere logical equivalence or regimentation without being mystified by them. The theorist is representing in a favoured, logically perfect, notation the types of semantic element which figure in the original sentence, and representing as explicitly as possible the significance of the construction which it exemplifies.

However, there is no need to make semantic proposals in this way; and if one is going to go for it one must be sure that one has a canonical syntax rich enough to accommodate all the types of semantic element one wishes to discover, and that one has an interpretational semantics for the language specified by the canonical syntax. It would seem preferable to short-circuit the canonical language, constructing an

²² See, e.g., G. Ryle's notion of 'adverbial verbs' (including 'hurry' and 'think') in 'Thinking and Reflecting', *Collected Papers*, Vol. 2 (London: Hutchinson, 1971), p. 467.

interpretational semantics for the natural language directly, where this is possible.

I said earlier that translation into a first-order language did force one to confront the right questions, but led one to answer them in a mildly perverse way. I want to end by explaining why I think that this is so.

We have put into the centre of the picture Frege's idea that some expressions can be regarded as introducing functions. This means that we are obliged to ask, concerning the function introduced by some expression like 'fake' or 'carelessly', 'What are its inputs and outputs?' Now if we translate such expressions into first-order languages, we will probably represent them as relational expressions whose terms designate objects of the type determined to be suitable as inputs. Thus instead of introducing a function from sets, 'large' becomes the relational expression 'large-for' holding between individuals and sets. In a way we register the extensionality of the function, but it is a different way.

Why is it perverse? The perversity lies in the fact that we attribute to the speakers of the language the ontological commitments which properly belong to the *theorist*. We are in fact no more justified in holding that the speakers' ontology encompasses sets, on the basis of the existence of expressions of theirs which introduce functions on sets, than we are in supposing that they require an ontology of truth values in order to have expressions for the truth functions. Delicacy on this issue will pay off when we come to consider intensional areas of language.

APPENDIX

Fragment of a homophonic truth theory for a regimented language containing extensional attributive adjectives as predicate modifiers.

The syntax is as for standard first-order theories save, in addition: where A is any one-place open sentence, $\hat{x}_i[A]$ is a one-place predicate abstract. A complex predicate of degree one is any one-place predicate abstract possibly preceded by one or more modifiers $M, M' \dots$. If ϕ is a complex predicate of degree one, ϕx_j is a wff.

We let a, b, \dots be metalinguistic variables over objects, and s, s', \dots metalinguistic variables over infinite sequences, and we write ' $s \approx_i s'$ ' for ' s differs from s' in at most the i th place'. We presuppose the standard two-place function $*$ from infinite sequences and object-language variables to objects such that $s^*(x_i) =$ the i th member of s .

The recursive definition of the two-place relation of satisfaction ('Sats') holding between sequences and object-language expressions is enriched by the following clauses:

$$\text{Sats}(s, \phi x_i) \leftrightarrow \text{SAT}(s^*(x_i), s, \phi).$$

$$\text{SAT}(a, s, \text{'large'} \widehat{\phi}) \leftrightarrow \text{large } \hat{b}[\text{SAT}(b, s \phi)]a.$$

$$\text{SAT}(a, s, \hat{x}_j[A]) \leftrightarrow (s')[(s' \approx_j s \ \& \ s'^*(x_j) = a) \leftrightarrow \text{Sats}(s', A)].$$

Note

We are obliged to define satisfaction for modified wffs by way of the three-place relation SAT for the following reason. If we attempted the most direct manoeuvre

$$\text{Sats}(s, \text{'large'} \widehat{\phi x_j}) \leftrightarrow \text{large } \hat{b}[\text{Sats}(b, \phi)]s^*(x_j)$$

then the predicate abstract $\hat{b}[\text{Sats}(b, \phi)]$ would have to be true of individuals, not sequences. But in general ϕ cannot be assumed to be replaced by an expression which an individual can be regarded as satisfying directly. For we need to keep relativity to the sequence s to deal with any free variables which the predicate abstract replacing ϕ may contain.

Pronouns, Quantifiers, and Relative Clauses (I)

I Introduction

Some philosophers, notably Professors Quine and Geach, have stressed the analogies they see between pronouns of the vernacular and the bound variables of quantification theory. Geach, indeed, once maintained that 'for a philosophical theory of reference, then, it is all one whether we consider bound variables or pronouns of the vernacular'.¹ This slightly overstates Geach's position since he recognizes that some pronouns of ordinary language do function differently from bound variables; he calls such pronouns 'pronouns of laziness'. Geach's characterization of pronouns of laziness has varied from time to time, but the general idea should be clear from a paradigm example:

- (1) A man who sometimes beats his wife has more sense than one who always gives in to her.

The pronouns 'one' and 'her' go proxy for a noun phrase (here: 'a man' and 'his wife') in the sense that the pronoun is replaceable in paraphrase by simple repetition of its antecedent.²

From *The Canadian Journal of Philosophy* 7, 3 (Sept. 1977), pp. 467-536. Reprinted by permission. I would like to thank the following for reading the paper and offering comments and encouragement: M. K. Davies, G. H. Harman, P. F. Strawson, B. Taylor, D. Wiggins, D. Wilson and M. J. Woods. [The minor changes Evans made to the original version of this paper when it was reprinted in *Reference, Truth and Reality: Essays on the Philosophy of Language*, Mark Platts (ed.) (London: Routledge & Kegan Paul, 1980) have been incorporated here. Ed.]

¹ P. T. Geach, *Reference and Generality* (Ithaca, NY: Cornell University Press, 1962), p. 112.

² For a record of the change in definition see first P. T. Geach, op. cit., pp. 124 ff., then 'Referring Expressions Again', in *Logic Matters* (Oxford: Blackwell, 1972), pp. 97-8, then 'Back-Reference', *Philosophia* 5 (1975), p. 194. The change turns out to be important.

However, if we leave such cases out of consideration for the time being, we are left with two main kinds of situation in which pronouns occur, and Geach appears to want to say that in both of them pronouns are functioning in the way bound variables of quantification theory function. In the first kind of situation, the pronoun has a singular term as its antecedent, as, for example, in

- (2) John loves his mother
(3) John is happy when he is in love

In the second kind of situation, the pronoun has a quantifier expression, or what Geach calls 'an applicational phrase', as its antecedent; as, for example, in

- (4) Some man loves his mother
(5) No man is happy when he is in love

From Geach's writings on pronouns it becomes clear in what the analogy between pronouns and bound variables is taken to consist. Time and again, in those writings, his target is the 'lazy assumption' that we can understand the functioning of pronouns by labelling them 'referring expressions' and inquiring into what they refer to. Just as it makes no sense to ask about the reference of any particular occurrence of the variable 'x' in the sentence

- ($\exists x$) (Fx & Gx),

Geach's idea is that it equally makes no sense to inquire into the reference of English pronouns. Correspondingly, just as there are many sentential contexts containing variables which cannot be regarded as having truth values, on Geach's view there will be many English sentences containing pronouns which cannot be regarded as complete sentences with a truth-value.

For at least some occurrences of pronouns in English, these parallels with bound variables appear quite striking. Geach is surely right that it does make no sense to inquire into the reference of the pronoun in

- (4) Some man loves his mother.

And, just as it makes no sense to ask for the truth value of the sentence 'Gx' in the formula

$$(x) (Fx \supset Gx),$$

surely Geach is right that we cannot assign a truth value to the sentence 'he admires Mozart' as it occurs in

If any man loves music he admires Mozart.

Despite this, I find myself in considerable disagreement with the *general* equation between pronouns and bound variables which Geach has put forward. My disagreement consists of two independent points which I can best summarize as follows.

First, take any sentence containing a pronoun which has a quantifier antecedent, and which is admittedly functioning like a bound variable, e.g. (4) and (5). Paired with any such sentence is a well-formed sentence in which a singular term stands in place of the quantifier expression, as (2) and (3) are respectively paired with (4) and (5). Now, Geach appears to want to insist that pronouns function like bound variables even when they have singular terms as antecedents, so that they too cannot be assigned a reference, nor can their most immediate sentential context always be assigned a truth-value. For example, Geach appears to want to regard the sentence:

If John loves music he admires Mozart

as the result of attaching a complex conditional predicate to the name 'John', so that the function of the pronoun 'he' cannot be said to be that of referring to John nor can the sentence be regarded as the conditional compound of two sentences each of which has a truth value.

I do not wish to claim that Geach's way of looking at pronouns with singular antecedents is incorrect, though I shall mention one or two advantages of looking at them another way. My first disagreement with Geach is on the question of whether it is *necessary* to adopt his way of looking at such pronouns. Geach believes that if we are to recognize the status of pronouns which have quantifier antecedents as akin to bound variables—as expressions whose function is

not to refer—then we must say the same about the pronouns in those singular sentences from which the quantified sentences may be regarded as got by substitution. This seems to me to be a mistake. I shall argue that the semantical significance of these pronoun-antecedent constructions can be exhaustively stated in terms of a simple principle according to which a pronoun refers to whatever its singular antecedent refers to. Such an account is entirely adequate, not merely in the sense that it explains the functioning of pronouns with singular antecedents, but in the sense that no further explanation of the functioning of pronouns with quantifier antecedents is called for.

The fact that there are two equally viable approaches to the semantics of these pronoun-antecedent constructions—Geach's approach and the co-referential approach—is of no great moment in itself. It is an entirely trivial consequence of the fact that there are two different approaches to the semantics of quantified sentences, which I will outline in section II below. As with other points that I shall make in this paper, this point is only worth making because Geach has denied it, many times and with great vehemence.

For the purpose of this first dispute, then, it is common ground that certain pronouns with quantifier antecedents function just like bound variables—the issue is rather how this effect is to be achieved. But my second, and much more important, disagreement with Geach arises over the question of just how many pronouns with quantifier antecedents can be seen as functioning in this agreed way. I want to try to show that there are pronouns with quantifier antecedents that function in a quite different way. Such pronouns typically stand in different grammatical relation to their antecedents, and, in contrast with bound pronouns, must be assigned a reference, so that their most immediate sentential contexts can always be assigned a truth value. The relevant grammatical relation appears to be Klima's relation of 'in construction with'.³ When the pronoun is in construction with its antecedent, as in (4) and (5), the result is a bound pronoun. But when it is not, as in

³ E. Klima, 'Negation in English', in J. Fodor and J. Katz (eds), *The Structure of Language* (Englewood Cliffs: Prentice-Hall, 1964), p. 297.

(6) Mary owns a donkey and John beats it

(7) John owns many sheep and Harry vaccinated them last July,

the pronouns must be regarded as having a reference, so that the second conjunct in both sentences may be assigned a truth-value. I call these pronouns 'E-type pronouns'.

E-type pronouns will occupy us for the bulk of this paper. But first I want to consider the other issue, since it will help us in thinking about pronouns that do not function like bound variables to have considered some that do. And before I can embark upon the question of the proper treatment of pronouns that do function like bound variables, I must say first something about two different approaches—the Tarskian and Fregean approaches—to the semantics of quantified sentences.

II Two approaches to the semantics of quantified sentences

The major problem posed for semantic theory by quantified sentences of both natural and artificial languages arises because of the curious dual role which connectives and quantifiers play. The connectives 'and', 'not', 'if . . . , then', etc., both have the role of forming complex sentences from sentences, and of forming complex predicates from predicates. Thus we have

It is not the case that snow is white
as well as

Some men are not bald,
and

Snow is white and grass is green
as well as

Some men are young and bald.

The quantifiers also have a dual role. Sometimes they form sentences from predicates, as 'Someone runs' is formed from 'runs', and sometimes they form predicates of degree $n-1$

from predicates of degree n , as 'loves someone' is formed from 'loves'.

This poses the following problem. Our first instinct would be to provide an account of the sentence-forming role of these expressions in terms of the *truth* conditions of the resulting expression, and of their predicate-forming role in terms of the *satisfaction* conditions of the resulting expression. So, for example, if we followed our first instinct with the expression 'and', we should have the principle

(A) Any sentence \widehat{S} 'and' $\widehat{S'}$ is true iff S is true and S' is true

to deal with the sentence-forming role, and the principle

(B) An object satisfies a predicate of the form \widehat{F} 'and' \widehat{G} iff it satisfies F and it also satisfies G

to deal with the predicate-forming role. But, by having two independent principles for the single semantical unit 'and', we deem it to be ambiguous, which our second instinct would be to say is absurd.

There appear to be two logically possible ways of solving this problem, and theories along both lines have been constructed. Either a theory takes the sentence-forming role as basic, keeps principles like (A), and somehow or other contrives to explain the predicate-forming role in terms of it. Or alternatively, the predicate-forming role is taken as basic, and the sentence-forming role is somehow derived from principles like (B). The first approach is associated with the name of Frege, and the second with that of Tarski.

Tarski was able to manage with principles of the form (B) alone, because he assimilated closed sentences to predicates, assigning to them, as well as to predicates, conditions under which objects satisfy them.⁴ The essentials of the trick can be seen if we suppose Tarski's formal language had also contained certain unstructured propositional constants P and Q, with, for example, the meaning of 'Snow is white' and 'Grass is green'. Then, instead of explaining their semantic significance in a natural way,

⁴ A. Tarski, 'The Concept of Truth in Formalized Languages', in *Logic, Semantics, Metamathematics* (Oxford: Clarendon Press, 1956).

P is true iff snow is white

Q is true iff grass is green,

but in a way which would not integrate with the clause for conjunction when the sentence 'P and Q' had to be dealt with, Tarski would have had clauses which effectively assigned them an extension:

An object satisfies P iff snow is white

An object satisfies Q iff grass is green.

Since these clauses are of the form ' $(x) (Fx \equiv R)$ ' we know immediately that either every object satisfies these sentences or no object satisfies these sentences according to whether or not they are true or false. So it will be open for us to define a true sentence as a (closed) sentence with the universal extension. With this definition of truth, it is easy to show that the truth-functional role of the connective 'and', for example, as forming truths when and only when flanked by truths, is a special case of its role of forming an expression which is satisfied by an object iff that object satisfies both of the expressions which flank it.

Perhaps the point comes out most clearly if we look at the matter model-theoretically. 'And' is assigned a function from pairs of sets to their intersection, 'not' is assigned a function from a set to its complement, and so on. It is clear that $f_{\text{and}}(a, b) = \text{the universal set}$ iff $a = b = \text{the universal set}$, and $f_{\text{not}}(a) = \text{the universal set}$ iff $a = \text{the empty set}$. So if we define $T = \text{the universal set}$, and $F = \text{the empty set}$, the truth-functional role drops out as a special case.

This will achieve the desired results, but only so long as it is arranged that true closed sentences are satisfied by every object and false closed sentences are satisfied by none. It is easy to arrange this case for unstructured propositional constants, but it remains to be seen how Tarski arranged it for structured and closed atomic sentences. Pretend for a moment that the language under consideration contains only monadic predicates, and that the only way of forming a closed atomic sentence is by combining a quantifier with a predicate. Then, the most natural way of

stating the semantic effect of the quantifiers would be in clauses which spoke of truth, along the lines of

(C) A sentence of the form 'Something' \widehat{A} is true iff something satisfies A.

To give closed sentences the properties Tarski requires, (C) must be replaced by a principle which states the impact of the quantifiers in terms of satisfaction:

(D) An object satisfies 'Something' \widehat{A} iff something satisfies A.

Once again, this has the form of ' $(x) (Fx \equiv R)$ ', and the effect that a closed sentence is satisfied by all objects iff it is true, and by no objects iff it is not true.

So long as we consider languages all of whose atomic predicates are monadic, the form of (D) could only be explained by a desire to assign closed predicates an extension, in order that the predicate-forming role of the connectives can be taken as basic. But once the language contains polyadic predicates, the quantifier also doubles as a predicate-former, so that a clause of the form of (C) is not adequate by itself, and a clause, like (D), dealing in terms of satisfaction has some independent advantages. In fact it has to be more complex than (D), speaking in terms of satisfaction by ordered n -tuples, or sequences, of objects, and comprising some device for keeping track of which position in a complex predicate goes with which other. But it will have the same effect as (D), in that, when the quantifier is initial and the sentence closed, the conditions under which a sequence of objects satisfies it have nothing to do with the particular properties of that sequence, so that either every sequence will satisfy it, or none will.

Frege's alternative strategy is less well known, and certainly less widely appreciated as a genuinely alternative solution to the problems posed by the dual status of the connectives and quantifiers.⁵ The Fregean strategy is to take

⁵ My attribution of this theory to Frege rests upon Dummett's. See M. A. E. Dummett, *Frege* (London: Duckworth, 1973), chap. 2 and pp. 516-17. I disagree with Dummett by holding that the Tarskian approach is not just a notational variant of Frege's.

Fregean treatment of quantifiers may be found, for example, in B. Mates,

the sentence-forming roles of the connectives as basic, with principles like (A) which deal in terms of *truth* as exhaustive statements of them. The theory is able to deal with sentences in which connectives operate upon predicates, or expressions which cannot be assigned a truth value, because in the course of evaluating such sentences, and by the time the contribution of the connective is to be accounted for, the sentence will have undergone a metamorphosis, as a result of which the constituents upon which the connective operates are, once again, complete sentences.

The main idea of a Fregean truth theory for quantified sentences is that 'in the case of a complex predicate, the notion of a predicate's being true or false of an object is derivative from that of the truth or falsity of the sentence which results from filling the argument-place of the predicate with a name of that object'.⁶ We may use a simple principle for the quantifiers like (C) but the relation of satisfaction which holds between an expression and an object to which that clause directs us is, in the case of a complex predicate, defined in terms of the truth value of the sentence which results when a singular term referring to that object is substituted in the predicate, or, if the language contains no name for the object, in terms of the truth value, in some extension of the language, of a sentence which results when a singular term which refers to that object upon that extension of the language is substituted in the predicate. We must assume that for every object there is an extension of the language which contains a name for that object, although at no stage are we obliged to assume that there is an extension of the language which contains a name for every object.

It should be clear how such a conception of satisfaction enables a theorist to dispense with any explanation of the role of connectives and quantifiers other than that stated in terms of truth. Thus, for example, 'Some man is such that he is young and he is bald' is true iff there is a man that satisfies the predicate 'he is young and he is bald'. Now, an

Elementary Logic (New York: Oxford University Press, 1965), p. 54; E. L. Keenan, 'Quantifier Structures in English', *Foundations of Language* 7 (1971), p. 262 and *passim*, and throughout Geach's writings.

⁶ Dummett, *op. cit.*, p. 405.

object y satisfies this predicate iff upon that extension of the language on which β denotes y , ' β is young and β is bald' is true.⁷ At this point we may invoke the simple principle (A) for sentential conjunction, and derive the result that y satisfies 'he is young and he is bald' iff upon that extension of the language on which β denotes y , ' β is young' is true and ' β is bald' is true. Since we know that a sentence of the form ' β is young' is true iff the denotation of β is young and a sentence of the form ' β is bald' is true iff the denotation of β is bald we derive the conclusion that y satisfies the complex predicate iff y is young and y is bald, so that the whole sentence is true iff there is an object that is young and bald.

There are similarities between Fregean truth theories and the substitutional truth theories familiar from the work of Professor Marcus.⁸ Both run their recursions directly on truth, and both take the sentence-forming roles of operators as basic. But there are crucial differences which I have tried to bring out by separating the clauses for the quantifiers (like (C)) and the clause giving a general explanation of the notion of satisfaction.⁹

Unlike the substitutional truth theory, the Fregean truth theory introduces no new concept of existence—the principle (C) uses the perfectly ordinary, objectual, concept of existence. And for every object that we deem to exist in this sense, we are obliged to consider as relevant to the truth value of quantified sentences, the truth value of a substitution instance that may be formed with the use of a term denoting that object, while at no point are we permitted

⁷ β is a name assumed not to occur already in the sentence. It is convenient to define the relation of extension holding between languages so that, as a limiting case, each language extends itself. We define truth not just for English but all the members of a family of languages which extend the stock of English singular terms.

⁸ See, e.g., R. B. Marcus, 'Interpreting Quantification', *Inquiry* 51 (1962), pp. 252-9. For an excellent discussion of substitutional quantification see S. Kripke's 'Is there a problem about Substitutional Quantification?' in G. Evans and J. H. McDowell (eds), *Truth and Meaning* (Oxford: Clarendon Press, 1976), pp. 325-419.

⁹ Now that this heuristic purpose has been discharged, I shall in later pages collapse the two principles into something along the more familiar lines of:

A sentence of the form ' $\text{Something}^{\wedge}A$ ' is true iff, upon some extension of the language, there is a substitution instance of the form $\beta^{\wedge}A$ which is true.

to consider as relevant the truth value of substitution instances formed with the use of non-denoting names.

The net effect of these two provisions is to deprive the Fregean truth theory of any ontological interest whatever. But, the fact that a Fregean truth theory is not an ontological rival to a Tarskian theory should not lead us to think that it is not a semantical rival. Using the leading idea of the Fregean theory, it is easy to construct homophonic theories for standard first-order languages which unquestionably take the sentence-forming roles of the quantifiers and connectives as basic. And the idea can be generalized to deal with quantification into any context. From the Fregean point of view, once an operator has been given a clear sense when attached to a closed sentence (containing singular terms), then no additional explanation is required for quantification into the context created by that operator.¹⁰

Although this is not the place to argue the matter, I do not think that the existence of the Fregean alternative is of merely technical interest. It is true that the Fregean theory with its direct recursion on truth is very much simpler and smoother than the Tarskian alternative, whose mechanism of infinite sequences differing in at most this or that place is dispensed with. But its interest does not stem from this, but rather from examination at a more philosophical level. It seems to me that serious exception can be taken to the Tarskian theory on the ground that it loses sight of, or takes no account of, the centrality of sentences (and of truth) in the theory of meaning; Tarski's assimilation of sentences to a certain kind of complex predicate is open to objection along just the same lines as the later Frege's assimilation of sentences to a certain kind of complex name.¹¹ Further, in the case of some expressions which double as both sentence and predicate formers, a direction of explanation which takes their

¹⁰ It is interesting to note that some of the delicacy of substitutional quantification into opaque contexts can be retained by Fregean quantifiers despite the ontological burden—that is to say, despite the fact that we are given licence to consider, for every object, a substitution instance involving reference to it. However, once we are dealing with opacities, we must interpret an object's satisfying a complex predicate $A(x)$ in terms of the truth of *some* (potential) singular sentence of the form $A(t)$ in which t refers to it, and not the truth of any such singular sentence.

¹¹ For cogent statements of this criticism, see Dummett, *op. cit.*, pp. 3-7, 194 ff.

sentence-forming role as basic seems to be the only possible one,¹² while in all cases it is more natural.¹³ This greater naturalness has a lot to do with the fact that the interpretation of these operators is empirically more determinate in their sentence-forming roles.

Important though these points are for the general theory of quantification, they are not of immediate importance for us.¹⁴ For us, all that matters is that the Fregean theory of meaning for quantified sentences should at least be deemed to be a coherent theory. Certainly Geach's opposition to the position on pronouns which I wish to defend as coherent does not rest upon doubt about this point, for he himself is prone, in his writings, to give truth conditions for quantified sentences along Fregean lines.¹⁵

Let us turn directly to the bearing these matters have upon the proper treatment of pronouns.

III Bound pronouns and pronouns with singular antecedents

What function do the pronouns have in the sentences

(2) John loves his mother

(3) John is happy when he is in love?

In addressing myself to this question, I am going to assume that there exists no compelling syntactic reason for supposing

¹² I particularly have in mind operators like 'It is certain that', 'John believes that'. I disagree with J. Wallace's paper 'Belief and Satisfaction', *Noûs* 6 (1972), p. 85, in which the converse, Tarskian, direction of explanation is defended.

¹³ This naturalness has certainly struck Quine; see the account in W. V. Quine, *The Roots of Reference* (La Salle, Ill.: Open Court, 1973), pp. 93-5, of the child's understanding the satisfaction by an object of a complex predicate in terms of the substitution of singular terms.

¹⁴ Those who are interested in the strengths and weaknesses of Fregean truth theory for quantifiers would benefit from reading T. Baldwin's paper, 'Quantification, Modality and Indirect Speech', in S. Blackburn (ed.), *Meaning, Reference and Modality* (Cambridge: Cambridge University Press, 1975). Baldwin dispenses with the idea of enlarging the singular terms of the language by exploiting the machinery independently needed for dealing with sentences containing demonstratives.

¹⁵ See, for example, the truth conditions for the quantifiers given throughout Geach's *Reference and Generality*, and also the discussion in 'Quantification Theory and Objects of Reference', *Logic Matters*, pp. 141 ff. It is true that it is not always possible to tell whether Geach has in mind purely substitutional or Fregean truth theories.

that sentences containing pronouns are derived transformationally from underlying structures in which there are repeated occurrences of singular terms. I do not make this assumption because I have been persuaded by the arguments against the existence of such a transformation, but because it seems to me that, if such were the origin of pronouns in singular sentences, the dispute between Geach and myself would have an easy resolution in my favour.

One pretty obvious answer to the question is this: the pronouns are singular terms referring to whatever their antecedents refer to. Let me try to expand this answer a bit.

Let us suppose that the base component of a grammar for English generates what I shall call *sentence frames* of the form

() loves ()

() is happy when () is in love.

In addition, we are permitted to form sentence frames from sentence frames by linking together two or more singular term positions in any sentence frame. Any device for this linking will do. We could use repeated occurrences of the same letter to link a number of singular term positions together, and then our sentence frames would look like this:

(x) is happy when (x) is in love.

Alternatively, we could use a brace notation suggested by Quine,¹⁶ and then they would look like this:

() is happy when () is in love.

We can call a series of singular term positions linked together in this way a *chain* of singular term positions, and for convenience we regard a single position as a 1-link chain. We can form sentences from sentence frames by the insertion of singular terms into singular term positions in such a way that there is one and only one term in every chain. Thus

¹⁶ W. V. O. Quine, *Mathematical Logic* (Cambridge, Mass.: Harvard University Press, 1965), p. 70.

(John) loves ()

(John) loves () and (Mary) loves ()

are sentences. Chained, but empty singular term positions will be realized in surface structure as pronouns of the appropriate number and gender.¹⁷

A semantic theory which issues the simple answer which I have just given to the question 'What are these pronouns doing here?' would be one which contained the following principle as an exhaustive account of the significance of the pronoun + brace device:

- (F) If σ is a sentence containing the singular term positions p_i and p_j which are chained together, and p_i contains the singular term τ and p_j contains the pronoun κ , then the denotation of κ in σ is the same as the denotation of τ .

By the application of such a principle, the truth conditions of sentences (2) and (3) would be given in the (slightly non-homophonic) theorems:

'John loves his mother' is true iff John loves John's mother

'John is happy when he is in love' is true iff John is happy when John is in love.¹⁸

Let us call the treatment which rests upon a principle like (F) the *co-referential treatment of pronouns*. I should say right away that I regard this as a proposal quite different from the proposal that pronouns with singular antecedents are 'pronouns of laziness'. When we say that an expression 'goes proxy for' some chunk of text, we evaluate the sentence containing it just as if the expression was replaced by what it goes proxy for.¹⁹ In many cases, the two treatments will

¹⁷ In allowing unrestricted forward and backward 'pronominalization' this simple grammar is quite unrealistic, but the tricky syntactical question of demarcating where 'pronominalization' is allowed and where obligatory really does not affect the semantical issues I am dealing with.

¹⁸ I consider the significance of the departure from homophony below.

¹⁹ I discuss the merits of the 'going proxy for' idea below.

have the same results, but they nevertheless invoke different mechanisms which produce divergent results in some examples.

Now, it has been widely thought that the co-referential treatment of pronouns is simply inadequate to deal with pronouns whose antecedents are not singular terms but quantifier expressions, like those in (4) and (5). But this criticism is ill-informed. For, if a Fregean statement of the truth conditions of quantified sentences is adopted, nothing more needs to be said about the pronouns with quantifier antecedents—they simply look after themselves.

Let us call expressions like 'some', 'many', 'the', 'a', 'few', 'every' *quantifiers*, and a quantifier plus a common noun, like 'some man', 'a girl', etc. a *quantifier expression*.²⁰ We extend the grammar to allow that any sentence containing a singular term in a position, p_i , remains well formed when a quantifier expression is substituted for it in p_i . (Inserting the quantifier into the singular term position being generalized upon does appear to be the way we indicate in English which quantifier goes with which position.) Since left-right ordering is not reliably used to indicate scope, we had better indicate with numerical subscripts the order in which the quantifiers are inserted in the construction of a sentence. Thus we distinguish

(Every₁ man) loves (a₂ woman) = There is a woman
every man loves

from

(Every₂ man) loves (a₁ woman) = For every man, there
is a woman he loves.

The semantic account of the quantifiers will have the form with which we are already familiar, but in which some account is taken of the common noun restricting the quantifier. Thus, for example, we have the following principle for the quantifier 'Every':

(G) If σ is a sentence containing in its i th singular term position the quantifier expression 'every' $\overbrace{\quad}^j \delta$ (where δ is a common noun and j an index) than

²⁰ I do not introduce quantifier phrases with relative clauses until sect. V.

which no other index attached to any quantifier in σ is higher) then σ is true (in L) iff on every extension of L with respect to some singular term β which does not already occur in σ on which the object which β denotes on that extension satisfies δ , σ^β/p_i is true.

(I write ' σ^β/p_i ' as an abbreviation for 'the sentence which results when β is substituted for whatever occurs in the i th singular term position in σ '.)

It should be easy to see how clauses like (F) and (G) suffice to deal with the occurrence of pronouns with quantifier antecedents. The grammar enables us to form sentences like:

(Some₁ man) loves () mother

realized in surface structure as

(4) Some man loves his mother.

Certainly we would get nowhere if we attempted to apply the rule dealing with pronouns directly, for the quantifier expression 'Some man' has no denotation. But if we apply the rule for the quantifier first, as we should, since the order in which the sentence is constructed gives it widest scope, we find that the truth of (4) depends upon whether or not there is some true (potential) substitution instance of the form

(β) loves () mother.

Now the pronoun rule can apply, eventually yielding the result that such a sentence is true iff the denotation of β loves the mother of the denotation of β . By elementary manoeuvres which can easily be imagined, we will arrive at the result that (4) is true iff there is a man such that that man loves that man's mother.

Just as a semantic theory which adopts a Fregean explanation for the truth conditions of quantified sentences can take the truth-functional role of the connectives as basic, so such a theory can regard pronouns as devices for registering co-reference (understood strictly). Or at least, so it would appear. We shall have to consider Geach's arguments against

this way of understanding pronouns, but I hope it does not appear *obviously* absurd and unpromising.

I suspect that many philosophers and linguists have meant no more by talking of the existence of variables in deep structure, and of the parallelism between pronouns of the vernacular and variables of quantification theory, than that there exists in natural language a device for marking chains of co-reference of which pronouns are the superficial manifestations.²¹ But this is certainly not what Geach intends by his claim on parallelism. For Geach's claim to be correct, pronouns must be the manifestation of a device which is essentially a device for the formation of complex predicates. With the Tarskian semantics in mind, we might say: a device that registers co-assignment, not co-reference. On the semantic theory just sketched, this claim will not hold good, as we may see by considering a sentence like:

If John is here, he will be sorry.

On the one hand, the expression:

If () is here () will be sorry

receives no semantic interpretation, and is not acknowledged to constitute a semantical unit, while the expression which is realized in surface structure as 'he will be sorry' is treated as a semantical unit, and is assigned *truth* conditions.

As I have said, Geach believes that any proposal which treats pronouns with singular antecedents as referring expressions can be shown to involve a definite mistake. We had better look at his arguments.

(i) The first argument I want to consider occurs in many places in Geach's writings. At one point, he put the argument like this:²²

²¹ See, for example, the discussion in G. Lakoff, 'Linguistics and Natural Logic', in D. Davidson and G. H. Harman (eds), *Semantics of Natural Language* (Dordrecht: Reidel, 1972) p. 633; and E. Bach, 'Nouns and Noun Phrases', in E. Bach and R. T. Harms (eds), *Universals in Linguistic Theory* (New York: Holt Rinehart & Winston, 1968).

²² Geach, *Reference and Generality*, p. 128. The same argument is used many times; see 'Ryle on Namely-Riders', *Logic Matters*, pp. 89-90, 'Referring Expressions Again', *ibid.*, pp. 98 and 101.

Let us consider an example:

(22) If any man owns a donkey, he beats it.

(23) If Smith owns a donkey, he beats it.

The pronoun 'he' is replaceable by 'Smith' in (23) without changing the import of the proposition; it is not thus replaceable by 'any man' in (22); so it looks as if it were a pronoun of laziness in (23), but not in (22). All the same, (23) predicates of Smith precisely what (22) predicates of any man; both contain the same unambiguous complex predicable 'If — owns a donkey, he beats it' . . . On the other hand, the proposition:

(24) If Smith owns a donkey, Smith beats it

contains the completely different predicable 'if — owns a donkey, Smith beats it'; when attached to the quasi-subject 'any man', this gives us the proposition:

(25) If any man owns a donkey, Smith beats it

which is wholly different in force from (22). Thus the wholly different sense of the predicables 'If — owns a donkey, he beats it' and 'if — owns a donkey, Smith beats it' shows that even in (23) 'he' has a definite logical role of its own and is not a mere pronoun of laziness—not a mere device for avoiding the repetition of 'Smith'.

It is true that in this argument Geach is opposing the view that such pronouns are pronouns of laziness going proxy for their antecedents, rather than the view that they have the same reference as their antecedents. However, this does not matter since, if his argument is effective against the latter view also, and, on other occasions on which it is deployed, the conclusion is expressly stated as that the propositions do not refer.²³

Later in the same work, Geach offers an exactly parallel argument in connection with the reflexive pronoun.²⁴ In essence it is this:

(1) 'Everyone contradicts himself' says of everyone what 'Hegel contradicts himself' says of Hegel.

²³ On the first and second occasions mentioned in note 22.

²⁴ The argument with reflexive pronouns occurs in Geach's *Reference and Generality*, p. 132. I have made an alteration in my presentation of this argument in response to an objection from Professor Geach.

- (2) 'Hegel contradicts Hegel' says of Hegel that he contradicts Hegel.
- (3) If, in 'Hegel contradicts himself', the import of 'contradicts himself' were the same as that of 'contradicts Hegel', then 'Everyone contradicts himself' would say of everyone that he contradicts Hegel, which is absurd.
- (4) *Therefore* in 'Hegel contradicts himself', the import of 'contradicts himself' is not the same as that of 'contradicts Hegel', and hence the pronoun 'himself' has a definite logical role of its own; it is not a mere device for referring to Hegel.

This argument is unsound; its second premise is false. Ironically, it is Geach, more than anyone else, who has been concerned to emphasize that it is false. Consider, for example, the following passage:²⁵

We may in some instances recognise a common predicate in two propositions even though this predicate is not an identifiable expression that can be picked out; for example, 'John shaved John' propounds the very same thing concerning John as 'Peter shaved Peter' does concerning Peter, and thus we may regard the two as containing a common predicate but this is by no means identifiable with the mere word 'shaved' occurring in both.

The significance of Geach's observation is this: seeing the import of the proposition 'Hegel contradicts himself' as being the same as 'Hegel contradicts Hegel' does not preclude us from seeing the same predicable occurring in 'Hegel contradicts himself' as occurs in 'Geach contradicts himself' and in 'No one contradicts himself'. To maintain, as Geach's argument appears to require, that 'Hegel contradicts Hegel' contains *only* the predicable '— contradicts Hegel' is to reject the possibility of multiple equally correct ways of breaking down a proposition into its constituents without which, as Geach has often reminded us, logic would be so hopelessly crippled.

²⁵ A similar point is found at many places in Geach's writings. See, for example, 'Logical Procedures and the Identity of Expressions', *Logic Matters*, p. 112; 'Names and Identity', in S. Guttenplan (ed.), *Mind and Language* (Oxford: Clarendon Press, 1975), pp. 139–40.

If we follow Geach, the phrase 'What "Hegel contradicts Hegel" says of Hegel' is an expression which fails to pick out a unique predicable—there are three different candidates. Correspondingly, on the co-referential treatment of pronouns, the same would be true of the expression 'What "Hegel contradicts himself" says of Hegel'. Perhaps this is what troubles Geach; perhaps he thinks that, unlike the proposition with repeated occurrences of the proper name 'Hegel', the proposition with the reflexive pronoun admits of only one breakdown into subject and predicate. But it is difficult to see why this should be so. For while the logical relationship between 'Hegel contradicts himself' and 'No one contradicts himself' requires us to see a common predicable, a similar connection between 'Hegel contradicts himself' and 'No one contradicts Hegel' would seem equally to suggest that they share a predicable.²⁶

(ii) The second argument concerns the sentence 'Only Satan pities himself'.²⁷

Moreover, it is not even true that when the antecedent is a singular term, it can always take place of a reflexive pronoun. 'Only Satan pities himself' and 'Only Satan pities Satan' are quite different in their import.

This objection may also be dealt with by taking seriously a point which Geach himself makes, namely that 'Only Satan' is not a singular term, but rather an applicational phrase in its own right.²⁸ Consequently, a pronoun which has the expression 'Only Satan' as its antecedent is not a pronoun to which the co-referentiality principle may be directly applied. Like the expression 'Someone other than', 'Only' is an expression which takes a name to form a quantifier. The truth conditions of the resulting proposition are essentially those given by Geach (in Fregean style):²⁹

'F(only α)' is true iff no interpretation of 'x' as a proper name makes 'F(x)' true unless 'x' names something that is named in or by ' α '.

²⁶ This point is made in Dummett, *op. cit.*, p. 14, and, with explicit reference to Geach's treatment of pronouns, by B. H. Partee, 'Opacity, Co-reference and Pronouns', in Davidson and Harman (eds), *op. cit.*, p. 436.

²⁷ Geach, *Reference and Generality*, p. 132.

²⁸ *Ibid.*, p. 188.

²⁹ *Ibid.*, p. 186.

Applying such a principle to the sentence

(Only Satan) loves ()

we will be directed to consider the truth value of substitution instances of the form

(β) loves ()

to see if we can find one which is true, yet in which β does not refer to Satan. Such 'exclusive propositions' then, can present no more difficulty for the co-referential treatment of pronouns than is presented by any other quantified sentence.

It is worth taking note of the fact that certain such 'exclusive propositions' are ambiguous. For example, the sentence:

(8) Only John loves his mother

admits also of the reading on which it is equivalent to 'Only John loves John's mother'. This ambiguity can be explained quite easily.

To secure the reading on which the sentence asserts that only John is an own-mother-lover, it must be built from the sentence frame

() loves ()'s mother

by the insertion of the quantifier expression 'Only John', thus:

(Only (John)) loves ()'s mother.

On this reading, the quantifier has a scope wider than that of the co-referentiality brace. For the other reading, the sentence would be constructed from the sentence frame

(Only ()) loves ()'s mother

by the insertion of the brace and singular term 'John', yielding

(Only (John)) loves ()'s mother

in which the co-referentiality brace has a scope wider than the quantifier.

Similar ambiguities are found in a variety of sentences in which there occurs a modifier of a complex predicate containing a pronoun:

John is elderly for a man who loves his mother

John is the oldest man who loves his mother

John is too anti-semitic to love himself.

Sentences of this general character have been considered by B. H. Partee, who supposes that they constitute a difficulty for Geach's view of pronouns.³⁰ It is not difficult to see why this should be supposed. Capturing the reading of (8) on which it means that only John loves John's mother seems to require seeing the sentence as containing the predicable '(x) loves John's mother', which is then attached to the 'quasi-subject' 'Only John'. And so it might appear that at least *some* pronouns have to be regarded as referring to what their antecedents refer to, or at least as going proxy for their antecedents.

I do not think that the dispute between Geach's treatment and the co-referential treatment of pronouns can be settled by this kind of example. For, take that proposition which gives the truth conditions of a sentence upon the co-referential theory. Replace each occurrence of a recurrent proper name or singular term in that statement of the truth conditions with a variable, and thus form a predicate abstract. Now, envisage a semantic theory on which truth conditions are derived for the sentence by deriving satisfaction conditions for that complex predicate. Such is Geach's theory. For example, Geach can represent the ambiguity in (8) as that between the results of attaching to 'John' the different complex predicates:

\hat{y} [Only y : \hat{x} [x loves x 's mother]]

\hat{y} [Only y : \hat{x} [x loves y 's mother]]

³⁰ B. H. Partee, 'Deletion and Variable Binding', in E. L. Keenan (ed.), *Formal Semantics of a Natural Language* (Cambridge: Cambridge University Press, 1975).

(iii) The third argument I want to consider is not to be found in Geach.³¹ Consider the sentence:

(9) John thinks he's under suspicion.

For reasons which Castañeda has made his own,³² this proposition is not equivalent to the proposition 'John thinks that John is under suspicion'. I have heard it suggested that here at least we should see the pronoun 'he' as being used in the construction of the complex predicate 'x thinks that x is under suspicion'.

I want to concede right away that the pronoun 'he' in (9) does not have the function of indicating co-reference. This is not because, if it did have that function, (9) would then ascribe to John the notional belief that John is under suspicion, for it would not. The ascription of such a notional belief would be the result of seeing the pronoun in (9) as a pronoun of laziness, which is a different proposal from the one we are considering. But, nevertheless, it is true that the result of applying the principle (F) can amount to no more than the ascription to John of the belief of John that he is under suspicion. Interpreting pronouns with singular antecedents according to (F) renders the position occupied by them referentially transparent. Beyond acknowledging that the pronoun in (9) is not to be dealt with by principle (F), I have nothing definite to say about it; perhaps we should see (9) as somehow derived from the *oratio recta* sentence: 'John thinks "I'm under suspicion"'.³³

With that said, it is none the less true that the observation is quite irrelevant to the current dispute. For it is just as much a consequence of the suggestion that we see (9) as built up out of the complex predicate 'x thinks that x is under suspicion' and the proper name 'John', that the belief (9) ascribes to John is merely the transparent belief of John, that he is under suspicion. There is no magic in the recurrent

³¹ Hints of this argument are to be found on p. 29 of Partee's paper, 'Deletion and Variable Binding'.

³² H. N. Castañeda, "'He": a study in the logic of self-consciousness', *Ratio* 8 (1966), p. 130, and many other papers.

³³ This proposal is essentially made in G. E. M. Anscombe's paper 'The First Person', in S. Guttenplan (ed.), *op. cit.*, p. 47. See also Susumu Kuno, 'Pronouns, Reference and Direct Discourse', *Linguistic Inquiry* 3 (1972).

variable 'x' that somehow ensures that we assign a 'self-conscious' belief to John. The recurrent variable merely serves to ensure that, in considering whether or not a particular sequence satisfies the predicate, we assign the same object to both occurrences of the variable as its denotation-relative-to-the-sequence. Such an explanation makes no apparent sense of the recurrence of a variable both inside and outside an opaque context. An application of a mechanism for guaranteeing co-assignment to singular term positions in a complex predicate supposes those positions to be referentially transparent in just the way an application of a method for co-reference does.

(Since confusion upon this point has occurred in the literature, I think it is worth emphasizing that the relational formulation 'x believes of x that he is under suspicion' does not capture the idea of 'self-conscious' belief though, of course, if it did, so would the formulation '(John) thinks () is under suspicion'. Admittedly, it is unclear *what* the relational formulation captures, for it is unclear what additional premises are required to license the inference from the notational 'x believes that a is F' to the relational 'x believes of a (that is to say b) that it is F'. But all the principles that have been suggested would appear to allow that a man could believe of someone who was in fact himself that he is F, without knowing that it was he himself. If this is so, then it is possible for a, that is to say, b, to satisfy 'x believes of x that he is under suspicion' in virtue of a's possession of the notional belief that b is under suspicion (when he does not realize that he is b).³⁴

We have been able to find no reason for modifying the

³⁴ For an example of confusion on this point see e.g. Wallace's paper 'Belief and Satisfaction'. Wallace symbolizes a teleological principle as follows:

(x) (y) (z) [Wants (x, (y), z) and Can(x, (y), z) and Believes (x, (x, y, z), x̄ ȳ z̄ [Can (x, (y), z)] then Satisfies (x, z)]

and writes:

It is important to notice that this principle makes essential use of universal quantification into the argument place made available by the relational sense of belief. The reader may if he wishes give notional formulations . . . but I think he will find any such principles . . . distinctly implausible the reason being that it appears impossible to capture notionally the idea that Nelson believes of himself that he has the ability to perform the contemplated action.

view that a theory which sees pronouns as devices for marking co-reference (strictly speaking) is perfectly viable, provided that it incorporates a Fregean statement of the truth conditions of quantified sentences. In conclusion, I should like to mention a consideration which would give a decisive edge to one treatment of pronouns over its rival, *if* it applied, but which does not seem to me to apply, and some considerations which give the referential treatment a mild advantage over Geach's approach.

I come to semantic investigations with a preference for *homophonic* theories; theories which try to take serious account of the semantic and syntactic devices which actually exist in the language by deriving for each sentence of the object language a statement of truth conditions in which the very resources employed in it occur and are not analysed away in favour of resources which do not occur. To take a relatively trivial example, I would prefer a theory which was sensitive to the binary structure of the sentence 'All As are Bs' and which, being thus sensitive, was able to deduce the theorem that 'All As are Bs' is true iff all As are Bs, over a theory which is only able to deal with this sentence by 'discovering' hidden logical constants, and deducing the result that 'All As are Bs' is true iff all things are B-if-they-are-A. The objection would not be that such truth conditions are not correct, but that, in a sense which we would all dearly love to have more exactly explained, the syntactic shape of the sentence is treated as so much misleading surface structure.³⁵

Obscure though this formulation is, it is not necessary for the purposes at hand to make it any clearer. For while it must be admitted that the co-referential theory of pronouns does depart from homophony in that the truth conditions of the sentence 'John loves his mother' are given by the sentence 'John loves John's mother', no remotely homophonic theory constructed on Geach's lines seems in prospect. The introduction of variables and the parsing of singular sentences

³⁵ A consideration first stated explicitly in J. Wallace, 'On the Frame of Reference', in Davidson and Harman, *op. cit.*, p. 237. See also D. Davidson, 'In Defense of Convention T', in H. Leblanc (ed.), *Truth, Syntax and Modality* (Amsterdam: North-Holland, 1973), p. 83, and Kripke, 'Is There a Problem . . .?', p. 356.

of the kind we have been considering as involving complex predicates is as much a departure from the actual pronoun-antecedent construction which we find in English as one which eliminates pronouns altogether. Perhaps we just have to learn to live with the idea that this pronoun-antecedent construction is, in the relevant sense, just so much surface structure.³⁶

I will now mention some considerations which tell against Geach's way of treating pronouns, at least in so far as it is applied quite generally to pronouns with singular antecedents.

First, Geach's treatment requires such pronouns to be bound by names, which can be regarded, for this purpose, as singular quantifiers. He thus requires the scope of a name to include any pronoun which has that name as its antecedent. Now in extensional contexts names are scopeless, and thus no difficulties arise from this requirement since names can always be given maximum scope. But it does not appear to be true that names are scopeless in all contexts of natural language; such would be the case only if all name-containing contexts were referentially transparent. Now, take any opaque construction containing the proper name β , $O(\beta)$. If a pronoun which is outside the construction looks back to the name as antecedent, it cannot be dealt with as Geach requires. The only way of binding the pronoun would require the legitimate formation of the complex predicate

$$\lambda\{O(x) \dots x\}$$

which, as I have said, would render the context created by O transparent.

It does not appear difficult to construct sentences containing pronouns which look back to terms occurring inside opaque contexts; the following seem perfectly natural:

Oedipus thinks that Jocasta is childless, but she isn't

Giorgione was so-called because of his size, and he hated it.

Such sentences present no difficulty for the theory which

³⁶ It would be an interesting exercise for the reader to attempt to construct a strictly homophonic theory for the simple 'brace' notation for co-referentiality introduced above, or for that fragment of English with the devices 'the former' and 'the latter'.

treats pronouns in terms of co-reference strictly interpreted. Upon that view, all that is required is that names in opaque contexts have a referent; it is not required that their referent is all that is semantically relevant.

There is a second mild advantage to the referential treatment. Consider the following dialogue:

A: John came today.

B: Did he stay long?

It seems desirable to allow that B is using the same device of cross-reference to ask his question as we have seen used in the construction of a single sentence. (This seems particularly desirable when we observe that B may not have the epistemological resources to make a reference to John on his own account.)³⁷ It requires only a trivial modification of the grammar to allow the chaining of singular term positions to singular terms which occur in other sentences. No modification of the referential semantics is required at all, once we allow the units processed by our semantic theory to be chunks of dialogue, not just single sentences.

It does appear rather difficult to deal with such pronouns on Geach's view, since we would somehow have to see A and B engaged in the co-operative construction of a complex predicate in a way which appears inconsistent with assigning a truth value to either of their remarks taken independently. Do we not want to allow that B *contradicts* A when he continues

B: No he didn't?³⁸

Finally, I think that the suggestion that pronouns make a contribution which is to be explained by principle (F) can

³⁷ It is in this way that I would like to account for the fact, made much of by Kripke (see 'Naming and Necessity', in Davidson and Harman, op. cit.) that if A uses the proper name β with the intention to refer to whoever B was referring to when he used the name β , then the referent of β on A's lips will be the same as on B's. I do *not* wish to deal with it by so extending (and weakening) the concept of 'epistemological contact' that one is in such contact with an object x if one has simply come into contact with someone who uses a name to refer to x .

³⁸ In order to accommodate this simple logical relation between A's remark and B's, we must use the apparatus of co-referentiality we have been considering, and must not suppose that B's reference is *fixed by the description* 'the item A referred to by his use of the token "John"'.

claim to be somewhat more explanatory of the morphological shape which expressions that are used in this way actually have. First of all, we should realize that many expressions other than pronouns, strictly so-called, may be used exactly as pronouns are used. For example, 'that logician' is functioning like a bound pronoun in the sentence

Every logician was walking with a boy near that logician's house.

In fact, almost any singular term can be used to make a back-reference provided it is reduced in stress.³⁹ Thus consider the sentence

Amin was widely disliked, but the Ugandan president did not seem to mind.

We can see this sentence as resulting from the sentence frame:

(Amin) was widely disliked, but () did not seem to mind;

reduced stress being the superficial manifestation of the brace.

If we collect together all the devices that can occur in a singular term position chained to some other referring expression, we discover that they are all capable of being used, in other contexts, to make independent references. This little generalization will surely come as no surprise to one who holds a theory according to which expressions occupying such positions do refer (albeit with the aid of a co-referring device). But upon Geach's theory, this generalization must remain quite unexplained. For Geach, pronouns are part of a device for complex-predicate formation. Now, as I shall argue later, there are devices in English which have precisely the function which Geach assigns to pronouns—namely the relative pronouns 'who', 'which'. In the expression 'who loves Mary and whom everyone despises', the relative pronoun is being used to keep track of which position goes with which in the formation of a complex predicate.

³⁹ Not just expressions like 'the bastard' and 'the fool' as seems to be suggested by R. S. Jackendoff on p. 110 of *Semantic Interpretation in a Generative Grammar* (Cambridge, Mass.: MIT Press, 1972).

But if such is the function of pronouns, and of expressions which function like them, there would be no more reason to expect them to be capable of being used to make independent references than there is to expect 'who' and 'which', etc., to be capable of being used in this way.

IV 'E-type pronouns'

A *Introduction*

In this section, I wish to defend the view that some pronouns with quantifier antecedents are quite unlike bound variables; in particular they may be assigned a reference and their immediate sentential contexts can be evaluated independently for truth and falsehood. Such pronouns are not genuine singular terms in the sense in which ordinary proper names and demonstrative expressions are; rather they are singular terms whose reference is fixed by description.⁴⁰ How exactly we are to secure the right semantical results is a matter of detail which I will discuss later;⁴¹ to begin with I will be mainly concerned to establish the limited conclusion that such pronouns cannot be regarded as analogous to bound variables.

Consider the sentence

(10) John owns some sheep and Harry vaccinates them.

For all we have said up to now, we can see this sentence as built up in the following way:

() owns () and () vaccinates ()

() owns () and () vaccinates ()

(John) owns () and (Harry) vaccinates ()

(John) owns (some sheep) and (Harry) vaccinates (them).

⁴⁰ I borrow the notion of a description's fixing the reference of a singular term from Kripke's 'Naming and Necessity'.

⁴¹ See sect. IV (D) and sect. VII.

To see (10) as built up in this way is to see the quantifier 'some sheep' as having the whole sentence as its scope, and (10) as equivalent to

John owns some sheep which are such that Harry vaccinates them

or

Some sheep are such that John owns them and Harry vaccinates them.

Now, although it *may* be possible to construe the sentence in this way, it is not open to dispute that this is neither the only, nor the most natural, interpretation. Upon the most natural interpretation, the sentence would not be true unless Harry vaccinates *all* the sheep which John owns. A paraphrase of (10) upon that interpretation would be

John owns some sheep and Harry vaccinates the sheep that John owns.

In the same way, the sentence

(11) Few MPs came to the party, but they had a marvellous time

is not equivalent to

It holds good of few MPs that they both went to the party and had a marvellous time

both because (11) entails, while its supposed paraphrase does not, that few MPs went to the party, and also that *all* the MPs that came had a marvellous time. Similarly:

(12) Mary danced with many boys and they found her interesting

is not equivalent to

Mary danced with many boys who found her interesting.

What this strongly suggests is that we must see (10), (11) and (12) as the conjunction of two sentences with the scope of the quantifier going only to the end of the first conjunct.

So, the first piece of evidence that pronouns like those in

(10), (11) and (12) are not functioning like bound pronouns is that, if we interpret them as bound pronouns, we do not give the sentences the meaning they are most naturally interpreted as having. There is another piece of evidence. In none of the sentences can we substitute a quantifier of the form 'No' + common noun *salva congruitate*:

- * John owns no sheep and Harry vaccinated them
- * No MPs came to the party but they had a marvellous time
- * Mary danced with no boys and they found her interesting.

Now, upon the view that these pronouns are bound pronouns, this fact is inexplicable. For upon that view, if we remove the quantifier from these sentences, we are left with a complex predicate, which was affirmed to be satisfied in the case of some sheep, few MPs, or many boys, and which we ought to be able to affirm to be satisfied in the case of no sheep, no MP, or no boy. If the pronoun 'them' in (10) was genuinely within the scope of the quantifier 'some sheep', as it is in the sentence

John owns some sheep such that Harry vaccinates them,
then the ungrammatical sentence ought to have just the interpretation of

John owns no sheep such that Harry vaccinates them.

Essentially the same point can be made with the quantifier expression 'every' + common noun, once we move to singular sentences of the same syntactic structure. If the sentence

(13) Socrates owns a dog and it bit Socrates

was the result of attaching the complex predicate 'Socrates owns x and x bit Socrates' to the quantifier 'a dog', as is the sentence

A dog is such that Socrates owns it and it bit Socrates,
then the sentence

*Socrates owns every dog and it bit Socrates

ought to be well formed and have the same meaning as

Every dog is such that Socrates owns it and it bit Socrates.

What appears to be going on is this. The most important determinant of the scope of a quantifier is its syntactic position in the sentence. Roughly, and not inflexibly, the scope of a quantifier is naturally interpreted as constituting the smallest sentence which includes all the constituents which are *in construction with it*. (Klima defines the notion 'in construction with' so that a constituent is in construction with another iff the former is dominated by the first branching node which dominates the latter.) In (10), for example, the scope of 'some sheep' extends only to the end of the first clause. If we wish to say that some man is both bald and tall, we must not insert the 'some man' quantifier into the sentence frame:

() is bald and () is tall

in which its scope will reach only to the end of the first clause, but we must transform the tree so that the singular term position into which we propose to insert the quantifier *governs* 'tall'.⁴² For example, we may use the conjunction reduction transformation, to produce

() is bald and tall

or the 'such that' construction

() is such that () is bald and () is tall.⁴³

⁴² I follow the example of K. Wexler, P. Culicover, and H. Hamburger in calling the converse of the 'in construction with' relation, 'governs'. See *Learning Theoretic Foundations of Linguistic Universals*, Social Sciences Working Paper No. 60 (Irvine, Calif.: University of California, 1974), p. 42. I am grateful to Mr Geoffrey Pullum for the reference to this and other relevant literature in Linguistics.

⁴³ There is a range of transformations—passivization, conjunction reduction, Neg-placement, among them—which are only 'meaning preserving' when they apply to singular sentences. (See, for example, B. H. Partee, 'Negation Conjunction and Quantifiers: syntax vs. semantics', *Foundations of Language* 6 (1970), pp. 153-65.) This strongly suggests to me that the best course is to restrict such transformations to singular sentences, and to allow quantifiers insertion to take

We must suppose a prohibition to be in force against the insertion of a quantifier expression into a singular term position to which another singular term position which it does not govern is chained—the result would be an unbound pronoun.

Let us consider a few examples:

The man who owns Fido vaccinates him ('Fido' does not govern 'him').

Fido loves his master ('Fido' governs 'his').

Fido loves Mary and also loves his master ('Fido' governs 'his').

Fido loves Mary and he also loves his master ('Fido' does not govern 'he', nor 'his').

Mary owns Fido and beats him with her broom ('Mary' governs 'her', 'Fido' does not govern 'him').⁴⁴

Fido barks when he is happy ('Fido' governs 'he').

If Fido barks then he is happy ('Fido' does not govern 'he').

Either Fido is unhappy or he barks ('Fido' does not govern 'he').

I think it will be discovered that when an existential quantifier 'some dogs' is substituted for 'Fido' (and 'he', 'his' are changed to 'they', 'them' etc.) then the pronouns are naturally interpreted as bound pronouns, and the quantifier as having wide scope if and only if the quantifier expression governs the pronoun. Similarly with 'many dogs', 'few dogs', 'most dogs', etc. Equally, a well-formed sentence results when 'no dogs' is substituted for 'Fido' (and appropriate

place at any stage in the transformational cycle. If Fregean truth conditions are given for the quantifiers, this will enable us to give the meaning of any sentence affected by these transformations in terms of the equivalence of meaning between transformed and untransformed singular sentences. This is simply an extension of the strategy we have adopted for singular pronouns.

⁴⁴ It was this example which showed that the relevant relation is 'in construction with' rather than Langacker's notion of 'command', for 'Fido' does command 'him'. I am very grateful to Deirdre Wilson for pointing this out to me, and for suggesting that the relevant relation might be 'in construction with'.

changes are made to the pronouns) only when the quantifier governs the pronoun.

I said 'roughly and not inflexibly' for two reasons. First of all, we can often just about hear quantifiers which are not in a governing position as having wide scope. For example, it is just about possible to hear the sentence

If a friend of mine comes, we are done for
as being equivalent to

A friend of mine is such that if he comes we are done for.

Second, there are quantifiers which we almost always interpret as having maximum scope; 'any' is one, and 'a certain' is another. Thus:

If any dog is happy he barks

If a certain friend of mine comes, we are done for

are both naturally interpreted in a way which gives the quantifier wide scope.

As far as the present topic is concerned, these points do not matter. What matters is simply the fact that there are sentences containing pronouns, whose antecedents are quantifiers, but which are not naturally interpreted in the way that would result if the pronouns were bound by those quantifiers. For the semantic role of these pronouns, another account must be provided. It really does not matter that those sentences should also be capable of another interpretation, or that other sentences, of the same grammatical pattern but with different quantifiers, should not be capable of the troublesome interpretation. Nevertheless, I think that if we exclude the wide-scope-seeking quantifiers 'any' and 'a certain', the generalization I have offered, as to which pronouns will *naturally* be interpreted as being bound by a quantifier, is substantially correct.

I should say that I have adopted the less radical and possibly the less interesting explanation of the phenomenon of E-type pronouns. According to the explanation I have adopted, sentence frames of the form

() loves () and () loves (),

in which 'pronominalization' takes place across co-ordinate structures, can be generated, and underlie sentences of the form

(14) Mary loves John and he loves her.

But, on my account, we are prevented from inserting quantifiers into the chained singular term positions because their scope will not be interpreted as reaching across the co-ordinate structure to bind the pronouns. A more radical explanation would be one which supposed 'pronominalization'—or the drawing of braces between two singular term positions—to be restricted to those structures in which the singular term position to be occupied by the pronoun is governed by the singular term position to be occupied by the singular term. We would then see the pronouns in singular sentences, like (14), to be E-type pronouns, but we will so account for the semantic contribution of the E-type pronouns that when their antecedents are singular terms, the net effect is simply that of co-reference.

The attraction of this position is a measure of harmony it offers between the conditions on forward and backward 'pronominalization'. This *rapprochement* is particularly appealing since it turns out that Klima's notion of 'in construction with', which seems to account for the distribution of bound and E-type pronouns, has been called upon by those who are looking for constraints upon backward 'pronominalization'.⁴⁵ Unfortunately, I do not have sufficient competence in linguistics to be able to assess the plausibility of this more radical suggestion, for surely, if two fundamentally different processes are at work in the range of data which have hitherto been collected together as examples of 'forward pronominalization', then this fact should have countless ramifications and consequences of a purely syntactic character.

Nothing that I say of any importance hinges upon the truth of the more conservative explanation which I have offered. Nor, in fact, does it depend upon whether, in

⁴⁵ See, e.g., T. Reinhart, 'Syntax and Coreference', *Papers from the Fifth Annual Meeting of the North Eastern Linguistic Society* (Cambridge, Mass.: Harvard University Press, 1974), p. 92, and P. W. Culicover, 'A Constraint on Coreferentiality', *Foundations of Language* 14 (1976), p. 109.

Klima's notion of 'in construction with'. I have correctly identified the relevant syntactic relation. The semantic and syntactic properties of E-type pronouns are sufficiently well defined for the type to be recognized independently of its final location in the theory of syntax. It is important to my account that there be a syntactic distinction between bound and E-type pronouns; it is not essential that there should be the distinction between pronouns which are, and pronouns which are not, governed by their antecedents. However I shall proceed upon the assumption that 'in construction with' is the relevant relation.

The view that pronouns which are not governed by their quantifier antecedents are not functioning as bound pronouns rests, so far, upon two pieces of evidence. The first is purely semantic: interpreting the pronouns as being bound by the quantifiers does not enable us to capture the most natural interpretation which these sentences have. The second piece of evidence is the ill-formedness of sentences like

*John owns no sheep and Harry vaccinates them

*Every Londoner was there and he had a wonderful time.

Both pieces of evidence seem to me to point in the direction of treating these pronouns (E-type pronouns) as singular terms whose denotation is fixed by a description recoverable from the clause containing the quantifier antecedent. Thus the denotation of the pronoun in (10) is fixed by the description 'the sheep that John owns', in (11) by the description 'the MPs that came to the party', in (12) by the description 'the boys who danced with Mary' and in (13) by 'the dog that Socrates owns'. Roughly, the pronoun denotes those objects which *verify* (or that object which verifies) the sentence containing the quantifier antecedent. (This idea is made considerably more precise in section VI.) The first piece of evidence points in this direction simply because such appears to be the interpretation we put upon these sentences. The second piece of evidence points this way because if it is the role of such pronouns to denote the verifier(s) of the sentence containing its quantifier antecedent, then we can explain why E-type pronouns cannot follow sentences

whose quantifier is 'No' + common noun, and why *singular* E-type pronouns cannot follow sentences whose quantifier is 'Every' + common noun.

I do not believe that anyone has identified the class of E-type pronouns in just the way that I am suggesting, but philosophers have often suggested treating this or that pronoun in what amounts to an E-type way. In so doing, they have drawn upon themselves the vituperation of Professor Geach, who believes that any such proposal can be shown to involve a definite mistake. For example, Geach maintains that any analysis of the sentence

(13) Socrates owns a dog and it bit Socrates

as a conjunction of two propositions with a truth value would be 'inept'.⁴⁶ Elsewhere the proposal is described as 'quite absurd', 'a prejudice or a blunder'.⁴⁷ It is therefore with some trepidation that I confess to thinking that a conjunction of two propositions is precisely what (13) amounts to. Before we consider Geach's arguments and satisfy ourselves that there is nothing in them, let us fortify our spirits by looking at one or two additional advantages which an E-type treatment has over its bound-pronoun rival.

B Some advantages of the proposed account over its bound-pronoun rival

1 It is a feature of the account I am suggesting that the quantifiers which are the antecedents of E-type pronouns do not have to be given wide scope in order to bind them. There are several contexts in which giving these quantifiers narrow scope allows a simpler semantical treatment of those contexts, and there are others in which it appears absolutely inescapable. I shall deal with several such contexts in turn.

(a) Conditionals

Consider the sentences:

If Mary has a son, she will spoil him

If someone comes in this room he will trip the switch.

⁴⁶ Geach, 'Quine's Syntactical Insights', *Logic Matters*, p. 118.

⁴⁷ Geach, *Reference and Generality*, pp. 125 and 126.

These sentences appear to contain existential quantifiers, yet do not have a sense which would result from giving that quantifier wide scope; we are not saying that there is some boy such that if Mary had *him* as a son she would spoil him, nor that there is someone such that if *he* comes he will trip the switch.

The natural way to understand these sentences is as being built out of two propositions joined by a conditional; the antecedent being, for example, that some man comes, the consequent, said to be conditional upon the truth of an antecedent, being that the man who comes will trip the switch. (Obviously, if a paraphrase is to be given which uses descriptions explicitly, they must be understood as having a scope narrower than the connective.) To see the sentence in this way requires seeing the pronouns in the consequent as E-type pronouns, as would anyway be suggested by the grammatical relation in which they stand to their antecedent.

The problem presented in these sentences is familiar to students and teachers of logic—and it is normally got around by supposing in an *ad hoc* fashion that, in the antecedents of conditionals, the words 'some' and 'a man' are surface forms of the universal quantifier 'any'. This suggestion is adopted as a solution to the problem by Harman who writes: 'One plausible solution is to suppose that the deep structure quantifier in (77) [the sentence "If some arrows are green they will hit the target"] is not *some arrows* but rather *any arrows*'.⁴⁸ But as Harman points out, it is very difficult to characterize the contexts where the change from 'any' to 'some' takes place. Furthermore, as Harman also points out in the same place, a similar problem arises when any quantifier which is existential in force⁴⁹ occurs in the antecedent of a conditional such as in

⁴⁸ G. H. Harman, 'Deep Structure as Logical Form', in Davidson and Harman (eds), *op. cit.*, p. 45.

⁴⁹ A quantifier is *existential in force* iff, if a sentence of the form A(Q + Common Noun + Relative Clause), in which Q is the quantifier with maximum scope, is true, then A('Some' + Common Noun + Relative Clause), with 'Some' as the quantifier with maximum scope, is also true. 'Any' and 'No' are the most important quantifiers which do not have existential force; 'many', 'few', 'most', 'all' and each of the numerical quantifiers has existential force (as used in English).

If several/few/many/two/three . . . /men come they will be disappointed.

Can it be seriously proposed that 'any men' can turn into each of these? If not, how is the 'solution' to be generalized?

(b) *'Just one' sentences*

Consider the sentence:

Just one man drank champagne and was ill.

If we did not have to worry about such an occurrence of the pronoun 'he', easily the most simple treatment of the expression 'just one man' is to see it as a quantifier which, when attached to a predicate A, yields a truth just in case just one man satisfies A. To use such a treatment in the sentence above, we would have to suppose the scope of 'just one man' ends at the conjunction; to let it extend beyond the conjunction would be to generate the different proposition

Just one man drank champagne and was ill.

Treating the pronoun 'he' in the second clause as an E-type pronoun enables us to adopt this simple treatment of the expression 'just one man'.

If, on the other hand, we are to treat it as a pronoun bound by the quantifier, the best we can do is to adopt the suggestion Geach has recently made⁵⁰ that 'just one man' is a binary quantifier taking two open sentences, or predicates, to make a sentence.

The trouble is, there is absolutely no evidence that 'just one man' is, in this way, a binary quantifier; we seem to have no difficulty in forming a complete sentence by attaching the 'just one man' quantifier to a single open sentence, as in

Just one man opened the box.

Geach suggests that we can see that sentence as 'got by deletion' from

Just one man opened the box and he opened the box.

That Geach could make such a proposal suggests to me that

⁵⁰ Geach, 'Back-Reference', p. 204.

he and I are engaged on different enterprises. I am interested in the quantifiers and pronouns that occur in the English natural language (and in a good many others, I bet). I am not interested in the quantifiers and devices of back reference which exist in logically possible languages which we might speak but do not.

(c) *Relative scope difficulties*

An abstract description of a type of sentence that will present *relative scope difficulties* for Geach's view of pronouns is this: it contains some operator O within whose scope a quantifier expression must fall, and outside whose scope a pronoun looking back to that quantifier must fall. Schematically:

. . . O (. . . Quantifier + CN . . .) . . . it/he/etc. . . .

I shall present two examples of this kind of sentence, though no doubt many others can be found.

Consider the sentence:

Just one man owns a donkey and he beats it.

I am not now concerned with the pronoun 'he' which we have just been considering, but with the nexus: 'a donkey' . . . 'it', for taking 'Just one man' as our operator O, we have here an example of the kind described. Whether or not the sentence admits of a reading on which it is equivalent to

A certain donkey is such that just one man owns it and he beats it

(which I doubt), it also admits of another interpretation, on which the property said to be uniquely exemplified is that of being a donkey-owner. For this latter reading, the 'a donkey' quantifier has to have a scope narrower than that of 'just one man', which would then leave the pronoun 'it' unbound, if it were a bound pronoun. However, it is open to us to give 'a donkey' narrow scope if we treat the pronoun 'it' as an E-type pronoun—a treatment which is anyway suggested by the grammatical relation in which it stands to its antecedent.

Geach's reaction to this kind of sentence can be gauged

from a discussion he offered of a sentence which presents a very similar problem; namely:

The only man who owns a donkey beats it.⁵¹

He wrote:⁵²

We still have not an acceptable analysis of (1) [the sentence 'The only man who ever stole a book from Snead made a lot of money by selling it']; for the use of 'The only man who ever . . .' precludes our taking the initial ten word phrase in (1) to mean the same as 'The man who stole a (certain) book from Snead' . . . I think the right account of the initial ten word phrase in (1) is that it neither simply means the same as 'The only man who ever stole *any* book from Snead' as it does in (2) nor simply means what 'The man who stole a (certain) book from Snead' means . . . , but rather corresponds in force to a combination of the two: 'The man who stole a (certain) book from Snead, in fact the only man who ever stole any book from Snead.'

On Geach's analysis, then, we find *two* quantifiers 'a book'; one which is given wide scope, and which is therefore conveniently there to bind the troublesome pronoun, and one with narrow scope to give the intended uniqueness condition upon the description. It must be acknowledged to be desperately *ad hoc* to suggest that two quite different quantifiers,

⁵¹ This sentence certainly does present difficulties for Geach, but actually belongs with the sentences like 'Most men who own a donkey beat it,' discussed under (d) below. However, Geach's remarks, if appropriate at all, belong at this point, since his 'two-quantifier' solution obviously does not deal with the general problem presented by the sentences we shall consider under (d).

⁵² Geach, 'Referring Expressions Again', *Logic Matters*, p. 100. It is worth pointing out to those who might otherwise be misled that Geach's views on the treatment of pronouns in such sentences have undergone a complete change. In 1963 ('Referring Expressions Again') the suggestion made by L. Cohen ('Geach on Referring Expressions—A Rejoinder', *Analysis* 23 (1962-3), pp. 10-12) that such pronouns should be treated as pronouns of laziness was rejected with a certain amount of brusqueness. In 'Back-Reference' (1975), p. 195, without a word of acknowledgement, Geach makes the same proposal himself. The pronoun-containing sentence for which Cohen proposed a 'pronoun of laziness' account was:

The only man who ever stole a book from Snead eventually made a lot of money by selling it.

The sentence for which Geach proposes a 'pronoun of laziness' account is

The youngest man who brought a girlfriend to the party kissed her.

I do not myself favour the laziness account, but agree with Cohen that some other account than the bound-variable one must be given.

with different scopes, collapse into one surface structure. Is there any kind of sentence which exhibits this kind of collapsing? Are we not interested in how someone might understand such a sentence upon the basis of its structure?⁵³

Other examples conforming to the schematic description offered above can be constructed with the use of adverbial modifiers. Consider:

John stupidly touched some snakes and they bit him

when it is being used to assert that it was stupid of John to touch some snakes—any snakes—not that there were some particular snakes that it was stupid of him to touch. Equally:

John slowly kissed all the guests and they hated it

can mean that John was a slow all-the-guest-kisser, not that he slowly kissed each of the guests. But plainly, in neither of these two sentences can the pronoun fall within the scope of the adverb.

(d) *Quantifiers in clauses restricting quantifiers with higher scope*

Sentences which have the form of:

Most men who own a car wash it on Sundays

Every man who owns a donkey beats it

appear to conform to our schematic description of sentences which provide relative scope difficulties for Geach's approach to pronouns. If the sentence is to express the intended restrictions upon the major quantifier—that of being a car-or

⁵³ Geach is not alone in proposing a 'two-quantifier' solution to these and related difficulties; it is also to be found in N. W. Tennant's contribution to his joint paper with H. E. H. Altham, 'Sortal Quantification', in E. L. Keenan (ed.), *Formal Semantics of a Natural Language*, op. cit., pp. 46-60. (See especially examples (4) and (6) on pp. 53-4). Tennant's claim is that adopting a 'sortal logic' (apparently a binary structure for quantified sentences of natural languages) enables him 'to provide many English sentences with more congruous logical forms than they would receive in the classical predicate calculus'. In so far as this claim concerns sentences which are problematical because of the occurrence of E-type pronouns, it is entirely spurious. No essential use is made of the binary structures in dealing with these pronouns; the *ad hoc* introduction of an additional quantifier is a manoeuvre available to those working within the unary structures of the classical predicate calculus.

donkey-owner—it would appear that the second quantifier must be given a scope which does not extend beyond the relative clause, and this rules out a bound variable interpretation of the later pronouns. Further, with sentences such as these, Geach's 'two-quantifier' proposal does not even get off the ground, since they do not entail the sentence which results when the existential quantifier is given wide scope. This provides yet another reason against adopting his proposal in the case of a sentence like:

The only man who owns a donkey beats it

which should be seen as sharing a form with the initial pair of sentences. It entails a wide-scope sentence:

A donkey is such that the only man who owns it beats it, not in virtue of its form, but in virtue of particular semantic properties of the quantifier 'The'.

That the pronouns in these sentences are not functioning like bound variables is exactly what we should expect, since their quantifier antecedents most certainly do not govern them.⁵⁴ However, though it may be clear that they are not bound pronouns, what should *not* yet be clear is how they can be regarded as E-type pronouns. For on the account of E-type pronouns I have suggested, they are referring expressions; yet surely it is as silly to inquire after the reference of 'it' in our examples as it is in the case of a bound pronoun.

I cannot provide a complete answer to this objection until section V, in which relative clauses restricting quantifiers are studied in some detail. But I can indicate the main lines on which the answer will run by considering simpler sentences.

I am putting forward the view that, in the sentence:

John owns a donkey and beats it,

⁵⁴ It should be pointed out that the difficulty presented for the 'pronouns as variables' view by sentences like these was mentioned in two important papers by Lauri Karttunen: 'Pronouns and Variables', in *Papers from the Fifth Regional Meeting of the Chicago Linguistic Society* (Chicago, Ill.: University of Chicago, 1969), and 'Definite Descriptions and Crossing Co-Reference', *Foundations of Language* 7 (1971). See especially fn. 12 of the latter paper where Karttunen writes: 'Thus he [Geach] completely overlooks the fact demonstrated above that pronouns are sometimes used in a way which is not possible with variables in the more restricted syntax of the predicate calculus.'

the pronoun 'it' has the function of designating the object (if any) that verifies the antecedent clause containing the existential quantifier. However, the process of substituting quantifiers into singular term position can be indefinitely iterated: by substituting a quantifier for the remaining singular term, we may construct the well-formed sentence:

Every man owns a donkey and beats it.

Once the singular term 'John' is supplanted by the quantifier, the E-type pronoun can no longer be regarded as having a reference. But *no new explanation of its role in the resulting sentence is called for*. For, once again, provided the derivation of truth conditions for the quantified sentence runs through a stage at which the truth conditions of singular substitution instances are considered, the existing explanation of the role of E-type pronouns—that which assigns them a denotation determined by certain conditions—can be drawn upon. Since the 'every man' quantifier has a scope wider than that of the E-type pronoun, we do not begin evaluating the sentence by inquiring into the denotation of the E-type pronoun.

The situation is exactly the same as we found with pronouns that *were* governed by their antecedents. There, we discovered that not being able to assign 'himself' a denotation in the sentence:

Every man loves himself,

neither meant that we could not say that 'himself' refers to John in

John loves himself

nor that we need say anything more.

In a parallel way, I shall argue that the pronouns in the initial pair of sentences are E-type pronouns, but that they have a scope less than that of the main quantifier, and can (and need) only be interpreted relative to some substitution instance of the main quantifier. And, when we are considering such substitution instances, it is clear that we put an E-type interpretation on the pronoun. Once again, this comes

out most clearly with plural quantifiers. The evaluation of the sentence:

Every man who owns some sheep vaccinates them
in the spring

directs us to consider the relative truth values of pairs of sentences:

John owns some sheep; John vaccinates them in
the spring

when the latter sentence is true only if John vaccinates all the sheep he owns.⁵⁵

2 We have seen how Geach treats 'There is just one man who Fs and he Gs' as involving a binary quantifier. Since all the numerical quantifiers give rise to the problem to which this purports to be a solution, it is reasonable to suppose that Geach will adopt a similar account of the sentence:

Exactly two men got off their bicycles and then they
fainted.

But now consider the sentence:

Exactly three men got off their bicycles and they
pushed the Volkswagen up the hill.

Now, perhaps this sentence does have a reading on which it entails that each of the men who got off their bicycles separately pushed the Volkswagen up the hill—the reading which would result if 'Exactly two men' was a binary quantifier whose second open sentence is 'x pushed the Volkswagen up the hill'. But there is clearly another reading which is, in the circumstances, more likely, and according to which the men are said to have together pushed the Volkswagen up the hill. This reading cannot be captured along Geach's lines. In

⁵⁵ The fact that I consider pairs of sentences rather than a single sentence is a reflection of the fact that I regard these quantifiers as binary in form. But the current point does not depend upon that; we could impose a unary structure in the normal way, in which case the substitution instance would be

If John owns some sheep then he vaccinates them in the spring,
a sentence of the form considered under (a) above.

order to capture it, the scope of 'Three men' has to be closed off at the end of the first conjunct, and 'they' must involve a plural reference.

3 Consider the sentence:

Socrates owned a dog and it bit Socrates.

On the bound variable view, which Geach favours, the whole sentence is strictly equivalent to a sentence of the form:

(For some dog x) (Fx & Gx)

with its logician's paraphrase,

Some dog is such that Socrates owned it and it bit
Socrates.

If this is indeed what our original sentence means, what could explain our unwillingness to express the thought:

Some finger is such that it is Socrates' and it hurts
Socrates

by uttering

Socrates has a finger and it hurts Socrates?

What can explain our reluctance to report the existence of at least one woman doctor in Manchester by uttering:

There is a doctor in Manchester and she is a woman,

or the existence of a number with an even successor by uttering:

A number has a successor and it is even?

Suppose someone deduces that at least one and possibly several of the people at the meeting smoked, upon the basis of the fact that the room was filled with smoke. If Geach's 'wide-scope' rendering of these sentences was accurate, what could explain the oddness in reporting the result of this deduction in the sentence:

Someone came to the meeting and he smoked?

It is a quite clearly marked feature of the use of pronouns

that are not governed by their quantifier antecedents that one does not utter a clause containing such a pronoun unless one is in a position to answer the question: 'He? Who?' or 'It? Which?'⁵⁶ Perhaps this fact is not so striking to one who makes no distinction between pronouns that are, and those that are not, governed by their antecedents, for of course, no such requirement is made upon the use of a bound pronoun. But even so, it remains a pretty striking fact about the use of certain pronouns, not less striking for being at odds with the bound variable treatment of them. It is therefore surprising that nowhere in any of Geach's quite extensive writings on the subject of pronouns is it ever mentioned. This omission cannot be explained by Geach's professed lack of interest in the nuances of idiom; the fact I have just mentioned is no more a matter of nuance than is a blow from a sledge-hammer.

C *The arguments against*

It cannot be over-emphasized that the proposal that I am making concerns a limited, and syntactically identified, class of pronouns. I acknowledge that many pronouns whose antecedents are quantifier expressions do correspond to bound variables at least in the sense that it does not make sense to inquire into their denotation, but I doubt that all do. The possibility of a principled, syntactic demarcation of E-type pronouns from bound pronouns is really the strongest weapon in our defensive armoury. For I am sure that the consideration that has most influenced philosophers to falter in their defence of the view that this or that pronoun with quantifier antecedent has a reference, has been an incapacity to see how to differentiate such pronouns from genuine bound pronouns, in regard to which such a view is absurd. 'So I suppose you are going to say that "his" refers to some man, in "Some man beats his dog"', thunders the opponent, and, aghast at the prospect, one is shamed into silence. Well, that move can no longer be made. Let us see what other moves can be made.

Against my proposal, appeal has sometimes been made to

⁵⁶ This is implicitly concluded by Geach in the reply (19) given to B in an imaginary dialogue in 'Back-Reference', p. 199.

an *intuition* that pronouns in the two syntactic positions are functioning in the same way. Harman, for example, considers the sentence

If some arrows are such that those arrows are green,
those arrows will hit the target,

and observes that the second occurrence of 'those arrows' may be replaced by the phrase 'those arrows that are green' (as we would expect, if it is an E-type pronoun) and the first occurrence cannot be.⁵⁷ But against this he appeals to the fact that 'the phrase "those arrows" seems to have the same function in both its occurrences'. But, instead of using the observation to undermine the intuition, he uses the intuition to undermine the observation, concluding that, contrary to first appearances, there is no special connection between the second occurrence of the phrase 'those arrows' and the description 'those arrows that are green' after all. One wonders how the intuition of similarity of functioning would respond to the observation that the substitution of the quantifier 'No arrows' has a different effect upon the two occurrences:

If no arrows are such that those arrows are green, those
arrows will hit the target.⁵⁸

I *'A man' does not refer*

Geach tends to assume that anyone who holds that the pronoun 'it' in the sentence:

John owns a donkey and he beats it

has the role of referring to some particular donkey, must hold that its job is that of 'picking up the reference made by the expression "a donkey"'. And Geach has no trouble in showing that the expression 'a donkey' *never* refers.⁵⁹

⁵⁷ Harman, 'Deep Structure as Logical Form', p. 45.

⁵⁸ Geach makes a similar appeal to intuition in 'Referring Expressions Again', *Logic Matters*, p. 100. See the sentence: 'All the same the relation of the dangling pronoun "it" to its antecedent "a book" is pretty clearly the same in (1) as it is in (10)'.

⁵⁹ See, e.g., the discussion in Geach, *Reference and Generality*, pp. 6 ff. and in 'Back-Reference', p. 199.

The position which Geach refutes so decisively is a position which fails to make any discrimination between my two classes of pronouns and crudely extends the explanation in terms of co-reference which is appropriate for members of one class, to members of the other.

Such is not my position—and I wish to emphasize that *one is in no way committed, by saying that E-type pronouns have a referential role, to the view that their quantifier antecedents refer.* (In general the pronouns denote the items (if any) that *verify* the quantified sentence.)

2 'Buridan's Law'

Geach considers the sentence:

(15) Just one man broke the bank at Monte Carlo, and he has recently died a pauper

and writes the following:⁶⁰

Supposing the first half of (15) were true, it seems plausible to take 'he' as referring to the man who broke the bank at Monte Carlo; and then the second half of (15) would be true—and thus (15) as a whole would be true—iff it were true of that man that he had (at the time of the statement) recently died a pauper. But if the first half of (15) is false, there is no plausible way of specifying a reference for 'he'; yet (15) does not then cease to be a proposition with a truth value because an ostensibly referring expression in it fails to refer—(15) is then simply false.

It is a plausible principle that no proposition whatever is expressed by purportedly genuine singular terms which lack a referent. But the same does not hold for singular terms whose reference is fixed by description. Precisely because the term has its reference fixed by description, its reference may be specified, and therefore the truth conditions of any sentence containing it may be specified, whether or not it has a referent. Of course, we cannot specify the reference in the simple form

In the second clause of (15) 'he' refers to the man who broke the bank at Monte Carlo,

⁶⁰ Geach, *Reference and Generality*, p. 126. I have changed the number of the example to agree with our ordering in this and subsequent quotations.

for such a specification might reasonably be taken to commit us to there being such a man. The specification must rather take the form

For any x , the denotation of 'he' in the second clause of (15) is x iff x is the only man who broke the bank at Monte Carlo.

Thus Geach's claim that, if the first conjunct of (15) is false, there is no plausible way so to specify the reference for the pronoun, is just wrong.⁶¹

Whether we say that the smallest sentence containing an empty singular term whose reference is fixed by description is false or that it has a third but non-designated truth value is a matter which need not concern us here, for there will anyway be no difficulty in securing the result that (15), which is a conjunction of a false proposition with a proposition, is false.

What of 'Buridan's Law'? Geach states the law in various ways. On one occasion he stated it like this:⁶²

But as Buridan pointed out long since, the reference of an expression can never depend upon whether the proposition it occurs in is true or false.

If 'the reference of an expression' is interpreted as 'whether or not an expression has a referent', this principle can indeed be used against an E-type analysis of the pronoun in (15) but, upon that interpretation, it is wholly unacceptable, for it is then simply equivalent to a denial that there can be expressions whose reference is fixed by description. For, if an expression, α , has its reference fixed by the description ϕ , then whether or not α has a referent depends upon the truth value of the proposition 'There is something uniquely ϕ ', and then the conjunction 'There is something ϕ and $F(\alpha)$ ' would infringe the 'Law'. The 'Law' remains unacceptable even when it is restricted to atomic sentences, for under the same conditions, the perfectly acceptable proposition ' $\phi(\alpha)$ ' would infringe it.

⁶¹ For denotation clauses of this character see, e.g., M. A. E. Dummett, 'What is a Theory of Meaning?', in Guttenplan (ed.), *op. cit.*, pp. 110-11, and T. Burge, 'Truth and Singular Terms', *Noûs* 8 (1974), pp. 309-25.

⁶² Geach, *Reference and Generality*, p. 52.

If there is an acceptable version of the principle, it concerns the *specification* of the reference of an expression. We might put the principle like this:

It is unacceptable for the specification of the reference of an expression occurring in any sentence to explicitly mention the truth value of that sentence; so that the only way of determining whether or not some object x is the referent of that expression would explicitly require a prior determination of the truth value of that sentence.

The word 'explicitly' in the principle is doing some work. There can be no objection to so fixing the reference of an expression that a determination of whether or not some object is its referent should *as a matter of fact* involve a determination of the truth value of the sentence in which it occurs; such would be the case for any sentence $\phi(\alpha)$ when α has its reference fixed by the description ϕ . In such a situation, a clear way has been laid down for determining the referent of the expression and thus the truth value of the sentence; there just happens to be an overlap between the different stages of the operation of discovering its truth value. We find a genuine infringement of the principle in the paralogism of a kind which, according to Geach, prompted Buridan to state the principle: In 'Is A a donkey?' 'A' shall stand for you if the right answer is 'yes' and for Brownie, the donkey on the village green, if the right answer is 'no'. Here the *specification* of the reference of an expression *explicitly* involves the truth value of the very sentence in which it occurs, and Buridan is right to object to it. It is difficult to see what connection a principle concerned with such a case could have with the proposal that the pronoun in (15) refers. On that proposal, a clear route has been laid down for the determination of whether or not something is the referent of the pronoun, a route which does not *explicitly* involve the truth value of the first conjunct, let alone the truth value of the sentence in which it immediately occurs.

3 *Treating E-type pronouns as referring expressions involves assigning the wrong truth conditions to sentences containing them*

Considering the sentence:

(16) Socrates owned a dog and it bit Socrates

Geach writes:⁶³

A medieval would treat this as a conjunctive proposition and enquire after the reference (*suppositio*) of the pronoun 'it': I have seen modern discussion that made the same mistake. For mistake it is. If we may legitimately symbolize (16) as ' $p \wedge q$ ' then a contradictory of (16), correspondingly symbolizable as ' $\neg p \vee (p \wedge \neg q)$ ', would be (17) Socrates did not own a dog, or else: Socrates owned a dog, and it did not bite Socrates.

But (16) and (17) are not contradictories; a moment's thought shows that they could both be true. So ' $p \wedge q$ ' is an inept schema to represent (16).

Presumably Geach's idea is that (16) and (17) can both be true when Socrates owns at least two dogs. But, if Socrates owned two dogs, on the proposal which I am defending (16) is not true; the second conjunct would not be true for failure of reference of 'it'. So this is really not an argument at all, but a counter-assertion. Geach claims that the sentence like (16) means no more and no less than

Socrates owns a dog which (such that it) bit Socrates.

Should we be moved by this?

I have already given strong *prima facie* evidence against this claim of equivalence in meaning. It is easy to envisage circumstances in which someone might accept as true the sentence:

John has a finger which hurts him,

while rejecting its supposed paraphrase:

John has a finger and it hurts him.

We can easily envisage circumstances in which someone might accept as true the sentence:

⁶³ 'Quine's Syntactical Insights', in *Logic Matters*, p. 118.

John owns some sheep which he vaccinates,
while rejecting its supposed paraphrase

John owns some sheep and he vaccinates them.

Now of course it is always open to philosophers or linguists to reject such *prima facie* evidence as to the meaning of sentences in the light of a systematic theory, when that theory not only works more smoothly by assigning a meaning to those sentences other than that suggested by the evidence, but also explains why we react to the sentences in the way that we do (by showing how independently well-attested conversational factors deform and modify the sentences' strict and literal meanings). But such considerations of system and theory there are seem not to tell against, but rather to reinforce, the *prima facie* evidence of non-equivalence; in the preceding sections I have presented a mass of interlocking evidence that suggests that the pronoun in (16) is functioning in a way which is quite unlike the pronoun in its supposed paraphrase. And how can 'conversational factors' be invoked to account for the divergence sentences containing E-type pronouns have from their Geachian paraphrases, when the divergence depends crucially upon quite specific *grammatical relations* in which the pronoun stands to its antecedent?

In the light of all that has gone before, one who presses *this* argument of Geach's resembles, not so much someone executing an aggressive manoeuvre, but rather someone burying his head in the sand.

4 *We are obliged to introduce psychologizing*
Suppose someone says:

(18) A Cambridge philosopher smoked a pipe and he
drank a lot of whisky.

Suppose further that there were two pipe-smoking Cambridge philosophers, X and Y, one of whom did, and the other of whom did not, drink a lot of whisky. Now, given that there was a Cambridge philosopher who smoked a pipe, the truth value of (18) will be that of the second conjunct. And if we

treat 'he' in the second conjunct as a referring expression, then the truth value of the entire remark will be determined by whether or not it is X or Y we fix on as the referent of 'he'.⁶⁴

And so we might find ourselves trying to determine the truth value of (18) by asking who a man would have in mind when he uttered or wrote down the sentence (18) . . . Such psychologizing is really not necessary. . . .

The first point to observe is this: on the view of E-type pronouns I have so far outlined there is equally no licence to engage in 'psychologizing'. On that view, the second conjunct, being equivalent to the sentence:

The Cambridge philosopher who smoked a pipe drank
a lot of whisky

will not be true, because the pronoun lacks a referent. Though, as we shall see, this position involves a certain divergence from idiom precisely because it contains no psychologizing, if psychologizing is indeed to be deplored, it still represents a much better position than any that results from an attempt to read the pronoun in (18) as a bound variable. A conviction that the bound variable approach is superior can only come from a conviction that, in the circumstances outlined above, (18) would be true. And this, in its turn, must rest upon the view that sentences of the structure of (18) are equivalent to sentences of the form:

A Cambridge philosopher both smoked a pipe and
drank a lot of whisky.

And as we have seen, this view simply cannot stand up.

However, although this seems to me a perfectly adequate fall-back position, there does not seem to be any great harm in liberalizing the account we give of the truth conditions of sentences containing E-type pronouns with a dash of psychologizing, in the interests of a greater realism. For, when the speaker is manifestly *talking about* something,⁶⁵

⁶⁴ 'Logical Procedures and the Identity of Expressions', in *Logic Matters*, p. 11.

⁶⁵ I use the concept 'talking about' in a way quite different from the concept 'referring to'. One talks about an item *x* in uttering a sentence *S* which contains

for example, in narrating an episode, it is acceptable to continue with the use of an E-type pronoun even when the antecedent containing sentence or clause has not provided the basis for a unique specification. One might begin a story:

One day, a man and a boy were walking along a road,
and the man said to the boy: 'Would you like to be
king?'

One does not want to be committed, by this way of telling the story, to the existence of a day on which just one man and a boy walked along a road. It was with this position in mind that I stated the requirement for the appropriate use of an E-type pronoun in terms of having answered, *or being prepared to answer upon demand*, the question 'He? Who?' or 'It? Which?'

In order to effect this liberalization we should allow the reference of the E-type pronoun to be fixed not only by predicative material explicitly in the antecedent clause, but also by material which the speaker supplies upon demand. This ruling has the effect of making the truth conditions of such remarks somewhat indeterminate; a determinate proposition will have been put forward only when the demand has been made and the material supplied.

Actually, this way of 'fixing the reference' of an E-type pronoun can involve cancellation of explicit predicative material in the antecedent. Consider the exchange:

A: A man jumped out of the crowd and fell in front
of the horses.

B: He didn't jump, he was pushed.

It is tempting to see, in B's remark, an application of the same use of E-type pronouns as we have been considering, especially since it is quite difficult to make sense of it while construing 'he' as a bound pronoun.

If this liberalization is made, it is important to see that

the predicate F in such a way that S entails that something is F, iff, in uttering S, one is *expressing* a belief about x to the effect that it is F. Thus one may be talking about something even though one manifestly refuses to let one's audience know which item it is that one is talking about, and this is inconsistent with my, and I think any decent, concept of (speaker's) reference.

such psychologizing as it involves infects merely the truth conditions of the sentences containing the E-type pronouns. The truth conditions of the simple, unquantified sentence

A man jumped out of the crowd

can and should remain as given before; the undeniable fact that we may have particular individuals in mind in uttering such sentences must not be used to tamper with their truth conditions, which can obtain in virtue of the condition of some individual the speaker did not have in mind.⁶⁶

In attempting to formalize the treatment of E-type pronouns in the succeeding section, I shall ignore the wrinkle introduced by this liberalization. I hope it is obvious how it can be incorporated into the final product.

D Are E-type pronouns 'pronouns of laziness'?

It appears that any treatment of E-type pronouns that does justice to all the considerations we have mentioned will involve recovering a description from the sentence containing its antecedent. The ease and uniformity with which native speakers supply descriptions in answer to the questions 'He? Who?', 'It? Which' etc., when they are raised in connection with E-type pronouns, is certainly indicative of a rule-governed process. But should we see the E-type pronoun as 'going proxy for' this recoverable description, or as a semantic element whose reference is fixed by it? This question encapsulates a summary formulation of two different anaphoric processes. If an E-type pronoun is going proxy for the description, this would mean that the semantic evaluation of the sentence containing it proceeds exactly as if the description stood in its place. Now, there certainly are sentences in English for which we appear to need to invoke

⁶⁶ This is ignored by C. Chastain ('Reference and Context', in K. Gunderson (ed.), *Minnesota Studies in the Philosophy of Science*, 7 (Minneapolis, Minn.: Minnesota University Press, 1975), pp. 194-269), who invokes the concept of reference to give the truth conditions of sentences containing expressions like 'A man', at least partly because of the possibility of subsequent E-type pronouns. For reasons mentioned in note 65, I would also reject Kripke's suggestion (made in the John Locke Lectures for 1973) that succeeding pronouns should be dealt with by invoking the concept of speaker's reference in connection with sentences containing expressions like 'A man'; Kripke quite rightly insists upon leaving their truth conditions unchanged. Neither proposal seems necessary.

such a mechanism; the sentence (1), with which we started, is a good example, as are sentences which result from verb-phrase deletion, like:

John listened to music and so did Harry.

However, there seem to me to be at least two arguments against regarding E-type pronouns as going proxy for descriptions.

The first is purely semantic; the sentence which results when the description takes the place of the E-type pronoun (the 'prolix sentence') is often ambiguous in a way in which the original sentence is not. The trouble arises because definite descriptions give rise to scope ambiguities when interacting with almost all operators. As a consequence, prolix sentences have interpretations, not possessed by the original sentences, which result when the description is not given maximum scope (in its clause). Examples illustrating this point can be found with each of the familiar operators that give rise to scope ambiguities with descriptions.

Negation

John owns a donkey but it is not the case that it is male is unambiguous, while:

John owns a donkey but it is not the case that the donkey John owns is male

is ambiguous and might be asserted on the ground that there is no such thing as *the* donkey John owns, i.e. giving wide scope to the negation sign.⁶⁷

Modality

John owns a donkey and it likes carrots though it might not have been the case that it liked carrots

seems to me to be unambiguous, with just the referentially rigid reading, while its prolix version is ambiguous:

⁶⁷ Incidentally, this shows that Geach's two sentences (16) and (17) are not contradictories because they can both be false; it is not possible to use a sentence containing an E-type pronoun to form the full contradictory of another sentence containing an E-type pronoun, because the mere use of an E-type pronoun carries with it a commitment to the existence of a referent. E-type pronouns are like definite descriptions which insist upon widest scope.

John owns a donkey and the donkey John owns likes carrots although it might not have been the case that the donkey John owns likes carrots.

Time

Boston has a Mayor and he used to be a Democrat is unambiguous, while its prolix version is ambiguous:

Boston has a Mayor and the Mayor of Boston used to be a Democrat

Psychological attitudes

A man murdered Smith, but John does not believe that he murdered Smith

attributes to John merely a non-contradictory belief of the murderer that he is not the murderer, while:

A man murdered Smith, but John does not believe that the man who murdered Smith murdered Smith

is ambiguous, with one (unlikely) reading on which John is attributed the self-contradictory belief that the man who murdered Smith did not murder Smith.

There is another kind of argument against treating E-type pronouns as pronouns of laziness. It appears that those anaphoric devices which are evaluated as proxies actually originate transformationally; a deletion transformation being triggered by relatively superficial syntactic identity between the antecedent segment and the segment that is either deleted entirely, or replaced with a pro-form. Now, it is difficult to see how E-type pronouns could have had such a transformational origin. For the descriptions for which they are supposed to be going proxy do not correspond to any syntactically coherent unit in the antecedent sentence. The relevant description is 'reconstructible out of' the antecedent sentence but does not occur in it.⁶⁸ For example, in the sentence

⁶⁸ This is why the change Geach has made in the concept of 'pronoun of laziness' does matter. The original idea—of an expression 'eliminable in paraphrase by a repetition of its antecedent'—might correspond to some underlying syntactic reality, whereas with the later addition 'or by a repetitious phrase somehow reconstructible out of its antecedent', the possibility of such a correspondence seems to be ruled out.

John owns some sheep which bite their tails and they are on the mountain

the pronoun 'they' would have to stand for the description:

the sheep which John owns which bite their tails

the italicized portion of which certainly does not correspond to any syntactic unit in the antecedent sentence.

Syntactic theory is currently in too great a state of flux for much weight to be placed upon this argument. Nevertheless, I think we may claim that, when both these arguments are taken together, at least a *prima facie* case has been established against treating E-type pronouns as pronouns of laziness. More conclusive considerations can only be forthcoming in the context of a general theory of anaphora in English.⁶⁹

V Quantifiers with relative clauses

A *Relative clauses introduced*

We are now in a position to set about constructing a theory which incorporates a rather more rigorous treatment of E-type pronouns than exists in the informal remarks and hints I have offered so far. To do so, we must offer some account of the main devices with which E-type pronouns interact; this means that we must extend both the syntax and semantics of quantified sentences to allow quantifiers to be restricted by relative clauses. Once the extension is made we will have a fragment of language rich enough to allow pronouns to get up to all their distinctive tricks, and a theory rich enough to deal with most of the sentences which philosophers and linguists studying pronouns have found troublesome.

The main lines of the treatment are these. I regard all

⁶⁹ It was comforting to read the recent paper by Jorge Hankamer and Ivan Sag, 'Deep and Surface Anaphora', in *Linguistic Inquiry* 7 (1976), pp. 391-428. In it, they propose as generally applicable a distinction between anaphoric processes which correspond to the distinction I have made in the case of pronouns. In their terminology I am proposing that E-type pronouns are deep anaphors, and pronouns of laziness are surface anaphors. By their tests, which include an ambiguity test similar to that used in the text, deep anaphors are certainly what E-type pronouns turn out to be. See also O. Dahl, 'On So-Called Sloppy Identity', *Synthese* 26 (1973), pp. 81-112.

natural language quantifiers as binary, taking two predicates, or open sentences, to make a sentence. This is partly because of my desire to construct a homophonic theory of meaning—a theory which stays as close as possible to the grammatical structures that are actually found in the language. Therefore, I look with a rather jaundiced eye at the reduction of the apparently binary structures of 'Some As are Bs' and 'All As are Bs' into unary structures—using connectives to join the two general terms A and B to form a single predicate. But, also, since some quantifiers, like 'Most', appear to *need* a binary analysis, considerations of simplicity strongly suggest adopting it for all.⁷⁰

Intuitively, the first predicate of the binary structure has the role of identifying the objects whose satisfaction of the second is relevant to the sentence's truth or falsity, or, in Fregean language, the role of identifying the *relevant* (potential) substitution instances. (The relevant singular terms are those which, when coupled to the first predicate, yield a truth.) Hitherto, this first predicate has always been a single common noun. We are now to allow it to be complex. Whereas before the relevant range was restricted by expressions like 'man', 'donkey', now it may be restricted by 'man who owns a donkey', 'donkey that can bray', and so on.

To understand the role of the common noun + relative clause in this way requires that the restricting clause on the quantifier with maximum scope should yield, when attached to a singular term, a complete sentence assessable as true or false. This is just a matter of sensible logical procedure; if we did not follow it, no start could be made upon the assessment

⁷⁰ To guard against confusion it is worth distinguishing the binary structure here adopted from the binary structure suggested by Geach for the 'just one man' quantifier, and criticized above. The break in Geach's binary structure was to come at the point marked by 'and' in the sentence

Just one man opened the box and he went home

and there is no evidence whatever that 'Just one' sentences are ill-informed unless they have two such constituents. The binary structure I am suggesting for all quantifiers would discern in the initial conjunct the two constituent general terms 'man' and 'opened the box', and there is evidence that we need both of *these* constituents to have a well-formed sentence, though when we wish, in English, to approximate the effect of unrestricted quantification, the first constituent is the universal predicate 'thing' or 'object'.

of the whole sentence as true or false. It is also true that to understand the role of the expression common noun + relative clause in a quantified sentence along these lines brings me once again into headlong collision with Geach, who argues that the phrase cannot be understood as functioning in this way, and even its apparent unity is a kind of logical mirage. There is no more in Geach's arguments on this matter than there is to any of the arguments we have scrutinized on other matters, but for those who wish to rest their view of this subject on deductive rather than inductive grounds, I have included an Appendix specifically addressed to this point.⁷¹

I will treat the English relative pronouns, 'who', 'which', etc., as devices of predicate abstraction, enabling us to form a predicate:

(who) loves (John)

from a sentence frame with one free singular term position:

() loves (John).

Such predicates are satisfied by an object iff the sentence which results from substituting a singular term designating that object for the relative pronoun is true. (I shall ignore inflection and 'WH-movement' which together yield 'whom John loves' from '(John) loves (who)').

On this account,⁷² the use we make of the relative pronoun is very similar to the use made of variables in forming predicate abstracts—that of indicating the position in a sentence being abstracted upon. However, the relative pronouns are not used with the full generality of variables in formal languages, since chains of co-reference *within* a clause must be left to the pronominal apparatus. Thus we have:

$\overbrace{\text{(who) loves (his) father}}$

⁷¹ The Appendix was published in the following issue of the *Canadian Journal of Philosophy* 7, 4. [It is here reprinted as Chapter 5. Ed.]

⁷² This is essentially the account given by W. V. Quine in *Word and Object* (Cambridge, Mass.: MIT Press, 1960), pp. 110-14, and *Roots of Reference*, pp. 89-92.

rather than:

(who) loves (who)'s father.

However, when the sentence frame is truth-functionally complex we find the repeated relative pronoun doing exactly the job of recurrent variables:

(who) loves (John) and (who) does not love (Harry)

(who) loves (John) and (John) loves (who).

As we would expect, we have something parallel to 'lambda elimination':

(John) A-s iff John is one (who) A-s.

When we had one-word general terms restricting the quantifiers, there was no reason why the quantifier expression and the main clause into which it is to be inserted should not be generated separately, but if we are to generate, and evaluate, sentences with E-type pronouns in the main clause which look back to quantifier antecedents in the relative clause, the two constituents of a binary structure have to be simultaneously constructed. To see exactly why this is so, let us introduce E-type pronouns into the picture.

I shall indicate that a pronoun κ is being used as an E-type pronoun by prefixing it with the symbol '#'. Syntactically '# κ ' is a singular term. We also need a device for indicating which quantifier expression an E-type pronoun has as its antecedent. Since each quantifier acquires a numerical index as it is inserted into the main clause, to keep track of scope, we can attach the same index to the pronoun which has it as its antecedent. Following the old procedure, we would construct the sentence 'Socrates owns a dog and it bit Socrates' as follows:

() owns ()

() owns (a_1 [dog])

(Socrates) owns (a_1 [dog])

(Socrates) owns (a_1 [dog]) and () bit ()

(Socrates) owns (a_1 [dog]) and () bit (Socrates)

(Socrates) owns (a_1 [dog]) and (# it_1) bit (Socrates)

(Several steps of the construction could have been reversed without altering the overall effect.) So far so good. But now consider a sentence like:

Most men who own a car wash it on Sunday,

in which the E-type pronoun is not actually referring, since the process of substituting quantifiers into singular term position has been iterated after the stage at which the pronoun was generated. The construction of the quantifier phrase is not difficult:

() owns (a_1 [car])

(who) owns (a_1 [car])

Men: (who) own (a_1 [car])

Most [men: (who) own (a_1 [car])].

The main clause, into which we insert the quantifier, cannot yet contain '#it' for want of a suitable antecedent, so it must be:

() washes () on Sunday,

which after insertion yields:

(Most₂ [men:(who) own (a_1 [car])]) wash () on Sunday.

But, now, it is *too late* to insert the singular term '#it'. For to insert it at this stage would give it wider scope than 'most men' and have the consequence that the first operation in evaluating the sentence would be to inquire into the denotation of '#it'. This would be wrong, for at the first stage in the evaluation of the sentence, the pronoun cannot sensibly be assigned a denotation.

Put briefly, the problem is that the E-type pronoun in the main clause must be inserted *after* the 'a car' quantifier but *before* the 'most men' quantifier. This puts paid to the separate construction of the quantifier phrase and the main clause.

Adopting the binary form, the process of the construction of the sentence looks like this:

Car ; () owns ()

A [car] ; () owns ()

() owns (a_1 [car])

Man: (who) owns (a_1 [car])

Man: (who) owns (a_1 [car]) ; () washes () on Sunday

Man: (who) owns (a_1 [car]) ; () washes (#it₁) on Sunday

Most [man:(who) owns (a_1 [car])] ; () washes (#it₁) on Sunday

(Most₂ [man:(who) owns (a_1 [car])]) washes (#it₁) on Sunday.

(I shall continue to substitute the quantifier expression into a singular term position in the main clause because this does appear to be the way we indicate, in English, which position in the main clause is being quantified.)

The semantic evaluation will unpick what we have just knitted; the whole sentence will be true iff for most interpretations of β on which '(β) is a man (who) owns (a_1 [car])' is true, '(β) washes (#it₁)' is true.

B Pronouns in relative clauses restricting quantifiers

The first thing I want to establish concerns bound pronouns, namely that it is not possible on the account of bound pronouns which I have offered to generate sentences in which a pronoun in a clause which restricts a quantifier Q is bound by a quantifier Q' with a lesser scope than Q. It is an immediate corollary of this that there can be no pair of quantifiers each of which binds a pronoun in a clause which restricts the other. This is no weakness of expressive power, but on the contrary, a matter of correct logical procedure.

On my account, a pronoun is bound by a quantifier when that quantifier is inserted into a singular term position to which the position occupied by that pronoun is chained. So, if a pronoun in a clause restricting a quantifier Q is to be bound by another quantifier Q', the main clause into which Q' is to be inserted must look something like this:

$$(\) R(Q [CN:(who) R'(\)]) =$$

$$(\) \text{hurts}(\text{a woman}:(\text{who}) \text{loves}(\text{him}))$$

which would turn, with the insertion of a singular term into the unoccupied chain, into a sentence like 'John hurts a woman who loves him'. Now, the quantifier Q' could also have its restrictive clause, so that, before the insertion of Q', we might have a binary frame which looks like this:

$$[CN:(who) R'' (*)]; (\) R(Q [CN:(who) R'(\)]),$$

or:

$$[\text{Man}:(\text{who}) \text{despises}(*)];$$

$$(\) \text{hurts}(\text{a} [\text{woman}:(\text{who}) \text{loves}(\text{him})]).$$

But there is no way in which the empty singular-term position in the clause restricting Q', which I have marked with an asterisk, could be bound by the quantifier already *in situ*. We cannot *now* connect with a brace the *-position and the singular-term position occupied by the quantifier Q, for, although we are permitted to draw such braces at any stage in the construction of a sentence, we may do so only between two as yet unoccupied singular-term positions. But equally, such a brace could not have been drawn between these two positions *before* the latter was occupied by the quantifier phrase containing Q, for while we are engaged in constructing the sentence frame:

$$(\) R(Q [CN:(\text{who}) R'(\)])$$

we have no legitimate place for a sentence which is going to restrict a quite separate quantifier, and thus no opportunity of creating relations of co-reference between positions like * in such a sentence, and positions in the sentence we are constructing.

There is thus no way of ending up with sentences which would look like this

$$\text{(Every}_j[\text{man}:(\text{who}) \text{despises}(\text{her})]) \text{hurts}(\text{Some}_i[\text{woman}:(\text{who}) \text{loves}(\text{him})])$$

($i > j$ or $j > i$); and a good thing too, for they are obviously meaningless.

Now, I do not deny that there are intelligible sentences which have the superficial form of this meaningless sentence—for example:

A boy who owned them ran down some sheep that were in his way.

But elementary logical considerations enable us to deduce that, in this sentence, *either* the relative clauses do not restrict the quantifiers, *or else* the pronouns cannot be bound-variable-type pronouns. Since we can produce some examples of the same construction in which the clauses must be understood as restricting the quantifier, such as

Every boy who asked him eventually got his father to agree,

sentences of this general type can be used in the construction of an elegant demonstration that at least some pronouns in English cannot be construed as bound-variable-type pronouns. Yet, paradoxically, these very same sentences have been triumphantly cited as examples which can be dealt with only if we adopt a bound-variable treatment of pronouns!⁷³ It is easy to demonstrate that there is no way of assembling the 'deep-structures' proposed for these sentences into a sentence of quantification theory (restricted or unrestricted) in which every variable is bound.⁷⁴

What of E-type pronouns in relative clauses? An E-type pronoun evidently cannot have as its antecedent a quantifier with wider scope. But, can an E-type pronoun, *in a clause restricting* one quantifier, have a quantifier with lesser scope as its antecedent? If we try to construct such sentences, we find them very odd:

Almost every man who loves her kills one of his sisters

⁷³ By J. D. McCawley, in 'Where Do Noun Phrases Come From?', in R. A. Jacobs and P. S. Rosenbaum (eds.), *Readings in English Transformational Grammar* (Waltham, Mass.: Ginn, 1970), pp. 176-7.

⁷⁴ That such sentences pose a *problem* for the 'pronouns as variables' position was clearly indicated by Harman ('Deep Structure as Logical Form', pp. 41-3).

Everyone who inherits it sells a house.

The reason for the oddness is not hard to discover. Attempting to evaluate the sentence for truth, we peel off the quantifier with maximum scope, and address ourselves to this question, for example: 'Are all of the interpretations of β , on which " β inherited #it" is true, interpretations on which " β sold a house" is true?' Well, which interpretations are these? Understanding the E-type pronoun in the only way possible, the relevant interpretations must be those on which ' β inherited the house which he sold' are true. And this means that the discovery of whether or not the interpretation is relevant already presupposes that the predicate in the main clause applies. There seems no point in allowing such sentences to be constructed.

Not all relative clauses appended to a quantifier expression need to be interpreted as restricting it; it is always possible to regard the relative clause appended to a simple existential quantifier as a non-restrictive clause. And it appears that we only find intelligible occurrences of E-type pronouns whose antecedents are quantifiers with lesser scope in relative clauses upon which a non-restrictive reading may be imposed. Thus, the sentence:

A boy who owned them ran down some sheep that were in his way

may be interpreted as:

A boy, who owned them, ran down some sheep that were in his way,

and thus as equivalent to:

A boy ran down some sheep that were in his way, and he owned them.

It is easy to verify that we place an E-type interpretation upon the pronoun in the original sentence—for it to be true the boy has to own all the sheep he ran down.⁷⁵ Since

⁷⁵ Harman considers a sentence of this kind, namely: 'A boy who was fooling them kissed many girls that loved him', and claims that it seems 'roughly equivalent to'

A boy who was fooling many girls that loved him kissed and was fooling many girls that loved him.

they do not appear to raise any new problems of interest or importance, I shall not bother further with non-restrictive relative clauses.

Finally, there are some sentences in which the pronoun in the clause restricting a quantifier cannot be interpreted either as a bound pronoun or as an E-type pronoun, but has to be seen as a genuine pronoun of laziness. Consider the sentence:

Every boy who plucked up the courage to ask him got his father to agree

or the sentence:

The only pilot that shot at it hit the MiG that was chasing him.

These sentences appear to be interpreted as equivalent to their prolix versions:

Every boy who plucked up the courage to ask his father, got his father to agree,

and

The only pilot that shot at the MiG that was chasing him hit the MiG that was chasing him.

There is every indication that these cannot be E-type pronouns. First, we noticed a general difficulty in interpreting

(Harman, 'Deep Structure as Logical Form', pp. 42-3). Notice here the lengths to which one must go to produce a reading which depends upon the occurrence of E-type pronouns without actually acknowledging them. Essentially Harman treats the pronoun 'them' in the original sentence as a pronoun of laziness, but this yields only

A boy who was fooling many girls that loved him kissed many girls that loved him,

which does not entail that he was fooling all the girls he kissed (nor, in fact, that he was fooling any of them). So, mysteriously, instead of 'kissed', the sentence's predicate somehow becomes 'kissed and was fooling'. It is no wonder Harman ruefully observes that 'it is not at all obvious what transformations would be used' to get away from his deep structure to the original sentence. Anyway, Harman's ruse does not work generally, as can be seen from the non-equivalence of:

A boy who was fooling them kissed exactly two girls that loved him

A boy who was fooling exactly two girls that loved him kissed and was fooling exactly two girls that loved him.

sentences in which quantifiers having wide scope were restricted by clauses with E-type pronouns with quantifiers with smaller scope as antecedents. Secondly, E-type pronouns normally occur quite happily when there is no specific description in the antecedent clause; the material in the *whole* of the clause is used to fashion a description (and this is indeed why there is the general difficulty of interpretation mentioned in the first point). But we cannot construct happy sentences of this general shape unless there is a description in the main clause immediately ready for substitution. Thus the following sentence is infelicitous:

*Every pilot that shot at it hit a MiG that was chasing him.

Finally, we observed that E-type pronouns were referentially *rigid*, so that, if 'it' in the clause:

The only pilot that shot at it

were an E-type pronoun, the whole clause would be equivalent to one in which an explicit description was given maximum scope,

The MiG that was chasing him is such that the only pilot that shot at *it*,

which, if it is interpretable at all, would require that there was a MiG at which only one pilot shot, rather than, what the original sentence requires, that there was only one pilot that shot at the MiG that was chasing him.

Obviously, such pronouns cannot be bound-variable-type pronouns, for the elementary logical reason which I mentioned earlier (which is, of course, not to say that whatever interpretation we decide the sentence has cannot be represented in the notation of the predicate calculus).

It therefore seems reasonable to see at work in the construction of these sentences the 'pronoun of laziness' device which secures the correct interpretation and for which we have independent evidence.

C The occurrence of E-type pronouns

As a final preliminary to formalization, we must state a principle about when an E-type pronoun can look back to a quantifier antecedent. In view of the ill-formedness of sentences like

*John doesn't own a car, and he drives it on Sunday

and

*Either John owns a car or he drives it on Sunday

it is obviously not sufficient to require that an appropriate quantifier antecedent (i.e. one with existential force) should occur as a syntactically coherent string in preceding material. But, in view of the well-formedness of sentences like

Either John does not own a donkey or he keeps it very quiet

If John owns a donkey he keeps it very quiet,

it is too stringent to insist that the sentence containing the quantifier antecedent should be embedded in the whole sentence in such a way that the truth of the whole sentence should require its truth.

The principle these last two examples suggest is this. Let $\Sigma(\sigma, \sigma')$ be some sentence embedding a sentence, σ , whose main operator is a quantifier of existential force, and a sentence, σ' , containing an E-type pronoun looking back to this quantifier. For $\Sigma(\sigma, \sigma')$ to be well formed, it must be so constructed that, although there may be other ways in which it may be true, if there is a situation in which its truth or falsity turns upon the truth or falsity of σ' , this will be a situation in which σ will also be true. (A conjunction of σ and σ' is just a special case of this.) So, intuitively speaking, if the truth value of σ' matters, there will be something for the E-type pronoun to refer to.

For truth-functional modes of embedding, we can define the required relation as follows. Let us say that a sentence σ is *affirmatively embedded* in Σ relative to σ' iff, on all valuations ν to the constituents of Σ on which $\nu(\sigma') = \nu(\Sigma) = \text{T}$ and which are such that, for all valuations ν' which

agree with ν save that $\nu'(\sigma') = F$, $\nu'(\Sigma) = F$, then, on those valuations ν , $\nu(\sigma) = T$.

It seems natural to extend this concept of affirmative embedding to the binary structures in which quantified sentences originate, so that we can say that the sentence, B(who), which contributes to the restriction on a quantifier, is affirmatively embedded in the whole sentence relative to the sentence which becomes the main clause and into which the quantifier phrase is inserted. Let the quantified sentence, Σ , originally be of the form:

$Q_i[\text{CN:B(who)}] ; A(\quad)$.

Now, although we cannot in general identify circumstances in which the truth value of each such quantified Σ turns upon the truth value of a substitution instance $A(\beta)$, the truth value of relevant substitution instances are always *germane* to the truth value of Σ . But only *relevant* substitution instances are germane in this way. So that, whenever an inquiry into the truth value of a quantified Σ obliges us to be interested in the truth value of a sentence $A(\beta)$ containing an E-type pronoun, the sentence $B(\beta)$ will also be true, and there will therefore be something for the pronoun to refer to.⁷⁶

So we shall allow for the insertion of an E-type pronoun into any singular term position in a sentence in relation to which some other sentence, whose main quantifier is existential in force, is affirmatively embedded. Notice that this rule of grammar must be understood as relating to the deep structures generated by the base rules; by the time the superficial form of the sentence is determined, the antecedent sentence may have been deleted or transformed, thus for example,

⁷⁶ Truth and well-formedness cannot be simultaneously and interdependently defined, on pain of ill-formed sentences being presented to the semantic theory for evaluation. In order to avoid this objection, presented to me by Barry Taylor, we should regard the notion of affirmative embedding being defined over a fragment of English that does not contain E-type pronouns, and for which truth and well-formedness are independently defined. Then the grammatical rule extending the fragment to allow for E-type pronouns will be understood as relating to contexts certified in the smaller fragment as being of a type in which one sentence is affirmatively embedded in relation to another. The semantical theory for the larger language will differ from that for the smaller only in containing a single additional clause for the evaluation of E-type pronouns.

John does not own a donkey, but Harry does and he beats it every day

is well formed even though there does not appear in the surface structure an appropriate existential antecedent for the E-type pronouns. And the deep structure underlying

Either John does not own a donkey or he keeps it quiet
can also yield

Either John owns no donkey or he keeps it quiet.

Despite these examples, it seems necessary to state the well-formedness rule for E-type pronouns in terms of the occurrence of a specific kind of *syntactical* antecedent; a purely semantic criterion would not be able to explain the differing acceptabilities of:

John has a wife and she hates him

*John is married and she hates him.

VI Formalization

The purpose of constructing the following mock-up of the syntax and semantics of quantified sentences in English is simply to lend plausibility and explicitness to the distinction between the two kinds of pronouns – a distinction which I hope can be incorporated into whichever particular approach to the syntax and semantics of quantified sentences seems, in the light of detailed syntactic investigation, to be the most plausible. Although I have been concerned to adopt, in this mock-up, a syntactic account of quantified sentences of English which is at least not known to be foreign to them, in the way in which it is generally agreed among linguists that the syntax of unrestricted first order quantification theory is foreign to them,⁷⁷ I have not tried to present something which may be expected to form part of that final, most plausible theory. I shall suppress almost all

⁷⁷ See, e.g., J. D. McCawley, 'A Program for Logic', in Davidson and Harman, *op. cit.*, especially p. 530, and E. Keenan, 'On Semantically Based Grammar', *Linguistic Inquiry* 3 (1972).

syntactic complexities which are not relevant to my main theme, indicating with an asterisk those points at which the most considerable divergences from English proper occur; where the asterisk is not self-explanatory, an amplification follows in parentheses. I do not include any of those transformations which, though introducing redundancy at the level of singular sentences, are indispensable for the expressive power of the quantified fragment of the language.⁷⁸ Another omission will be any attempt to deal with plural reference, made possible by E-type pronouns with plural quantifiers as antecedents. This is not the place to explain, in general, how plural reference is to be understood; when it is understood the modification to the theory presented here will be obvious.

A Syntax

We suppose the fragment to contain a stock of predicates, indexed as to their degree, some of which are called *common nouns*. It also contains a stock of singular terms, some of which are called *pronouns* and a stock of quantifiers, none of which is plural* and some of which are called *existential in force*. We also have a stock of numerals called *indices*.

(1) If π is a predicate of degree n , π followed by* n singular term positions (written thus: $\pi() () \dots ()$) is a sentence frame _{n} . (The numerical subscript is a record of the number of singular term positions free in the frame; a singular term position is free in a sentence frame iff no expression has been substituted in it in the construction of the frame.)

(2) If σ is a sentence frame _{n} , $\lceil \text{Not}(\sigma) \rceil$ is a sentence frame _{n} * and if σ is a sentence frame _{m} and σ' is a sentence frame _{n} then $\lceil (\sigma \text{ and } \sigma') \rceil$, $\lceil (\sigma \text{ or } \sigma') \rceil$, and $\lceil (\text{If } \sigma \text{ then } \sigma') \rceil$ are sentence frames _{$(n+m)$} . σ and both σ and σ' , are respectively said to be *constituents* of the complex sentences formed by application of these rules, and the constituent relation is transitive.

(3) A common noun is a simple predicate expression₀.

(4) If σ is a sentence frame _{m} ($m \geq 1$) with position p_i free,

and δ is a simple predicate expression (common noun), then $[\delta: \sigma^{\text{WH}}/p_i]$ is a predicate expression _{$(m-1)$} .* (*No differentiation of relative pronouns; no provision for more than one occurrence of a relative pronoun in a single complex predicate.) (As before we write ' σ^ϵ/p_i ' for the result of substituting the expression ϵ in the position p_i in σ .) σ is said to be a constituent of the resulting predicate expression.

(5) If σ is a sentence frame _{m} ($m \geq 1$) and π is a predicate expression _{n} , then $(\pi : \sigma)$ is a binary sentence frame _{$(n+m)$} of which π and σ are said to be constituents.

(6) If ρ is a binary sentence frame _{n} of the form $(\pi ; \sigma)$ with p_i free in σ , and if Q is a quantifier and j an index, then $Q_j[\pi]$ is a quantifier phrase and $\rho^{Q_j[\pi]}/p_i$ is a quantified sentence frame _{$(n-1)$} provided that no quantifier in ρ has an index higher than j , and that no singular term position to which p_i is connected by a brace is not governed by it. (Observe that π does *not* become a constituent of the resulting σ .)

(7) For any n , if σ is a sentence frame _{n} with positions p_i and p_j free, then the result of drawing a brace connecting p_i and p_j and substituting a pronoun in one or other of p_i and p_j is a sentence frame _{$(n-1)$} .* (*No restriction on backward pronominalization, no gender agreement of pronouns, no pro-forms other than pronouns.)

(8) If σ is a complex sentence frame _{n} with constituents σ' and σ'' , where σ' is a quantified sentence frame whose quantifier is a quantifier of existential force whose index is the numeral j , and σ'' is a sentence frame in which the i th empty singular term position p_i of σ occurs, and if σ' is affirmatively embedded with respect to σ'' in σ , then if κ is a pronoun, $\sigma'^{\#\kappa j}/p_i$ is a sentence frame _{$(n-1)$} .

(9) If σ is a sentence frame _{n} with a singular term position p_i free, and τ is a singular term, then σ^τ/p_i is a sentence frame _{$(n-1)$} .

(10) All and only sentence frames₀ are sentences.

⁷⁸ See Dummett, op. cit., pp. 12-14.

B Semantics

I shall only state the principles for the devices with which we have been concerned in this paper.

Quantifiers. I shall take 'Every' as an example; clauses for other quantifiers can be straightforwardly derived from this example. Though the clause is stated in a semi-formal meta-language, observe that, if formalized in the language of the mock-up, it could yield strictly homophonic theorems.

If σ is a sentence frame_o containing in its i th singular term position the quantifier phrase 'Every' $\widehat{\kappa}_i$ [$\delta : B(WH)$] (where δ is a common noun and j an index higher than any index attached to any other quantifier in σ , and the constituent represented by $B(WH)$ is optional) then σ is true iff on every extension of the language with respect to some singular term β (which does not occur in σ), on which the object which β denotes on that extension satisfies δ , and on which $B^\beta(WH)$ is true, if there is such a constituent, σ^β/p_i is true.

Co-reference. If σ is a sentence frame_o containing positions p_i and p_j which are braced together, with p_i containing the singular term τ and p_j a pronoun κ , then the denotation of κ in σ is the same object as the denotation of τ .

E-type pronouns. The idea is to construct from the sentence containing the antecedent quantifier a description which is to fix the reference of the E-type pronoun. (Let us call this 'the antecedent sentence'.) In those cases where the E-type pronoun and its quantifier antecedent occur in coordinate clauses, the antecedent sentence is easy to identify; it is the smallest sentence which contains the quantifier and everything which it governs. But we have also allowed for the construction of sentences like

Most men who own a car wash it on Sundays

where the antecedent quantifier is in a relative clause restricting a quantifier with greater scope. In such cases, the question of evaluating the sentence containing the E-type pronoun will only arise relative to some substitution instance of that quantifier with greater scope: ' β washes it on Sundays'; and

then the antecedent sentence is the smallest singular sentence containing the antecedent quantifier and everything which it governs, formed by substituting the same constant (under the same interpretation) in that relative clause (' β owns a car').

As we saw when considering a sentence like

John owns a sheep which bites its tail and he beats it,

the reference of an E-type pronoun is fixed by a description which is formed from the antecedent sentence by the conjunction of (a) the main clause into which the antecedent quantifier is inserted ('John owns ()'), (b) the common noun in the antecedent quantifier expression ('sheep'), and (c) any relative clause restricting the antecedent quantifier ('(WH)bites ()'s tail'). In the example, the relevant description is 'the sheep John owns that bites its tail'.

These provisions are captured by the following laborious formulation.

If σ is a sentence frame_o which is a constituent of a sentence frame_o, Σ , and which contains the term ' $\# \widehat{\kappa}_i$ ' in its j th singular term position, and where σ' is the smallest sentence frame_o containing the quantifier with index i which occurs in Σ , and which is of the form $A(Q[\delta : B(WH)])$ (with the constituent represented by 'B(WH)' optional), and where there is no larger sentence frame in Σ which has σ as a constituent and does not have σ' as a constituent.⁷⁹

OR

If σ is a sentence frame_o which is a substitution instance with respect to the constant β of a sentence frame σ^* which

⁷⁹ The point of the clause 'and there is no larger sentence frame in Σ which has σ as a constituent and which does not have σ' as a constituent' is to ensure that the description which fixes the reference of the E-type pronoun has as wide a scope as does not include the sentence containing its quantifier antecedent. This will secure the referential rigidity which we observed these pronouns to display. At the same time, the scope of the description is not the whole sentence; so we do not end up with the inaccurate result that a sentence like

Either John does not own a donkey or he keeps it very quiet
is true if

The donkey which John owns is such that either John owns no donkey
or he keeps it very quiet
(and thus false if John owns no donkey).

is a constituent of a sentence frame_o, Σ , and σ contains the term $\widehat{\#} \widehat{\kappa}_i$ in its j th singular term position, and where σ' is a substitution instance with respect to that same constant β of the sentence frame σ'^* which is the smallest sentence frame_o containing the quantifier with index i which occurs in Σ , and where σ' is of the form $A(Q[\delta : B(WH)])$ (with the constituent represented by 'B(WH)' optional) and where there is no larger sentence frame in Σ which has σ^* as a constituent and which does not have σ'^* as a constituent

THEN

Any object, x , is the denotation of $\widehat{\#} \widehat{\kappa}_i$ iff x is the unique object which satisfies $A(\quad)$, δ , and $B(\quad)$ (if there is such a constituent), and σ is true iff, upon any extension of the language with respect to a constant, γ (which does not already occur in σ or σ') on which the denotation of γ is the same as the denotation of $\widehat{\#} \widehat{\kappa}_i$, σ'/p_j is true.

Pronouns, Quantifiers, and Relative Clauses (II)

It is occasionally tempting, after climbing a mountain, to use the elevation one has gained to dash up to the top of a connected peak which does not have sufficient interest to induce one to climb so high for its sake alone. It is in this spirit that I turn to Geach's Latin Prose theory of relative clauses. The matter itself is of no very great moment, and some new ground will have to be covered in dealing with Geach's arguments. Nevertheless we shall primarily be applying the theory constructed in the body of the paper, and when one is in a position to expose bad arguments relatively rapidly, it is perhaps a good idea not to leave them unchallenged, especially when they appear to be gaining currency.¹

The issue concerns the relative clauses which are appended to quantifiers. In the sentence 'Any man who owns a donkey beats it', it is natural to take the relative clause 'who owns a donkey' as going together with 'man' to form the logical unit 'man who owns a donkey' which is a (complex) general term appropriately substitutable for the schematic letter A in the schema: 'Any A is B'. On the intuitive view, the expressions for which the letters A and B schematically stand in 'Any A is B', 'Most As are Bs', 'Just one A is B', etc., are general terms—simple (like 'man') or complex (like 'man who owns a donkey')—but in either case terms which have an extension, and whose extension is relevant to the truth or falsity of the resulting quantified sentences.

From the *Canadian Journal of Philosophy* 7.4 (1977), pp. 777-97. Reprinted by permission. This article is an appendix to 'Pronouns, Quantifiers, and Relative Clauses (I)', *Canadian Journal of Philosophy* 7.3 (1977), pp. 467-536. [Reprinted as Chapter 4 in this volume, and referred to as 'the body of this paper' in the present chapter. Ed.]

¹ See esp. W. V. O. Quine, 'Reply to Geach', in D. Davidson and J. Hintikka (eds.), *Words and Objections* (Dordrecht: Reidel, 1969), pp. 331-2 and *The Roots of Reference* (La Salle, Ill.: Open Court, 1973), pp. 90-1.

On the treatment of quantified sentences constructed in the body of this paper, expressions of the form 'man who owns a donkey' are definitely regarded as genuine logical units. If we look at that treatment model-theoretically, quantifiers are regarded as functions from *pairs* of sets (of degree 1) to truth values, and it is the role of the complex general term appended to the quantifier to identify the first of the two sets. This can be brought out more clearly by adopting a more perspicuous notation than that of ordinary English, which we have been trying to mirror as best we can. Quantified sentences would appear as follows:

$$Qx (Ax; Bx)$$

where 'Qx' is schematic for quantifiers 'any', 'many', 'most', 'a', 'just one', etc.; 'Ax' for any general term, simple like 'man' or complex like 'man who owns a donkey', and 'Bx' similarly. The 'most' quantifier, for example, is associated with that function from pairs of sets to truth values which yields truth iff more members of the first set are members of the second than are not. 'Just one' is associated with that function from pairs of sets to truth values which yields truth iff the intersection of the sets has exactly one member. The quantifier 'The' is associated with that function from pairs of sets to truth values that yields truth iff the first set has exactly one member and it is also a member of the second. And so on.

On the binary analysis which we have adopted, expressions of the form 'man who owns a donkey' are therefore regarded as unquestionable semantic units whose semantic properties, crucially their extension, are relevant to the truth value of the quantified sentences in which they occur. Nor is such a binary treatment at all eccentric, or the exclusive property of those who have my unsound views of pronouns. For it has recently been widely recognized that the way of reducing the superficial binary structures of 'Some As are Bs' and 'All As are Bs' to the unary structures which are familiar from the classical predicate calculus cannot be generalized to all quantifiers.

For 'some' and 'all', and, if we wanted it undefined, 'no', we may form a single general term with the aid of a connective

from the two general terms which appear in the surface structure. In the case of 'some' the connective is 'and', the general term is 'being both A and B', and the function from sets to truth values associated with the quantifier yields truth iff the set is non-empty. In the case of 'all', the connective is 'if', the general term is 'being B if A' and the function from sets to truth values associated with the quantifier yields truth iff the set is identical with the universal set.

However, it appears that the quantifiers of ordinary language for which this reduction can be effected, or can be effected with relative ease, are the exception rather than the rule. A tremendous reworking of the surface form of the sentence has to be undertaken if the quantifiers in 'Exactly two As are Bs', 'The A is B', etc., are to be given unary structures. And, as I believe N. Rescher was the first to point out,² in the case of the plurality quantifiers 'Many As are Bs', 'Few As are Bs', 'Most As are Bs', 'Almost all As are Bs', etc., a unary structure does not appear to be workable at all. Since Rescher's observation, Wallace,³ Altham and Tennant,⁴ Lewis,⁵ Dummett,⁶ and Geach himself⁷ have suggested binary analyses of ordinary language quantifiers.

² N. Rescher, 'Plurality-quantification', *Journal of Symbolic Logic* 27 (1962) p. 374.

³ J. Wallace, *Philosophical Grammar*, Ph.D. dissertation, Stanford University, 1964 (Ann Arbor, Mich.: University Microfilms, 1971), esp. pp. 136-54.

⁴ J. E. J. Altham and N. W. Tennant, 'Sortal Quantification', in E. L. Keenan (ed.) *Formal Semantics of a Natural Language* (Cambridge: Cambridge University Press, 1975). In saying that they propose a binary analysis, I am cutting through their confusing terminology, and regarding their 'sortalizer' as the first constituent in a binary structure. The contrast they have in mind between 'sortal' (or 'restricted') quantifiers on the one hand and binary quantifiers on the other is never adequately explained, and in my view is better dispensed with. This will help to clear up one contradiction in their paper; for on p. 52 'Many Ps are Qs' is said to be *irreducibly sortal* while on pp. 50 and 56 sortal quantifiers are said to be replaceable everywhere by non-sortal quantifiers. The point is that 'Many Ps are Qs' is irreducibly *binary*, and it is not threatened by the replacement of Altham and Tennant's 'sortal' quantifiers by other *n*-ary ($n > 1$) quantifiers.

⁵ D. K. Lewis, 'Adverbs of Quantification', in E. L. Keenan (ed.) op. cit. pp. 3-15.

⁶ M. Dummett, *Frege* (London: Duckworth, 1973), p. 162. It is not so much a suggestion Dummett makes on his own account as one he offers to one who is impressed by Russell's theory of descriptions.

⁷ P. T. Geach, 'Back-Reference', *Philosophia* 5 (1975), pp. 204-5. Geach offers a binary treatment of the 'Just one' quantifier and also, with acknowledgement to Prior, to the 'the' quantifier. Presumably to maintain consistency with his

Given that there appear to be some quantified sentences of ordinary language which *have* to be assigned a binary structure, considerations of simplicity combine with the desire for homophony of which I spoke earlier⁸ to encourage us to assign binary structures even to those sentences which can, with a certain amount of distortion and uncovering of 'hidden connectives', be regarded as involving a single complex general term. This is certainly the reasoning I would use to defend the theory which I constructed in the body of the paper.

However, in the present context, it is not necessary to try to construct a defence of the *superiority* of the binary treatment. Once again, Geach's aggressiveness gives us an easy target, for he claims that a treatment along the lines I have sketched is not even viable. He writes:

... the complex term 'A that is P' is a sort of logical mirage. The structure of a proposition in which such a complex term occurs can be clearly seen only when we have replaced the grammatically relative pronoun by a connective followed by a pronoun [i.e. transform the apparently binary structure into a unary structure with the aid of a connective]; when this is done, the apparent unity of the phrase disappears ...⁹

For, as Geach says earlier:

Whereas ... 'gentleman *who* is so grossly insulted' looks like a logical unit, the string of words ... 'gentleman, if he is so grossly insulted' has no such look at all.¹⁰

views on relative clauses, the binary structure Geach assigns to the sentence 'The only bachelor who was at the party was F' is 'the only bachelor x (x was at the party; x was F)', rather than 'The x (x was a bachelor at the party; x was F)'. But there is no getting around the fact that the property which is required to be uniquely exemplified for the truth of the sentence is 'being a bachelor at the party'. In view of this, 'bachelor' "goes with" 'was at the party' in a way in which it plainly does not go with 'was F', and it seems pointless not to have a notation which registers this fact. See, e.g. B. H. Partee, 'Some Transformational Extensions of Montague Grammar', in Partee (ed.), *Montague Grammar* (New York: Academic Press, 1976), p. 66: 'Only by making the major syntactic division between "the" and "boy who lives in the park" can a uniform semantic treatment of the be given.'

⁸ See 'Pronouns, Quantifiers and Relative Clauses (I)', [reprinted as Chapter 4 in this volume. Ed.].

⁹ Geach, *Reference and Generality* (Ithaca, NY: Cornell University Press), p. 118.

¹⁰ *Ibid.*, p. 115.

Elsewhere Geach writes:

... the apparent unity of a complex term 'D that is P' is delusive; ... such a phrase has no more logical unity than, say, 'Plato was tall' has in the context 'The philosopher whose most eminent pupil was Plato was tall'.¹¹

These are strong words. It is not merely that Geach prefers an analysis of quantified sentences upon which the unity of the phrase 'D that is P' disappears. Geach's view is that any treatment of quantified sentences as containing the logical unit 'D that is P' must be *doomed*—just as any treatment which attempts to discern the constituent sentence 'Plato was tall' in the sentence 'The philosopher whose most eminent pupil was Plato was tall' is doomed.

So, it is sufficient to refute Geach's view simply to construct a viable semantic theory which does not treat phrases of the form 'D that is P' as logical units. And by this I mean, somewhat more precisely, a deductive theory which derives truth conditions for quantified sentences by making essential use of a lemma in which a semantical property is assigned to the constituent 'D that is P'; most plausibly an extension (model-theoretically) or satisfaction condition (truth-theoretically).

Geach seems not to be aware of the strength of his claim and of the appropriate ways to defend it. He replied on one occasion to an extremely pertinent criticism of his views:

If someone wishes to construct a formalized language that will refute my view of complex terms, it will not be sufficient for the language to contain structures analogous to ordinary-language complex terms, nor even for the language to lack the *ML* sort of eliminative definitions; he must also ensure that no such procedures of elimination can consistently be added to the language as rules of inference. Nobody has come near to meeting this condition.¹²

This is absurd. It is not a sufficient defence of the view that the apparent unity of the expression e in the sentence $S(e)$ is a logical mirage simply to show that $S(e)$ is equivalent, or even provably equivalent, to some sentence S'

¹¹ Geach, 'On Complex Terms', *Logic Matters* (Oxford: Blackwell, 1972), p. 104.

¹² *Ibid.*, pp. 107–8.

which does not contain any unitary expression corresponding to *e*. We do not show that the apparent unity of the expression ‘-p’ in the sentence ‘-(p ∨ -q)’ is a logical mirage by pointing to the equivalence of the whole sentence to the sentence ‘p & q’ which contains no constituent corresponding to ‘-p’. It was not enough for Russell to defend his view that the description ‘the ϕ ’ is not a genuine constituent of the sentence ‘The ϕ is F’ that he be able to show that the sentence containing the description is equivalent to some sentence in which there is no constituent corresponding to it. He had to show that no semantic theory which treated ‘the ϕ ’ as a genuine constituent could be constructed—to show, that is to say, not only that there is an equivalence between description-containing and non-description-containing sentences, but in addition that the equivalence would have to be *used* in the construction of a systematic theory of meaning for description-containing sentences, since no viable theory could operate upon the unreformed structure.

If we maintain a proper grip upon the polemical situation induced by Geach’s claim that the apparent unity of phrases of the form ‘D that is P’ is a logical mirage, we can see that his position has already been refuted. We have already shown how to construct a viable semantic theory in which such expressions are treated as logical units (following, it must be admitted, other philosophers, including Geach himself). So we could leave matters there, secure in the knowledge that there must be something wrong with Geach’s arguments. But a certain amount of instruction and even amusement can be derived from showing exactly what is wrong with them.

As far as I can discover, there are four arguments which Geach has advanced for his position, each associated with one of the following four sentences:

- (I) Any man who owns a donkey beats it.
- (II) The one woman whom every true Englishman honours above all other women is his mother.
- (III) A boy who was only fooling her kissed a girl who really loved him.
- (IV) Only a woman who has lost all sense of shame will get drunk.

Insuperable difficulties are supposed to arise when a theory which treats ‘D that is P’ phrases as logical units is applied to each of these types of sentence. I shall discuss the arguments associated with these examples in turn.

I

The argument connected with the first example is succinctly stated by Quine as follows:

On this view, which Geach calls the *Latin prose theory* of relative pronouns, it is wrong to treat ‘that found him’, or ‘man that found him’ as a term or as a self-contained grammatical entity at all. For, to switch to a medieval example that he adduces, take ‘man that owns a donkey’—as if to say ‘donkey-owner’. The sentences

Any man that owns a donkey beats it

Some man that owns a donkey does not beat it

would reduce to nonsense:

Any donkey-owner beats it,

Some donkey-owner does not beat it.

On his analysis, which renders ‘that’ as ‘if he’ or ‘and he’ and includes changes in word order, the sentences remain coherent:

Any man, *if he* owns a donkey, beats it

Some man owns a donkey *and he* does not beat it.¹³

It certainly is good to know that Geach’s analyses of the sentences are coherent. But if one is to extract from this an *argument* against the theory which Geach opposes, it must rest upon the following principle:

If the syntactically complex string *e* in the sentence *S(e)* is a genuine logical unit with the role of a general term, then the sentence *S(e')* which results from replacing *e* by another possibly unstructured, but synonymous general term *e'*, is also well formed.

Perhaps this is an acceptable criterion for some concept of a *self-contained grammatical entity*, but, despite Quine’s desire

¹³ *The Roots of Reference*, p. 90. See *Reference and Generality*, p. 117. The longer argument in *Reference and Generality* contains a counter-argument against a way of trying to get out of this ‘difficulty’. Since I shall argue the ‘difficulty’ is quite spurious we do not have to go into these ramifications.

to agree with Geach, this is not the concept Geach is concerned with. As a criterion for an expression's occurring as a genuine logical unit, I see absolutely no reason to accept it. It is tantamount to the denial of the very possibility of E-type pronouns.

Consider the sentence:

John owns a donkey and Mary beats it.

I have tried to show that we should see this sentence as a conjunction of two propositions each with its own truth value, despite the fact that the second clause contains a referential device ('it') whose reference is fixed by a description derived, according to a rule, from the first clause. We have two independent propositions expressed by two grammatically interdependent clauses. Given the grammatical dependence of the clause 'Mary beats it' upon the first clause, it is not surprising that the well-formedness is not always preserved when we substitute an equivalent clause of a different grammatical structure, as in:

*John is a donkey-owner and Mary beats it.

Given all that has gone before, we are hardly going to be induced to accept this fact as a proof that we do not have, in the original sentence, a conjunction of two propositions, in which the sentence 'John owns a donkey', and the general term 'owns a donkey', occur as genuine semantical constituents. By the same token, we are not going to accept the mere fact that one cannot substitute 'donkey-owner' for 'man who owns a donkey' in (I) as a proof that the latter expression does not there occur as a genuine semantical constituent. (I) has exactly the same form as a sentence which we analysed in detail in 'Pronouns, Quantifiers and Relative Clauses (I)',¹⁴ namely:

Most men who own a car wash it on Sundays

in which analysis 'man who owns a car' was treated as a genuine constituent.

Apart from resting upon a dubious criterion of an expression's occurring as a genuine semantic constituent, Geach's

¹⁴ [Reprinted as Chapter 4 in this volume. Ed.]

argument is defective in another way, which can be brought out if we pretend to accept it. Geach is opposing the view that (I) contains the constituent 'man who owns a donkey' as it would on the analysis into the binary structure:

Any (x) [x is a man who owns a donkey; x beats it].

Instead, he offers the structure:

Any man (x) (if x owns a donkey then x beats it).

Now, although Geach's parsing reveals no constituent of the form 'man who owns a donkey', it does appear to contain the complex predicate 'owns a donkey'. So, by the criterion we have just adopted, we ought to be able to replace this complex general term with the equivalent 'is a donkey-owner' *salva congruitate*; but we cannot:

*If any man is a donkey-owner he beats it.

It would appear that Geach is hoist with his own petard.

What this shows is that even if we adopt the criterion, the considerations of permissible substitutions upon which it rests do not really bear upon the nexus of common noun and relative clause at all. The criterion rules out the recognition of a constituent 'man who owns a donkey' in (I) only because it rules out recognition of the sub-constituent 'owns a donkey' (with a narrow scope existential quantifier). A narrow scope existential quantifier forces an E-type interpretation upon the pronoun in the second clause, which the criterion effectively rules out. Since it would appear that Geach's analysis also purports to contain the constituent 'owns a donkey', his parsing is no more in harmony with the criterion than that which it was designed to replace.

To this it will be replied that Geach need not regard 'a donkey' in his analysis as the trace of an *existential* quantifier, but rather of the wide scope *universal* quantifier 'any donkey', which is then able to bind the pronoun in the apodasis. So more fully, his analysis of the sentence would be:

Any donkey (y) Any man (x) (if x owns y then x beats y).

Now it is true that there is no longer any constituent

corresponding to the complex predicate 'owns a donkey' which might be replaced by 'is a donkey-owner'. But this only serves to show how irrelevant to the relation between common noun and relative clause these considerations are. For it is equally open to those who wish to see the common noun plus following relative clause form a coherent unit to avail themselves of a wide scope universal quantifier 'any donkey'. Their analysis of (I) would then look like this:

Any (y) [Donkey y ;
Any (x)[x is a man who owns y ; x beats y]]

or:

Any donkey is such that any *man who owns it* beats it, which becomes, after the quantifier 'A donkey' has been inserted in the argument place being generalized:

Any man who owns a (any) donkey beats it.

In the italicized constituents we have the kind of unit which Geach deems to be a mirage.

It is true that Geach appears to think that a genuine semantical constituent cannot be bound into from the outside, but that view is the stuff of which the next argument is made.

II

Geach writes:

Consider the pair of sentences:

(19) The one woman whom every true Englishman honours above all other women is his mother.

(20) The one woman whom every true Englishman honours above all other women is his Queen.

In (20) it is tempting to construe the string 'woman whom every true Englishman honours above all other women' as a general term A ; but we surely cannot do this in (19), or else (19) would imply that the one and only A is the mother of each true Englishman. The noun phrase theory cannot resolve the difficulty.¹⁵

¹⁵ Geach, 'Quine's Syntactical Insights', *Logic Matters*, p. 122. See also 'On Complex Terms', pp. 102-5.

For a reason I will explain below, Geach's second example is in fact unsound for making his point, so I shall change it to:

(IIa) The one woman whom every true Englishman honours above all other women lives in Buckingham Palace.

(Since I have not introduced resources for dealing with adjectives in attributive position, I shall suppress 'true' in 'true Englishman'.) Let us first construct these two sentences in the grammar we have outlined.

Sentence (II)

() honours () above all other women
(Rule) 1; sentence frame₂

woman: () honours (who) above all other women
4; predicate expression₁

() is ()'s mother
1; sentence frame₂

woman: () honours (who) above all other women;
() is ()'s mother
5; binary sentence frame₃

(The₁ [woman: () honours (who) above all other women]) is ()'s mother
6; quantified sentence frame₂

(The₁ [woman: () honours (who) above all other women]) is ()'s mother
7; quantified sentence frame₁

Englishman; (The₁ [woman: () honours (who) above all other women]) is ()'s mother
5; binary sentence frame₁

(The₁ [woman: (every₂ [Englishman]) honours (who) above all other women]) is ()'s mother
6; quantified sentence frame₀

'The woman whom every Englishman honours above all other women is his mother.'

Sentence (IIa)

() honours () above all other women
1; sentence frame₂

Englishman; () honours () above all other women
5; binary sentence frame₂

(every₁[Englishman]) honours () above all other women
6; quantified sentence frame₁

woman: (every₁[Englishman]) honours (who) above all other women
4; predicate expression₀

lives in Buckingham Palace 1; sentence frame₁

woman: (every₁[Englishman]) honours (who) above all other women; () lives in Buckingham Palace
5; binary sentence frame₁

(The₂ [woman: (every₁ [Englishman]) honours (who) above all other women]) lives in Buckingham Palace
6; quantified sentence frame₀

'The woman whom every Englishman honours above all other women lives in Buckingham Palace.'

I confess that I do not see the difficulty. As the process of construction makes clear, the sentence which is supposed to present irresolvable difficulties for the noun phrase theory, (II), certainly contains a genuine logical unit formed from the common noun 'woman' together with its relative clause, but this relative clause contains an empty singular term position which is subsequently quantified upon by the wide scope universal quantifier 'every Englishman'. The underlying logical structure is masked by the convention, in English, of marking the singular term position being quantified upon by actually inserting the quantifier into it; the structure is better brought out in the paraphrase:

Concerning every Englishman, the *woman he honours above all others* is his mother

and even more clearly in the notation:

Every (x) [Englishman x; The (y) [Woman y & Honours above all etc. x, y; Mother y, x]]

and, in both, the italicized phrases constitute a unitary structure corresponding to the nexus of common noun plus relative clause.

Consequently, to argue upon the basis of sentences like (II) that a common noun plus its relative clause does not form a genuine logical unit seems to require the absurd assumption that a genuine logical unit cannot be quantified into. It is no better an argument than one which proceeds to the conclusion that an antecedent of a conditional does not constitute a genuine logical unit from consideration of such a sentence as:

If a (certain) colleague of mine comes we are done for when this is interpreted as having the structure:

There is a colleague of mine such that if *he comes* we are done for.

But upon neither of the semantic theories of quantified sentences which we have considered is it necessary to become perplexed by the semantic role of constituents like 'he comes' in my example, or 'woman he honours above all others' in Geach's. (One could almost say that it is precisely at this point that one can manifest the understanding which a genuine semantic theory for the quantifiers produces.)

On the Fregean semantics, a constituent with an unbound, or free, pronoun is assigned no semantic interpretation at all; for a Fregean, all antecedents of conditionals are closed sentences, and all (possibly complex) general terms restricting quantifiers are closed general terms. But, of course, Geach can take no comfort from this fact. For the Fregean need only consider this case; by the time the role of the conditional, or the innermost quantifier, is to be evaluated, the sentence will have metamorphosed itself so as to exemplify it. We will be considering the truth value of some potential sentence of the form

'If β comes we are done for' or 'The woman whom β honours above all other women is β 's mother'. Just as we observed that the Fregean could content himself with an account of the role of the conditional as it stands between constituents with a truth value, so he can content himself with an account of the role of the constituent formed from common noun and relative clause as that of introducing a complete general term which restricts the range of the quantifier.

If the semantic theory is constructed along Tarskian lines, one can recognize the constituent containing an unbound pronoun directly. The constituent which restricts the innermost quantifier 'woman he honours above all other women' is not a closed general term, satisfied or not satisfied by an object absolutely, but an open general term, satisfied or not satisfied by an object only relative to certain assignments to the pronoun 'he'. Since, upon a Tarskian theory, the semantic value of all expressions is computed relative to such assignments, the treatment of a quantifier restricted by an open general term is exactly the same as one restricted by a closed general term.

It may be thought that it is unnecessary to go over such elementary matters, for Geach writes as though all that he is concerned to emphasize is that 'woman whom every Englishman honours above all others' (with the quantifier in place) does not form a genuine constituent, just as the closed sentence 'a colleague of mine comes' does not occur as a genuine constituent in my example. Geach does say:

It may already have occurred to some readers that the puzzle about (II) could be resolved by considering the different scope of 'every Englishman', in (II) and (IIa). In (IIa) the scope is only the clause 'whom every Englishman honours above all other women'; in (II) it is the whole proposition (that is to say, the whole proposition is one to the effect that every Englishman . . .). But this explanation does not conflict with mine; for if you say that the scope of 'every Englishman' is not confined to the clause 'whom every Englishman honours above all other women', then you are in effect saying that upon logical analysis the unity of the phrase 'woman whom every Englishman honours above all other women' breaks up.¹⁶

¹⁶ Geach, 'On Complex Terms', *Logic Matters*, p. 105. I have changed the numbering and the constituents to make them appropriate to the examples we are considering.

This is either quite disingenuous, or an appalling muddle. For although it is true that interpreting the quantifier 'every Englishman' as having wide scope does involve splitting up the phrase 'woman whom every Englishman honours above all others', *it is a split quite different from that which Geach is supposed to be demonstrating*. This is shown by the fact that even after the wide scope quantifier is taken out, we are left with a constituent ('woman whom he honours above all other women') whose unity Geach is committed to denying. The splitting-up which is involved in removing the quantifier from the singular term position upon which it quantifies is not aided by, and does not involve, breaking up the relative pronoun into a connective and a pronoun.

In conclusion, I had better explain why I did not make use of Geach's example: 'The one woman whom every true Englishman honours above all other women is his Queen.' In fact, this sentence has exactly the same structure as the other sentence in Geach's pair, (II), with a wide scope 'every Englishman' quantifier, and a narrow scope 'the one woman' quantifier. It entails that there is one (and only one) woman whom every true Englishman honours above all other women, not in virtue of its structure, but in virtue of the particular semantical properties of the word 'Queen': we know that every Englishman shares one and only one Queen. If we try to give wide scope to 'the one woman' and narrow scope to 'every Englishman', along the lines of:

There is one and only one woman who has this property:
(she is honoured by every true Englishman above all
other women) and this woman . . .

then the 'his' in 'his Queen' will be unbound. (There is no hope of an E-type interpretation of this pronoun.) Geach conceals this from his readers, if not himself, by adopting the incorrect reading:

There is one and only one woman who has this property:
(she is honoured by every true Englishman above all
other women and she is his Queen).¹⁷

¹⁷ 'On the other hand, (20), is "G(one woman)", where G is proxy for: Every true Englishman honours—above all other women, and—is his Queen.' Quine's *Syntactical Insights*, *Logic Matters*, p. 124.

By incorrectly incorporating the 'she is his Queen' constituent into the general term being said to be uniquely exemplified, Geach provides a way in which the 'his' in 'his Queen' can be bound, but only at the expense of attributing to the sentence a reading it does not have. I am sure Geach will acknowledge that sentences of the form 'The ϕ is F' with 'the ϕ ' having wide scope are not generally paraphrasable as 'Just one thing is both ϕ and F'.

Although I felt this mistake worth pointing out, it is not essential to Geach's intended reasoning, which may be adequately refuted on other grounds. That is why I switched examples.

III

As Geach himself observes, the difference between restrictive and non-restrictive clauses is very well marked in English and 'certainly does correspond to a logical difference'.¹⁸ The sentences:

All Americans, who are lovers of money, are disturbed
by this development

All Americans who are lovers of money are disturbed
by this development

clearly differ in their import. And it would appear to be quite correct to say that in the case where the relative clause is non-restrictive, 'Americans, who are lovers of money' does not form a coherent logical unit. But it does not serve Geach's purposes to show that *some* expressions of the form common noun + relative clause do not form coherent logical units. This would have been granted by absolutely everyone without the need for any elaborate arguments. Geach's concern is with the relative clauses on the other side of this well-marked division; though, of course, if he was right in his view of those clauses, the division would have to be reinterpreted.

Consequently, it is to no avail for Geach to produce compelling reasons for not regarding the nexus common noun + relative clause as a genuine semantical unit in sentences of

¹⁸ Geach, *Reference and Generality*, p. 113.

a type which we have independent reasons for thinking may involve non-restrictive relative clauses. (III) is such a sentence, and when we considered a sentence exactly like it, we were able to 'insert commas', yielding:

A boy, who was only fooling her, kissed a girl who
really loved him.¹⁹

For, as Geach says, inserting commas around the relative clause in sentences of the form 'An F who is ϕ is G' makes no difference to the import of the proposition.²⁰ This reply would not be available if Bach-Peters sentences could easily be constructed involving quantifiers whose relative clauses could not be interpreted as non-restrictive (e.g. 'all', 'every', 'most') but we gave reasons for thinking that they could not be. We saw that any sentences involving these quantifiers which had crossing pronoun-antecedent relations would have to involve genuine pronouns of laziness as in:

The only *boy who plucked up courage to ask him*
got his father to agree

and once the 'him' is replaced by the material for which it is going proxy, there is no difficulty whatever in interpreting the resulting expression 'boy who plucked up courage to ask his father' as a genuine logical unit, a general term appropriately replaceable for 'A' in the schema 'The A is F'.

IV

Discussing our sentence (IV) Geach says:

A person who asserts (IV) is by no means implying that a man, as opposed to a woman, will not get drunk; so we cannot regard (IV) as obtainable by reading 'B' in 'Only a B will get drunk' as the complex term 'Woman who has lost all sense of shame'. Plainly the right way to construe (IV) is this:

A(ny) woman will get drunk only if she has lost all sense of shame.

¹⁹ See 'Pronouns, Quantifiers and Relative Clauses (I)', [reprinted as Chapter 4 in this volume. Ed.].

²⁰ Geach, *Reference and Generality*, p. 114.

The 'Only . . . who . . .' of (IV) can be replaced by a connective and a pronoun, 'only if she'; and the apparent unity of the complex term has vanished.²¹

Geach's observation about (IV) is intended to refute an analysis of it along the lines of:

Only (x) [woman who has lost all sense of shame x ;
will get drunk x]

here 'only' is associated with that function from pairs of sets a, b such that $f(a, b) = \text{True}$ iff $b \subseteq a$. He proposes to replace it with:

Any woman (x) [will get drunk x only if has lost all
sense of shame x].

Now, the first thing to notice about (IV) is that it is ambiguous; it has the reading Geach suggests and also the reading he rejects. It is quite proper to report the state of affairs in which the only people who ever get drunk are women who have lost all sense of shame by saying 'Only women who have lost all sense of shame will get drunk' or (IV). Geach chooses to use an example in which such an interpretation is unlikely, but nevertheless it exists. By itself this does not matter much; Geach could gain some comfort from the existence of an English sentence which at least has one reading which involves, upon analysis, the breakup of the common noun + relative clause nexus. However, the observation is suggestive. For it suggests that the relevant set which is identified for purposes of contrast is not syntactically determined by the shape of the sentence but is pragmatically determined, with some assistance from the intonation contour with which the sentence is uttered. If this is so, we should expect to find interpretations of 'only' sentences in which the identification of the relevant set requires breaking into expressions which even Geach regards as genuine logical units. This is exactly what we find. Consider the sentence:

Only boys who lose their mothers before adolescence
have a troubled adult life.

²¹ Geach, 'Complex Terms Again', *Logic Matters*, p. 107; see also *Reference and Generality*, pp. 118-19 and 'Back Reference', p. 202.

Not only can one utter this sentence without intending to imply that no girls have a troubled adult life, but one may also not intend to imply that the only boys who have a troubled adult life are boys who lose their mothers before adolescence. Context, and perhaps contrastive stress on 'before', can make it clear that one's intention is to assert that, of boys who lose their mothers, only those who lose them before adolescence have a troubled adult life. The break Geach advocates into 'boys' and 'has lost his mother before adolescence' will not get the right result. This should no doubt dampen Geach's enthusiasm to use the interpretation of 'only'-sentences as a test of what is a genuine logical unit, and what is not.

A similar point can be made with the help of attributive adjectives. As Geach himself has been concerned to emphasize,²² some expressions of the form adjective + common noun ('a good king') are not analysable conjunctively ('is a king and is good'). Attributive adjectives are predicate modifiers, which take a predicate to make a new predicate, and cannot sensibly stand on their own. One would have thought that there is no doubt that the expression 'good king' forms a coherent logical unit in the sentence:

Any good king is loved by his subjects.

But now consider the sentences:

Only a good king is remembered after his death.

Only a large woman will get drunk.

Both of these appear to have the reading which Geach supposes is troublesome for the 'noun phrase theory'; i.e. in which they do not entail that no one other than a king than be remembered after his death, and that no one other than a woman will get drunk, respectively. Once again it would appear that 'only' sentences require us to break into genuine logical units to interpret them.²³

This concludes my examination of Geach's arguments

²² See Geach, 'Good and Evil', *Analysis* 17 (1956), p. 33.

²³ For a discussion of a related sentence, 'Only moderate students left', see E. L. Keenan, 'Quantifier Structures in English', in *Foundations of Language* 7 (1971).

which are supposed to demonstrate that the unity of the common noun + relative clause is a logical mirage. As the viability of the theory we had constructed which treated these expressions as logical units led us to expect, not one of these arguments produces any genuine difficulty for such a theory. So, for all Geach has said, the noun-phrase theory is as viable as the Latin Prose theory. I want to end by giving some arguments for regarding the Latin Prose theory as inferior if not unworkable.

On the theory we have advanced, a uniform role is adopted for the relative pronouns; they are devices for marking the position in a sentence which are being abstracted upon in forming a complex predicate. This theory deals not only with relative pronouns appended to all quantifiers in a uniform way, but also with the occurrence of relative pronouns in other contexts. For, of course, complex predicates occur in other contexts, and it is not surprising that we find relative pronouns there doing the same job. So, just as we have:

He is young for an executive

we have:

He is young for a man who has climbed Everest;

just as we have:

He wanted to be king

we have:

He wanted to be a man who has read everything;

and just as we have:

He became a man

we have:

He became a man who did not know his own name.

Finally, just as any simple general term may be used to pin down a demonstrative reference, as in 'That cat is mine', so a complex general term may be used for the same purpose, as in 'That cat which has a white patch on its back is mine.'

In addition, we may hold out the prospect of eventually

providing a uniform account of relative and interrogative pronouns. It is a familiar enough point that WH-questions can be regarded as having the form of the interrogative quantification:

For which x : $A(x)$?

where the variable marks the place in a complex predicate which the audience is requested to fill in. It is obvious how a predicate abstraction device should be usable in this way.

As far as I can see, Geach's suggestion that relative pronouns always split up into a connective plus a pronoun offers no such prospect. Nor can I see a suitable Geachian rendering of the relative pronouns in the non-quantificational contexts. Certainly someone satisfies 'a man who has climbed Everest' iff he satisfies 'man' and satisfies 'has climbed Everest', and this might suggest that some analysis involving conjunction (as the connective) can be worked up. But a conjunctive rendering is the one analysis which really does not serve Geach's larger purpose, for it can hardly be used to show that the original expression does not form a coherent logical unit.

Even if we restrict attention exclusively to quantificational contexts, Geach's analysis must be deemed inferior. First, it must count for something that no uniform account of the pronoun is offered; that 'we must divine from the context which connective is packed in with the portmanteau word "that"'.²⁴ The awkwardness which this non-uniform treatment introduces has been noticed by the grammarian McCawley. Considering the two sentences:

(72) At least some Americans want Nixon to invade
New Zealand

(73) Some, if not all, Americans want Nixon to invade
New Zealand

he says:

If *some* combines semantically with the conjoined sentence ' x is an American and x wants Nixon to invade New Zealand', *at least* in (72) should refer to a scale on which the different things that that sentence

²⁴ Geach, *Reference and Generality*, p. 116.

can be combined with appear, and (73) implies that *all* is on that scale; but *all* does not combine with 'x is an American and x wants Nixon to invade New Zealand'. . .²⁵

More important is the fact that Geach does not appear to be able even to offer an account of all the relative clauses that go along with quantifiers. As I reported earlier, no analysis of quantifiers like 'almost all', 'most', 'many', 'few', etc., is known which joins the two constituents apparent in surface structure with a connective to form a single general term upon which the quantifier can operate. Geach has acknowledged the difficulty in the following words:

The one sort of case I know where the noun phrase theory seems to have the advantage over the Latin prose theory comes in a rather outlying field of logic, *pleonotetic* logic, as it might be called—the logic of majorities . . . Consider:

(35) Almost every man who drives a car dislikes the police

The noun phrase theory gives the correct truth conditions for this: (35) is true just in case 'Almost every motorist dislikes the police' is true. But clearly neither of the following is equivalent to (35)

(36) Almost every man, *if he* drives a car, dislikes the police

(37) Almost every man drives a car *and he* dislikes the police

It would take me too far into pleonotetic logic to deal with this objection; I think it can be dealt with . . .²⁶

What, then, is the situation? On the one hand, we have a theory which is capable of dealing in a uniform way with the occurrence of relative clauses not only in all quantified sentences, but also in other contexts. By staying close to the forms quantified sentences actually take in English, the theory offers the prospect of homophonic truth conditions. And on the other? We have a theory which limps in almost every way. Far from being able to deal with relative clauses in other contexts, it cannot even deal with relative clauses in all quantified sentences. Where it works, it works by introducing departures, different in different cases, and more or less extensive, from the surface forms

of the sentences which it treats.²⁷ And, finally, not one of the arguments which are supposed to induce us to overthrow the former theory and swallow the latter stands up to scrutiny.

²⁷ Geach should not be encouraged in his discovery of, e.g., a hidden 'if . . . then' in the sentence 'Any man who owns a donkey beats it' by the natural English paraphrase 'Any man, if he owns a donkey, beats it'. For after all, we have 'Most men, if they are married, are happy' where the 'if' is not a trace of the logician's conditional. See D. K. Lewis, 'Adverbs of Quantification', op. cit., p. 11: 'I conclude that the *if* of our restrictive if-clauses should not be regarded as a sentential connective . . . It serves merely to mark an argument-place in a polyadic construction.'

²⁵ J. D. McCawley, 'A Program for Logic', in D. Davidson and G. H. Harman (eds), *Semantics of Natural Language* (Dordrecht: Reidel, 1972), p. 530.

²⁶ Geach, *Logic Matters*, p. 125.

Can There Be Vague Objects?

It is sometimes said that the world might itself *be* vague. Rather than vagueness being a deficiency in our mode of describing the world, it would then be a necessary feature of any true description of it. It is also said that amongst the statements which may not have a determinate truth value as a result of their vagueness are identity statements. Combining these two views we would arrive at the idea that the world might contain certain objects about which it is a *fact* that they have fuzzy boundaries. But is this idea coherent?

Let '*a*' and '*b*' be singular terms such that the sentence '*a* = *b*' is of indeterminate truth value, and let us allow for the expression of the idea of indeterminacy by the sentential operator ' ∇ '. Then we have:

$$(1) \nabla(a = b).$$

(1) reports a fact about *b* which we may express by ascribing to it the property ' $\hat{x}[\nabla(x = a)]$ ':

$$(2) \hat{x}[\nabla(x = a)]b.$$

But we have:

$$(3) \sim \nabla(a = a)$$

and hence:

$$(4) \sim \hat{x}[\nabla(x = a)]a.$$

But by Leibniz's Law, we may derive from (2) and (4):

$$(5) \sim(a = b)$$

contradicting the assumption, with which we began, that the identity statement '*a* = *b*' is of indeterminate truth value.

If 'Indefinitely' and its dual, 'Definitely' (' Δ ') generate

a modal logic as strong as S_5 , (1)-(4) and, presumably, Leibniz's Law, may each be strengthened with a 'Definitely' prefix, enabling us to derive:

$$(5') \Delta \sim(a = b)$$

which is straightforwardly inconsistent with (1).

Reference and Contingency

'A logical theory may be tested by its capacity for dealing with puzzles, and it is a wholesome plan, in thinking about logic, to stock the mind with as many puzzles as possible, since these serve much the same purpose as is served by experiments in physical science.'¹ This paper is an attempt to follow Russell's advice by using a puzzle about the contingent *a priori* to test and explore certain theories of reference and modality. No one could claim that the puzzle is of any great philosophical importance by itself, but to understand it, one has to get clear about certain aspects of the theory of reference; and to solve it, one has to think a little more deeply than one is perhaps accustomed about what it means to say that a statement is contingent or necessary.

The idea that there might be truths which are both contingent and *a priori* was thrown up by Kripke in the course of his celebrated discussion of the modal and epistemic categories to which the notions of the contingent and the *a priori* respectively belong.² There has been some discussion of the idea since Kripke raised it, all of it based upon the assumption that the existence of a statement which is both contingent and *a priori* would constitute an intolerable paradox. For example, Michael Dummett has argued that the fact that Kripke's views on reference and modality appear to lead to the recognition of the existence of *a priori* truths shows that something must be wrong with those views.^{2a} In other recent discussions, attempts are made to dissolve the puzzle by showing that, properly understood, the problematical statements are not both contingent and *a priori*. There seem to me to be clear logical and semantical errors in all

From *The Monist* 62, 2 (April 1979), pp. 161-189. Reprinted by permission.

¹ B. Russell, 'On Denoting', *Mind* 14 (1905), pp. 484-5.

² S. Kripke, 'Naming and Necessity', in D. Davidson and G. Harman (eds) *Semantics of Natural Languages* (Dordrecht: Reidel, 1972), pp. 253-355.

^{2a} M. Dummett, *Frege* (London: Duckworth, 1973), p. 121.

of these attempts, but more importantly, their starting-point seems incorrect. There is no paradox in the existence of statements which are both contingent and *a priori*, at least, not in the sense in which the problematical statements may be claimed to be contingent. There are two quite different conceptions of what it is for a statement to be contingent; statements may be, as we might say, *deeply contingent* or *superficially contingent*. Whether a statement is deeply contingent depends upon what makes it true; whether a statement is superficially contingent depends upon how it embeds inside the scope of modal operators. While it would be intolerable for there to be a statement which is both knowable *a priori* and deeply contingent, I shall try to show that there is nothing particularly perplexing about the existence of a statement which is both knowable *a priori* and superficially contingent, which is the most that the problematical statements may be claimed to be.

In Kripke's original presentation of the puzzle, and in all subsequent discussions, the problematical statements were formulated with the use of a very special kind of singular term, which I shall call 'a descriptive name': a name whose reference is fixed by description.³ In fact, the puzzle about the contingent *a priori* does not have any special connection with the theory of reference, since it is easy to formulate statements with the same initially puzzling combination of characteristics without the use of singular terms at all. Indeed, it is one of the most serious deficiencies in the existing attempts at a solution to the puzzle that they do not generalize to these other cases. Nevertheless, though the puzzle is not peculiar to them, and though they occur only infrequently in natural language, descriptive names are of some theoretical interest, and an understanding of their properties is essential to a critical appraisal of previous approaches to the puzzle. For these reasons, I shall devote the first part of this paper to an account of how descriptive names function. In Part II, I shall use this account to examine the previous attempts to dissolve the puzzle. In Part III, I attempt to get clear about

³ The term 'descriptive name' is also used by Strawson to refer to a type of definite description. See *Subject and Predicate in Logic and Grammar* (London: Methuen, 1974), p. 60.

the modal properties of statements, and in particular about the relation between the modal properties of a statement and its content. Finally, in Part IV, I attempt to explain how it is possible for a statement to be both (superficially) contingent, and knowable *a priori*.

I

A descriptive name is a name whose reference is fixed by description. This formulation covers two points. First, a descriptive name is a referring expression; it belongs to that category of expressions whose contribution to the truth conditions of sentences containing them is stated by means of the relation of reference. Second, there is a semantical connection between the name and a description; the sense of the name is such that an object is determined to be the referent of the name if and only if it satisfies a certain description. If we borrow an idea of Frege's (as expressed by Dummett) and think of a statement of what an expression refers to as simultaneously showing, or displaying, its sense, then we may say that a descriptive name has a sense which is displayed by the statement that it refers to whatever it is that satisfies such and such a description. In this way, a descriptive name has a descriptive content.

Very few names which naturally occur in ordinary language can be regarded as descriptive names. It is difficult to hold of ordinary proper names that there is some particular description semantically associated with the name. It is more plausible to hold the view which Wiggins put as follows:

The sense of a proper name simply consists in its having been assigned whatever reference it has been assigned; to know the sense of *n* is to know to which entity *n* has been assigned, a single piece of knowledge which may be given in countless different ways by countless different descriptions.⁴

Even when there is a community-wide association between a name and one description, as perhaps 'Homer' is now associated with the description 'The author of *Iliad* and the *Odyssey*', it is more plausible to regard the association as

⁴ D. Wiggins, 'Identity, Designation, Essentialism, Physicalism', *Philosophia* 5 (1975), p. 11.

constituting a bit of information, or misinformation, inherited from people who purported to use the name as an ordinary proper name, rather than as manifesting a general intention to use the name to refer to whoever in fact satisfies the description. A natural example of a descriptive name will occur only when a name is introduced in connection with some description. Kripke mentions 'Jack the Ripper' and 'Vulcan' as examples of such names; another example might be 'Deep Throat', used as a name for whoever in the White House was the source of Woodward and Bernstein's Watergate-related information. Nevertheless, no matter how rare examples may be, it would appear always to be open to create descriptive names by stipulation. For example, we might stipulate:

(D) Let us use 'Julius' to refer to whoever invented the zip,⁵

and, governed by such a stipulation, 'Julius' would appear to have the properties of a descriptive name. For present purposes, it is not necessary to concern ourselves with the situation that would arise if the name became associated with other predicates as a result of discoveries made using the stipulation. We need only consider the simple case—the initial period during which the name is unquestionably a 'one-criterion' name.

It is fairly easy to see how the view expressed by Wiggins leads to the claim that ordinary proper names are 'Russellian'—if they have no referent, they have no sense. On that view, understanding an ordinary proper name requires knowing of the referent that the name refers to it, and this knowledge cannot exist in the absence of a referent. After all, if the knowledge is capable of being given in countless different ways, there must be something which unifies them, and this can only be the fact that they are all ways of identifying the same object; in the absence of an object this principle of unification, and hence the single piece of knowledge, does not exist. But, given the close semantical connection between a descriptive name and a description, no such problem arises,

⁵ In a slight regimentation, I shall take the English quantifier 'whoever' 'whatever' etc. to be free of existential commitment. By 'zip' I mean 'zip fastener'.

and it is plausible to hold that such names are 'Fregean'—they have a sense whether or not they have a referent.^{5a} It is sufficient to understand 'Julius' that one know that it refers to whoever invented the zip. This knowledge can certainly be possessed whether or not there is such a person, and possessing it, one is in a position to know exactly what conditions have to be satisfied for sentences containing the name to be true, and hence to understand them.

The only argument I know against the view that names introduced by description are Fregean was presented to me by Kripke. He supposed, I think correctly, that if one held that a speaker says something by uttering the sentence 'Julius is F' when the name is empty, then no better account could be given of what he said than this: he said that the man who invented the zip is F. But, Kripke argued, we cannot in general suppose that, when 'a' is a name whose reference is fixed by the description ϕ , that someone who utters the sentence 'a is F' says that the ϕ is F, since the statement that the ϕ is F may have different *modal* properties from the statement that a is F. I agree that sentences containing names embed differently under modal operators than do sentences containing descriptions, but it is perhaps the main point of this paper that the conclusion which Kripke draws from this fact follows only upon a questionable view of the connection between the content of an utterance and its modal properties. This important matter must be held over until Part III.⁶

^{5a} I no longer think the term 'Fregean' is appropriate for names which have a sense whether or not they have a referent, for reasons which I explain in my paper 'Understanding Demonstratives' in *Meaning and Understanding*, eds H. Parret and J. Bouveresse (Berlin and New York: W. de Gruyter, 1981), pp. 280–303 [and reprinted as Chapter 10 in this volume. Ed.].

⁶ As Kripke's inclination to argue in this way makes clear, my explanation of what it is for a name to have its reference fixed by description is almost certainly not the one intended by Kripke. I have made 'having a reference fixed by description' a semantical property of a word in a public language, whereas Kripke prefers to speak of a particular person, not necessarily the introducer of a name, fixing its reference by description. (I am grateful to John Dolan for reminding me of this.) It is a consequence of this that we cannot say that such-and-such a statement—here using a sentence in a public language—is knowable *a priori*, but only that such-and-such a sentence is *a priori* for X. If names like 'Julius' are Russellian, we will not even be able to state what X knows in a public language. Further, a sentence may be *a priori* for X, when it is not even true, namely when the name is empty. I am not sure what Kripke means by these relativized notions; they

Russell held that there could not be a referring expression with the properties we are taking 'Julius' to possess. He wrote:

Whenever the grammatical subject of a proposition can be supposed not to exist without rendering the propositions meaningless, it is plain that the grammatical subject is not a proper name, i.e. is not a name directly representing some object.⁷

Precisely because he held that sentences containing definite descriptions had determinate truth conditions whether or not the description was proper, he concluded that definite descriptions were not referring expressions, and he would surely have drawn the same conclusion about 'Julius'. Russell thought that it was a consequence of using the relation of reference to state the semantical contribution which an expression makes to sentences containing it that, in the event that the expression has no referent, those sentences would be deprived of truth conditions, and thus 'meaningless'. I think that Russell was wrong about this, but it is not a foolish view, and it is important for us to assure ourselves that there can be expressions of the kind the puzzle supposes names like 'Julius' to be. I feel this obligation particularly strongly since I agree with Russell, though not for his reasons, that definite descriptions are not referring expressions, and I also agree with him that other members of the category of referring expressions, to which I wish to regard 'Julius' as belonging, are such that, if they are empty, sentences containing them have no truth conditions. We must have at least some idea of what conception of reference includes 'Julius' as a referring expression, despite the fact that it is Fregean, and most referring expressions are not, and excludes 'the inventor of the zip', despite the very close similarity between it and the name 'Julius'.

seem to come to something like: *what X means by S* or, *the belief X expresses by S*, is *a priori* true. If this is the case, then I do not think that the puzzle can be formulated, since I hold that it makes sense to say that what X means, or what X believes is contingent only relative to some way of expressing that belief in a public language. Since I see no incoherence in the idea of a descriptive name, I shall avoid Kripke's relativized notions.

⁷ B. Russell, *Principia Mathematica* (2nd edn) (Cambridge: Cambridge University Press, 1968), p. 66.

Reference may be regarded as whatever relation it is between expressions and objects which makes the following principle true:

- (P) If $R(t_1 \dots t_n)$ is atomic, and $t_1 \dots t_n$ are referring expressions, then $R(t_1 \dots t_n)$ is true iff \langle the referent of $t_1 \dots$ the referent of $t_n \rangle$ satisfies R .⁸

Satisfaction is also whatever relation makes (P) true; (P) simultaneously and implicitly defines reference and satisfaction in terms of truth. (P) invites the semantic theorist to identify a class of atomic sentences in which he can discern expressions of two characteristic types, and to deal with expressions of each type by means of two different semantic relations which fit together, according to (P), to yield the truth conditions of those sentences. This is all you know, and all you need to know, by way of a *definition* of reference, or of satisfaction. Then, it is natural to regard as a referring expression any expression whose semantic contribution to the sentence in which it occurs is stated by means of the relation of reference which is found in (P). Now, we should expect Russellian singular terms to have their semantic contribution stated in clauses such as:

- (1) The referent of 'John' = John.

When there is no referent, no such clause can truly be stated, so that truth conditions for sentences containing the term cannot be derived: this is a formal representation of the fact that nothing is said by one who utters a sentence containing the term.⁹ But it is not necessary that a clause for a referring expression should take this simple form. It is equally true that a clause like:

- (2) (x) (Refers to ('Julius', x) $\equiv x$ uniquely invented the zip)

⁸ The status of the description 'the referent of t_i ' is *sub judice*, but to understand (P) you need only your working mastery of your language. A more realistic principle would need to take account of context dependence, but since our official business in no way depends upon context dependence, I have suppressed the complexities which stem from that source.

⁹ For some of the consequences of using clauses like (1) in the theory of meaning, see J. H. McDowell, 'On the Sense and Reference of a Proper Name', *Mind* 36 (1977), pp. 159-85.

uses only the relation of reference which is found in principle (P); taken together with the normal satisfaction clauses for atomic predicates, such a clause will enable us to derive truth conditions for sentences containing 'Julius' of the form:

- (3) 'Julius is F' is true iff the man who invented the zip is F.¹⁰

Neither (2), nor anything used in the derivation of (3), presupposes the existence of a referent for 'Julius'. Hence, according to this theory, 'Julius' is Fregean.

The truth condition stated in (3) is not homophonic, but this is quite inessential to the approach, once we suppose the metalanguage to contain a name with the same sense as 'Julius'. However, at this point we must pay attention to an important logical consequence of the semantic status which we are in the course of securing for 'Julius'. It is a consequence of the fact that 'Julius' is Fregean that, when it is empty, not only are atomic sentences containing it significant (though not true) but there may also be *complex* sentences containing the name which are *true*. For example, unless we specifically prevent it, the general rule for sentential negation will apply to the sentence 'It is not the case that (Julius is F)'; and, since it states that a negated sentence is true iff the embedded sentence is not true, the result will be that the given sentence will be determined as true when the name 'Julius' is empty. A similar point applies to any truth-functionally complex sentence containing an atomic constituent of the form 'Julius is F'; given other constituents with suitably chosen truth values, such complex sentences may be true. It is therefore an immediate consequence of the recognition of a name like 'Julius' that we must either modify the classical clauses for the truth-functional connectives, or we must modify classical logic—specifically by

¹⁰ I contend that this statement of the truth conditions of 'Julius is F' is perfectly acceptable as a statement of its meaning or content, despite the fact that the sentence used on the right hand side embeds differently inside modal operators than the sentence quoted on the left. A theory incorporating such a theorem certainly need not misstate the truth conditions of any modal sentence, and as to content, I hold that the two sentences do have the same content, despite their modal differences. This is defended in Part III.

restricting the rules of Existential Generalization (EG) and Universal Elimination (UE). Unrestricted, the rule of EG would enable us to pass from a true premiss, such as $\neg(F(\text{Julius}))$ to what might be a false conclusion, $(\exists x) \neg F(x)$, when the name 'Julius' is empty. Similarly, the rule of UE would enable us to pass from a true premiss, such as $(x) (x = x)$, to a false conclusion: 'Julius = Julius' when the name 'Julius' is empty. Logics with the required restrictions are well known under the title of Free Logics.¹¹ I shall not go into details here. Since it will be convenient to allow names like 'Julius' to take wide, as well as narrow, scope, I shall suppose that we are working with a language with explicit scope indicators, and a rule of Existential Generalization which is sensitive, at least for these names, to their scope. Thus, I shall follow Russell's square bracket scope-indicating device, and distinguish:

[a] $(P \vee F(a))$

from:

$P \vee [a] (F(a))$,

and so on, with the logical theory permitting $(\exists x)A(x)$ to be inferred only from $[a]A(a)$.¹²

With this background, we can see that there is no obstacle to using the name 'Julius' to state its own semantic contribution, in the way which is characteristic of homophonic theories:

(4) (x) (Refers to ('Julius', x) \equiv [Julius] ($x = \text{Julius}$)).

Using (4), we will be able to derive homophonic truth conditions for sentences containing the name 'Julius', but since the name has narrow scope in (4), and in the resulting

¹¹ For an excellent text, see R. Schock, *Logics Without Existence Assumptions* (Stockholm: Almqvist and Wicksell, 1968).

¹² Explicit scope-indicating devices are not necessary; in their absence, descriptive names would always be regarded as having narrowest scope, and wide-scope readings would be stated using the device: $(\exists x) (x = a \ \& \ \dots \ x \ \dots)$. See S. Kripke, 'Is there a Problem about Substitutional Quantification?', in G. Evans and J. H. McDowell (eds), *Truth and Meaning* (Oxford: Clarendon Press, 1976), pp. 373-4.

statement of truth conditions, the semantic theory is not itself committed to the existence of a referent.¹³

We are half-way towards answering our question. We have an account of what unifies 'Julius' with other referring expressions, despite its difference from them. We have not yet explained why definite descriptions should not be regarded as referring expressions. There is no formal obstacle to treating descriptions in clauses similar to (2), and such theories have been constructed.¹⁴ Since there are indefinitely many descriptions, we do not expect a clause in the theory of meaning giving the reference of each one. Rather, the theory would contain a recursive principle along the lines of:

(5) $(\phi) (x)$ (Refers to ('the' $\widehat{\phi}$, x) \equiv Satisfies (ϕ, x)),¹⁵

from which indefinitely many such statements of reference can be derived. Such a theory enables us to deduce a truth condition for sentences containing a description even when it is improper. Consequently, if the only objection to regarding descriptions as referring expressions stems from Russell's observation that, by uttering a sentence containing an improper description, one makes a perfectly intelligible move in the language game, then it would appear to have been met. However, it is not the only objection, for if we look at matters more closely, it becomes clear that a principle like (5) is not adequate to explain the behaviour of descriptions in all contexts.

The feature of the behaviour of descriptions which cannot be captured in this way comes out most clearly in modal contexts, and in order to explain it, we must work with some semantic theory adequate to deal with those contexts. Since possible-worlds semantic theories are both familiar and easy to work with, I shall put my points in their terms, but they

¹³ It is not necessary to get involved here in the details of the truth theory; for some of them, see T. Burge, 'Truth and Singular Terms', *Noûs* 8 (1974), 309-25. Burge applies the same treatment to descriptions, which I do not want to do. His complicated restricted identity-substitution principle, (A8), seems unnecessary; the derivation of truth conditions need only exploit the extensionality of the description operator.

¹⁴ See, e.g., T. Burge, *ibid.*

¹⁵ I use underlining as a uniqueness operator. Thus: ' $A(\tau_1 \dots \tau_i \dots \tau_n)$ ' abbreviates ' $A(\tau_1 \dots \tau_i \dots \tau_n) \ \& \ (x) (A(\tau_1 \dots x \dots \tau_n) \supset x = \tau_i)$ '.

should be capable of translation into any semantical framework. I certainly do not wish my use of possible-worlds semantics to be taken to indicate either that I believe it to be the correct semantical framework for modal sentences of natural languages, or that I believe it to be immune to philosophical objection.¹⁶

A possible-worlds semantic theory states the truth conditions of sentences of a language which contains modal operators in a metalanguage which dispenses with such operators in favour of explicit quantification over possible worlds. For each n -place predicate, R , of the object language, there is in the metalanguage an $n + 1$ place predicate, R' ; the additional argument-place being occupied by terms referring to, and variables ranging over, possible worlds. These predicates are connected *via* satisfaction clauses of the form:

$$(x_1) \dots (x_n) (\text{Satisfies}_w('R', \langle x_1 \dots x_n \rangle) \equiv R'(x_1 \dots x_n w))^{17}$$

For example, an object satisfies 'Bald' *with respect to* a world w iff it is bald *in* w . The theory is so constructed that we are able to derive, for each sentence S , a theorem of the form:

$$(w) (\text{True}_w(S) \equiv \dots)$$

with which the clauses for the modal operators connect in the familiar way:

$$(S) (w) (\text{True} ('It is possible that' \widehat{S}) \equiv (\exists w') (\text{Alt}(w, w') \& \text{True}_{w'}(S)))$$

('Alt' is some suitable 'alternativeness relation' defined over the set of possible worlds; its properties need not concern

¹⁶ For an alternative approach to the semantics of modality, see C. A. B. Peacocke, 'Necessity and Truth Theories', *Journal of Philosophical Logic* 7 (1978).

¹⁷ I subscript the satisfaction relation thus: 'Satisfies_w', and similarly with 'true_w', partly for ease of comparison with classical clauses, but also to emphasize that these relativized semantical relations are not got by the same process that gives us 'F' (x, w) from 'F(x)'. The statement that a is bald in w ('Bald' (a, w)) can be understood as equivalent to the simple counterfactual: if w had been actual, a would have been bald, but the statement that a certain statement is true with respect to a world ('True_w(S)) is not equivalent to the simple counterfactual: if w had been actual, S would have been true. See below.

us.) A sentence is true *simpliciter* iff it is true_{w*}, where 'w*' refers to the actual world.

In the context of this semantic theory, the principle (P) must be modified to connect reference with the notion true_w.

$$(P') \text{ If } R(t_1 \dots t_n) \text{ is atomic, and } t_1 \dots t_n \text{ are referring expressions, then } R(t_1 \dots t_n) \text{ is true}_w \text{ iff} \\ \langle \text{the referent of } t_1 \dots \text{ the referent of } t_n \rangle \text{ satisfies}_w R.$$

(From this principle, the principle (P) can be derived as a special case.) Once this change has been made, no other change needs to be made. Even in a modal language, all that is necessary to state the significance of names and other referring expressions is to state to what, if anything, they refer; the truth-with-respect-to-a-situation of a sentence containing a singular term depends simply upon whether or not its referent satisfies the predicate with respect to that situation. But, notoriously, this is not the case with definite descriptions. If we assign them a reference by means of a principle like (5), and connect this assignment with truth by means of (P'), if, in short, we treat them like other referring expressions, we capture only one of the readings of a sentence like:

The first man in space might have been an American,

namely that on which it is equivalent to the claim that Gagarin might have been an American. The whole sequence would be determined as true iff there is a possible world with respect to which the referent of the description, i.e. Gagarin, satisfies 'American'. In order to capture the other reading of this sentence, on which it is true iff there is a possible world in which the man who is first in space in *that* world is an American (in that world), some changes have to be made.

The only way of making those changes, while still attempting to treat descriptions as referring expressions, is by relativizing the relation of reference to a possible world. (5) must become:

$$(6) (\phi) (x) (w) (\text{Refers to}_w('the' \widehat{\phi}, x) \equiv \text{Satisfies}_w(\phi, x))$$

and (P') must become:

(P'') If $R(t_1 \dots t_n)$ is atomic, and $t_1 \dots t_n$ are referring expressions, then $R(t_1 \dots t_n)$ is true_w iff
 (the referent_w of $t_1 \dots$ the referent_w of t_n) satisfies_w R.

This can be done.¹⁸ But it is at a high price, due to the fact that we must relativize the relation of reference in all cases. Simply in order to assimilate descriptions to other referring expressions, we introduce a major change in the semantic apparatus in terms of which we describe the functioning of those other expressions. As a consequence of this change, we ascribe to names, pronouns, and demonstratives semantical properties of a type which would allow them to get up to tricks they never in fact get up to; since their reference never varies from world to world, this semantic power is never exploited.

A similar point can be made when we take account of the existence of ambiguities which definite descriptions generate in tensed sentences, like:

The leader of the Conservative Party will be courageous.

To deal with this, essentially similar, ambiguity, the relation of reference must also be relativized to a time. Once again, this enrichment of the type of semantical assignment made to singular terms is unnecessary for all terms other than descriptions. Finally, the fact that a position inside a description can be bound by a higher quantifier, as in the sentence:

The father of each girl is good to her,

requires a relativization of the relation of reference to a sequence, or an assignment, π , to the empty singular term positions which the description may contain.¹⁹

¹⁸ Has been done, by, e.g., R. Thomason and R. Stalnaker, 'Modality and Reference', *Noûs* 2 (1968). 'An expression like " $(\exists x)\phi(x)$ " is assigned a referent which may vary from world to world'. On such a view, the first reading considered is captured by giving the descriptive singular term wide scope with respect to the modal operator.

¹⁹ This point is made in B. Mates's paper, 'Descriptions and Reference', *Foundations of Language* 10 (1973), pp. 409-18. In fact, the objection can be dealt with if one uses 'Fregean' rather than 'Tarskian' treatments of the quantifiers; for this distinction, see my paper 'Pronouns, Quantifiers and Relative Clauses (I)', *Canadian Journal of Philosophy* 7 (1977), 471-7. [Reprinted as Chapter 4 in this volume. Ed.] It should be stressed that the relativity to a time

Thus, if we are to include definite descriptions in the category of referring expressions, we are forced to describe the behaviour of all the members of that category in terms of the relation 'Refers to_{w,t,\pi}(τ, x)' rather than the simple relation 'Refers to (τ, x)' which is otherwise perfectly adequate. This certainly does not constitute a knock-down argument against treating descriptions as referring expressions, but it does rather strongly suggest that the grouping that results on this treatment may not correspond to any natural, semantical, kind. This case is strengthened when we look at the other side of the story, and examine alternative approaches to descriptions. One approach looks especially promising. Every semantical theory must recognize the category of quantifiers, members of which occur in the sentences 'Every ϕ is ψ ', 'Some ϕ is ψ ', 'No ϕ is ψ ', etc. We find that if we suppose the sentence 'The ϕ is ψ ' to be built up in exactly the same way out of exactly the same type of semantic elements as the quantified sentences which it so closely resembles, we achieve a remarkably good fit with the behaviour it is observed to display. Even on a theory which attempts to treat descriptions as singular terms, we could introduce, with no complication of theory, a quantifier 'The', and the resulting sentences would be indistinguishable from those containing the supposed singular term. Such a theory could then be considerably simplified if it made do with just the quantifier 'The', and allowed the relation which deals with referring expressions to revert to its simple, unrelativized form. In other disciplines, such a consideration would strongly recommend the resulting theory, and I am not sure why the theorist of meaning should be unmoved by it.²⁰ In what follows, I shall treat 'The' as a quantifier;

required to deal with the temporal flexibility of descriptions is quite different from that introduced by context-dependence, so that, treating descriptions as referring expressions imposes on the relation of reference a double relativity to a time.

²⁰ I discuss the aspect of the methodology of semantic theories which I am here relying on in my paper 'Semantic Structure and Logical Form', in Evans and McDowell, op. cit. For those who think that the proposal to treat 'The' as a quantifier need be accompanied by the butchering of the surface structure of English in which Russell so perversely delighted, see the treatment of quantifiers in my paper 'Pronouns, Quantifiers and Relative Clauses (I)', op. cit.

specifically, as a binary quantifier, taking two open sentences to make a sentence. 'The ϕ is ψ ' is therefore formalizable as '(Ix) ($\phi(x)$; $\psi(x)$)', which I shall suppose to be provably equivalent to the Russellian expansion ' $(\exists x) (\phi(x) \& \psi(x))$ '.²¹

I have assumed, in the argument I have just presented, that proper names, pronouns, and demonstrative expressions function as 'rigid designators'. I shall not defend this claim here; I shall only remark that it seems to be a fairly strongly marked feature of even such a name as 'Julius' that we do not use it so that the following comes out true:

If you had invented the zip, you would have been Julius,

If Julius had never invented the zip, he would not have been Julius.

However, if we adopt the recommendation of the previous paragraph, the behaviour which Kripke labels 'rigid designation' emerges as simply that of designation. A referring expression does not designate the same thing with respect to each possible situation; it simply designates, and the truth value of any sentence containing it depends upon what, if anything, it designates. The term 'rigid designation' carries with it the suggestion of 'non-rigid designation', and hence only really belongs in a theory in which the designation/reference relation is relativized and used in the treatment of both names and descriptions.

I began this line of thought with a doubt about the very possibility of a Fregean referring expression. I have tried to show that this doubt is groundless. If a name like 'Julius' is treated in clauses like (2), or (4), it will be Fregean, while if the relation of reference used in those clauses is connected to truth_w via the principle (P'), it will behave as a 'rigid designator'. With this background, we can now consider the puzzle.

II

Given that the name 'Julius' is introduced into the language by means of the stipulation (D), it appears that someone can know that the sentence:

²¹ I have with these arguments finally come round to the position urged on me some years ago by Mr M. K. Davies.

(S) If anyone uniquely invented the zip, Julius invented the zip

is true, simply in virtue of knowledge he has as a speaker of the language. The problematical sentence is given this conditional form because the simple sentence:

(sub-(S)) Julius invented the zip

requires for its truth something which, it is supposed, (S) does not, namely that someone did uniquely invent the zip, and since this cannot be known *a priori*, neither can sub-(S).

At the same time, (S) appears to be contingent; there are possible worlds with respect to which it is false. Because 'Julius' is a referring expression, the truth with respect to a world of the consequent, sub-(S), requires the satisfaction with respect to that world of the predicate 'invented the zip' by the referent of 'Julius'—i.e. the man who actually invented the zip, if there is such a man. Hence, a world in which someone who did not actually invent the zip invents the zip is a world with respect to which the antecedent of the conditional is true, but the consequent, and thus the whole conditional, is false.

In order to bring out the puzzlement which this combination of characteristics ought, at least initially, to invoke, let me quote a passage of Kripke's:

I guess it is thought that . . . if something is known *a priori* it must be necessary, because it was known without looking at the world. If it depended upon some contingent feature of the actual world, how could you know without looking? Maybe the actual world is one of the possible worlds in which it would have been false.²²

And, in a recent paper devoted to the puzzle, Donnellan poses the problem in similar terms:

If a truth is a contingent one then it is made true, so to speak, by some actual state of affairs in the world, that, at least in the sort of example we are interested in, exists independently of our language and linguistic conventions. How can we become aware of such a truth, come to know the existence of such a state of affairs, merely by performing an act of linguistic stipulation?²³

²² Kripke, op. cit., p. 263.

²³ K. S. Donnellan, 'The Contingent *A Priori* and Rigid Designators', *Midwest Studies in Philosophy*, 2 (1977), p. 13.

Donnellan attempts to solve the puzzle by showing that the problematical statement is not both contingent and capable of being known *a priori*. He summarizes his strategy as follows:

I am going to invoke a distinction between knowing that a sentence expresses a truth and knowing the truth of what is expressed by the sentence. I am going to suggest that, as a result of the introduction of a name as a rigid designator by means of a description fixing the referent, we can come to know, perhaps even *a priori*, that certain sentences express truths, but we do not come to know *a priori* the truth of what they express.²⁴

There undoubtedly exists a distinction between knowing that a sentence is true and knowing the truth it expresses—the passage from one bit of knowledge to the other is mediated by knowing what the sentence means. It is Donnellan's contention that someone who knows only the reference-fixing definition (D) does not understand the name 'Julius', nor sentences containing it. According to Donnellan, to understand the name, one must know *of* some object that it is the referent of the name, when such knowledge *of* an object requires a causal connection with it. (This last, causal, feature is inessential to his general strategy, which works provided any *a posteriori* knowledge is made a precondition of understanding the name. It is consistent with the strategy to adopt a criterion of 'knowledge *of*' according to which one knows *of* the shortest spy that he is a spy simply by knowing that the shortest spy is a spy; it remains the case that understanding the name will require *a posteriori* knowledge — that there exists an inventor of the zip—and so that one cannot have *a priori* knowledge of what (S) says.)²⁵

This way of solving the puzzle obliges us to make sense of the idea that an expression can have a meaning in a language even though no past, present (and, possibly, future) speaker of that language knows what its meaning is. As

²⁴ Ibid. p. 18. As he acknowledges, Donnellan's solution to the puzzle follows that of A. Plantinga, *The Nature of Necessity* (Oxford: Clarendon Press, 1974), pp. 8–9, n. 1, and M. Levin, 'Kripke's Argument against the Identity Thesis', *Journal of Philosophy* 72 (1975), p. 152, n. 2.

²⁵ This option is, in essentials, taken in S. Schiffer's paper 'Naming and Knowing', *Midwest Studies in Philosophy* 2 (1977), pp. 28–41. See the passage cited on p. 198 below.

Donnellan himself admits, this is not an idea with which we feel very comfortable. But apart from any reservations we may have about this idea in general, there is room for doubt about the use Donnellan makes of it to solve the puzzle. I shall make three points: one *ad hominem*, one substantive, and one promissory. First, the solution to the puzzle seems to me to take a needlessly complicated form. Given the strength of the premisses from which Donnellan proceeds, there is open to him, and indeed he is really forced to take, a much, much shorter way with the puzzle. This comes clear when we take note of an obvious, but previously unmentioned, point: *there simply is no puzzle unless the use of free logic is accepted*. Unless a sentence containing the name 'Julius' can be formulated which is free of existential commitment, there is not even a candidate for the status of the contingent *a priori*, but within a classical framework, there are no such sentences. No matter how a name may be embedded in a sentence, in a classical language that name is accessible to existential quantification, and the truth of the whole sentence requires that the name refer. Since it is quite pointless, within a classical framework, to make the problematical sentence conditional in form, it must be presumed to be the intention of anyone offering these conditional sentences that they be taken to be sentences governed by a free logic in which the names take narrow scope. Using our earlier notation, and an otherwise obvious symbolization, (S) must be intended to be understood as having the form:

$$(\exists x) (\phi(x)) \supset [a] \phi(a)$$

It is a presupposition of the use of a free logic that there exist Fregean names, and hence a presupposition of the usual formulation of the puzzle, that names like 'Julius' be Fregean names. But Donnellan is prepared to impose upon understanding the name 'Julius' conditions which preclude it from being Fregean—the knowledge which Donnellan requires for really understanding the name cannot be in anyone's possession if the name is empty. Given that he was prepared to impose these requirements, I cannot see why he played along with the conditional form of (S), rather than exposing it as

pointless, and I cannot see why he did not dismiss the supposed puzzle in a paragraph, stating that an unrestricted rule of Existential Generalization applies to any proper name, and that, for his part, he cannot see what all the fuss is about.

Anyway, both the position which Donnellan advanced, and this brusquer version of it, depend upon the claim that there cannot be descriptive names, and this brings me to my substantive point: I do not think that this is true. Donnellan does not advance any reason for his claim that a knowledge of (D) is not sufficient for understanding the name 'Julius'. On this crucial point, as though appealing to an unchallengeable datum, he says simply:

It is rather . . . that as these stipulations introduce names, they give the names no descriptive content.²⁶

This will not do. It is one thing to hold that ordinary proper names have no descriptive content, quite another to hold that there can be no such thing as a descriptive name. Attempting the only demonstration to which these matters are susceptible, I have tried to show how a theory of meaning can be constructed which treats 'Julius' as a Fregean referring expression, and Donnellan gives no reason for thinking that this demonstration is defective.

By omitting to give reasons for one's views, one runs the risk that possibly quite erroneous speculations will be made as to what those reasons might be. I am going to suggest a line of thought which may have influenced Donnellan, though it must be stressed that there is only slight evidence of it in his paper. Suppose that x in fact invented the zip, and hence, is the referent of 'Julius'. Now, let us ask what makes a sentence like 'Julius is F ' true. The only answer appears to be x 's *being F*; the existence of that state of affairs in any world is a necessary and sufficient condition of the truth of 'Julius is F ' with respect to that world. No state of affairs of the type y 's *being the inventor of the zip and y 's being F* when $y \neq x$ can make the sentence true,

²⁶ Donnellan, *op. cit.*, p. 21. Donnellan's claim that names like 'Julius' have no descriptive content is particularly perplexing in the light of his earlier statement that such names are 'pegged to' descriptions. He also strongly suggests that the traditional description theory of names holds good of these names. See p. 17 of Donnellan.

since it is not true with respect to a world incorporating any such state of affairs. Equally, it is not x 's *being the inventor of the zip and being F* which makes the sentence true, since x 's *being F* is enough to make the sentence true with respect to a world even if x is not the inventor of the zip in that world. (So much is simply to restate the rigidity of 'Julius'.) Consequently, if 'Julius is F ' is true, what makes it true is the existence in the actual world of the state of affairs of x 's *being F*. Now it seems reasonable to hold that someone who understands the sentence will know what state of affairs makes it true. In that it is made true by x 's being in such-and-such a condition, the sentence is *about* x , and hence no one who is not aware of the connection between 'Julius' and x can understand the sentence, for he will not know what it is about. Since knowledge of (D) makes one aware of no such connection, knowledge of (D) is not sufficient for understanding the name.

I mention this argument now, though I will not try to explain what I think is wrong with it until the next section. It will suffice at this point to observe that the argument rests upon a connection between the content of a statement and its modal properties which no one who holds that there can be descriptive names has any reason to accept, and which there are good, independent reasons to reject. Those independent reasons stem from the fact that there are other sentences which have, at least the appearance of being both contingent and *a priori*; a solution to the conundrum in their case, and, I would argue, in all cases, depends upon rejecting the connection between content and modality upon which this argument turns. And this brings me on to my third, promissory point. Although I shall not give examples until Part IV, I claim that there exist examples of the contingent *a priori* which do not even appear to be amenable to treatment along Donnellan's lines. With these examples, it is out of the question to hold that those who are in a position to know that they are true *a priori* do not really understand them. This should make us even more reluctant to deny that there can be descriptive names, since it shows us that such a denial is not even part of a generally effective strategy on the puzzle we are trying to solve.

In what follows, I shall assume that a knowledge of (D) is sufficient for understanding the name 'Julius'. Donnellan may point out that, on such an assumption, (S) certainly does not say anything informative about the world, and this is undoubtedly correct. But saying this merely serves to identify the right puzzle, and does not solve it. The puzzle is not how we can know *a priori* something informative or interesting about the world, but how we can know *a priori* something contingent, and hence, how something contingent can be uninformative.

I have stressed that the puzzle does not even get off the ground unless we are allowed to formulate (S) in a free logic, and in a way which gives the name 'Julius' narrow scope. It may be thought that, understood in this way, (S) is no longer contingent. This seems to be the position of Stephen Schiffer, who wrote the following brief remarks on the puzzle. (I have altered the passage merely by substituting my example for his.)

Kripke cannot have it both ways. If (2) ['Ralph knows that Julius invented the zip if anyone did'] attributes to Ralph knowledge of a proposition that is contingent because the person who in fact invented the zip might not have, the 'Julius' in (2) refers to that person, and (2) attributes to Ralph *de re* knowledge of that person—specifically the knowledge that that person invented the zip if anyone did. Now I do not mind saying that the inventor of the zip is known by Ralph to have invented the zip, and if Ralph can reason a little, to have invented the zip if anyone did. But if Ralph has this knowledge, he has it just by virtue of knowing that some one person invented the zip, and not at all by virtue of having named the inventor of the zip 'Julius'. Yet Ralph's knowledge that some one person invented the zip is certainly something that he has only *a posteriori*, and so, therefore, is his knowledge, of the inventor of the zip, that the person invented the zip—if Ralph has that knowledge.²⁷

This passage clearly implies that if there is a proposition which is contingent it must be existentially committing, and hence, not *a priori*. Contrapositively, if (S) is interpreted in such a way that it is not existentially committing, then it is not contingent. *But this is just not true*. The sentence in which the name takes narrow scope, is certainly contingent, at least in the sense that:

²⁷ Schiffer, *op. cit.*, p. 29.

$$\neg \Box [(\exists x) (\phi(x)) \supset [a] \phi(a)]$$

i.e.,

$$\Diamond [(\exists x) (\phi(x)) \ \& \ \neg [a] \phi(a)]$$

It is no doubt true that, if (S) were contingent *because of* the contingency of the possession by an individual of a certain, presumably conditional, property, then (S) would have to be interpreted in a way which gave the name wide scope, and Schiffer's way of solving the problem would become applicable. But I am not aware that Kripke stipulated the *source* of the contingency of his problematical sentence, and even if he had done so, the stipulation should be ignored, since it is sufficient to generate the puzzle that the sentence be contingent, never mind how or why it is contingent. It is true that the contingency of (S) crucially depends upon the fact that 'Julius' is a referring expression; if we replace the name with the description 'the inventor of the zip' in such a way that it too receives narrow scope, then the sentence is no longer contingent. But it does not follow from the fact that a sentence is contingent because it is formulated with the use of a referring expression, that its contingency is due to the contingent possession of a certain property by the object to which that expression refers.

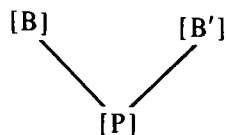
The puzzle presented by (S) is a little more resistant to solution than it might first appear. The name 'Julius' can be understood by anyone who knows the reference-fixing definition, so a solution along Donnellan's lines is ruled out, while at the same time, simply observing that this is so does not dissolve the puzzle, for even when 'Julius' is interpreted as a descriptive name with narrow scope, (S) remains both contingent and knowable *a priori*.

III

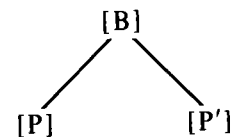
As we have seen, a modal semantics associates with each sentence conditions under which it is true with respect to an arbitrary possible world. The notion 'true_w' is specifically designed to account for the way sentences embed inside modal contexts. If $\Diamond(Q)$ is true, for example, then the true_w relation for (Q) must be so characterized that there

exists a world w such that (Q) is true _{w} . If we wish, we can think of the true _{w} theorems as associating with each sentence a *function* from possible worlds to truth values, and it is customary to call such functions 'propositions'. As a preliminary to solving the puzzle presented by (S), we must get clear about the relation which exists between the *proposition* associated with a sentence and its *content*, or *what it says*. I shall not attempt to give an analysis of the notion of content here; I want to rely upon the intuitive sense according to which, if two sentences have the same content, then what is believed by one who understands and accepts the one sentence as true is the same as what is believed by one who understands and accepts the other sentence as true. On this, very strict, view of sameness of content, if two sentences have the same content, and a person understands both, then he cannot believe what one sentence says and disbelieve what the other sentence says. When two sentences meet this condition, I shall say that they are epistemically equivalent. Naturally, great difficulties arise when one attempts to apply the criterion in ordinary cases, but we shall be dealing exclusively with rather extraordinary cases—with 'one-criterion' words.

It is immediately evident that two sentences which express the same proposition can have different contents; after all, any two necessary truths are associated with the same function from worlds to truth values, as are the two sentences 'F(a)' and 'F(b)' when 'a = b' is true. However, I think that it has generally been assumed that, if two sentences express different propositions, they must have different contents, or say different things, and this is the view I want to challenge. Just as we accept the situation which we may represent as follows:



in which distinct statements are modally indistinguishable, so we must accept the situation:



in which epistemically equivalent sentences are modally distinguishable. This may sound very grand, but it amounts to nothing more than this: that sentences with the same content might embed differently inside the scope of modal operators.

There is a position concerning negation and falsity which is parallel to the one I want to adopt, and although I am not at all sure I want to hold it, I think it is sufficiently familiar, and clear, to cast light upon the position I do want to hold. Upon at least one interpretation of Strawson's reply to Russell, he held that the sentence 'The ϕ is not F' is the unambiguous result of applying a negation operator, which I shall symbolize as 'N', to the sentence 'The ϕ is F', and that the sentence 'N(The ϕ is F)' is not true when the description is improper. However, Strawson pointed out, when 'N' is applied to the Russellian expansion 'Something is uniquely ϕ and it is F', the resulting sentence is true. Hence, 'The ϕ is F', and its Russellian expansion, embed differently under the operator 'N'. If one holds this view, one can continue to maintain that 'N' is a truth functional operator by adopting a three-valued logic, in which 'The ϕ is F' gets the value 1/2 ('neither true nor false') when the description is improper, and 'N' maps the value 1/2 on to itself.

On this view, the two sentences are associated with different functions from states of the world into truth values in the set $\{1, 1/2, 0\}$, and one can imagine a use of the notion of 'proposition' rather similar to that in the modal case, according to which this difference would mean that the sentences express different propositions. Nevertheless, it would be a great mistake to conclude from the fact that they are associated with different propositions, that the two sentences have different contents, or say different things. To draw this conclusion would place one in the hopeless position of casting around for an account of what it is that one who understands and accepts the sentence 'The ϕ is F'

believes, other than that something is uniquely ϕ and F. There is simply no such account; the two sentences are epistemically equivalent. One should not allow oneself to be so misled by one's semantic theory. The division, among non-designated values, into the strictly false, and the neither true nor false, is simply designed to register the differences in the ways sentences embed under 'N', and there is no reason why those differences should not be due to differences in the way the sentences are constructed which are consistent with an identity of content.²⁸

Whatever the merits of this position as an account of the way negation interacts with descriptions, it provides a parallel to the position I want to adopt in the case of modality. We know that the sentences 'Julius is F' and 'The inventor of the zip is F' are associated with different propositions—they have different true_w -conditions. Nevertheless, it seems clear that the two sentences are epistemically equivalent. Remember that we are interpreting the sentence 'Julius is F' in such a way that it is capable of being understood by one who knows only the reference-fixing stipulation (D), and that what it says must be capable of being specified in the absence of any referent. Given these conditions, I cannot imagine how the belief that Julius is F might be characterized which is not simultaneously a characterization of the belief that the inventor of the zip is F, i.e., that one and only one man invented the zip and he is F. Belief states are individuated by the evidence which gives rise to them, the expectations, behaviour, and further beliefs which may be based upon them, and in all of these respects the belief states associated with the two sentences are indistinguishable. We do not get ourselves into new belief states by 'the stroke of a pen' (in Grice's phrase)—simply by introducing a name into the language.²⁹

²⁸ My understanding of this position is due to Dummett's elaboration of it in Frege, Chap. 10. In fact, Strawson may have been arguing for the position which I believe he holds today, namely that the designation 'neither true nor false' is not an indication of a third truth value, but of a sentence's failing to say anything at all. However, it was precisely a reluctance to accept this that led Russell to formulate his theory of descriptions in the first place. I hope it is unnecessary to add that here and throughout when I speak of definite descriptions, I am concerned with 'pure' uses, when no link-up with antecedently existing identificatory knowledge is intended.

²⁹ I am ignoring what I take to be irrelevant complications which might arise

In view of this, it seems to be just as much a mistake, leading to just as fruitless a search, to argue from a difference of 'proposition' to a difference in content in this case, as it was seen to be in the previous case. Rather, we should accept that the two sentences are composed out of different parts of speech—a quantifier versus a name—and that this is a difference in their construction to which modal operators are sensitive even though it leads to no difference in content.

If we had asked those working with a modal semantics why they believed that the situation represented by the second diagram could not arise, I think they might have given the following reply.

'One cannot simply say that two equivalent sentences can embed differently inside modal contexts, and leave matters there. The absolute notion of truth, and the relative notion, truth_w , cannot be as independent as this casual talk encourages us to imagine. A modal semantics must validate the theorem $Q \vdash \diamond Q$, and this means that it is not only simpler, but theoretically indispensable, to be able to regard absolute truth as a special case of truth_w —i.e. as truth_{w^*} , where w^* designates the actual world. Only if there is this connection between the concepts will it follow from the fact that a sentence is (absolutely) true, that there is a world with respect to which it is true. Hence we cannot have two separate parts of the semantic theory, one giving sentences absolute truth conditions, and thereby their content, and the other assigning true_w conditions, and thereby explaining how the sentence will embed under modal operators. On the contrary, there is a single theory, assigning true_w conditions; the truth conditions, and the content of the sentence, must be derivable from that assignment. Hence, we must be able to regard each sentence as formulating a single property, or condition, which it demands of an arbitrary possible world—this is the function 'in intension'—and regard the content

if we take into account beliefs which arise about the name, or the stipulation. G. Harman argues in his paper 'How to Use Propositions', *American Philosophical Quarterly* 14 (1977), pp. 174–5, that the introduction of a 'mental name' by a stipulation such as (D) enables someone to have thoughts and beliefs about the referent although his previous beliefs would have been merely existential or general. I cannot see any point in this suggestion other than the perhaps laudable one of utterly discrediting the notion of 'belief about'.

of a simple assertion made with the sentence as tantamount to the claim that the actual world satisfies this condition. Now, if two sentences are associated with different propositions, so that there are worlds with respect to which the one is true and the other is not, the associated properties which determine the functions from worlds to truth values must be different. How, then, could the two sentences have the same content, since one is tantamount to the claim that the actual world satisfies one property, and the other is tantamount to the claim that it satisfies a different property?

The considerations offered in this reply certainly preclude any large-scale, or even very interesting, detachment of content and proposition. Nevertheless, there is a flaw in the argument—a gap which it leaves open, and which all the examples of the contingent *a priori* exploit. Consider for a moment the properties determined by the following two monadic predicate expressions:

(ξ) is as tall as John,

(ξ) is as tall as (ξ), or (ξ) is as tall as himself.

These are certainly different properties, in that there are objects which satisfy the second but which do not satisfy the first. But it is not correct to infer from this that, in all cases, the ascription of one property to an individual yields a statement with a different content from that which results from the ascription of the other property to that individual. When the two properties are ascribed to John, the results are the same, namely:

John is as tall as John.

Correspondingly, the argument advanced above will be fallacious just in the cases in which, to formulate the property of an arbitrary world which is associated with a sentence, one must make reference to what is actually the case. If a sentence is associated with such a property, then the way is open for it to be epistemically equivalent to a sentence which is associated with a distinct (non-co-extensive) property, and hence for sentences which have the same content to be associated with different propositions. In the case of 'Julius

is F' and 'The inventor of the zip is F', this is exactly what we find. The former sentence requires of a world w that the man who invented the zip *in the actual world* be F in w , while the latter sentence requires of a world w that the man who invented the zip in w be F in w . Schematically the properties are:

$$\lambda w [(Ix) (\phi'(x, w^*); F'(x, w))]$$

and

$$\lambda w [(Ix) (\phi'(x, w); F'(x, w))]$$

and these are certainly distinct. But the two sentences are epistemically equivalent, because, when applied to the actual world, they yield the same result:

$$(Ix) (\phi'(x, w^*); F'(x, w^*)).$$

I conclude from this that the notion of epistemically equivalent sentences being associated with different propositions is a coherent one; that it has application to certain pairs of sentences in natural language, and that the sentences 'Julius is F', and 'The inventor of the zip is F' are such a pair.

If this conclusion is correct, it follows that we must use the notion of *what makes a sentence true* with a great deal of care, for, if it is used without care it will collapse the distinction between content and proposition which we have been at pains to describe. There are two ways in which one can use the notion of what makes a sentence true. One can follow traditional practice, as I shall do, and tie the notion to the content of a sentence, so that if two sentences are epistemically equivalent, they are verified by exactly the same state of affairs, and what one believes, in understanding the sentence and accepting it as true, is precisely that some verifying state of affairs obtains. On this conception, the same set of states of affairs makes the sentence 'Julius is F' true as makes the sentence 'The inventor of the zip is F' true. If $x, y, z \dots$ is a list of all objects, then any member of the set:

{*x's being the inventor of the zip & x's being F; y's being the inventor of the zip and y's being F; z's being the inventor of the zip & z's being F . . .*}

will suffice to make either sentence true.³⁰

Alternatively, one can tie the notion of what makes a sentence true to the proposition it expresses by means of the principle: σ makes (Q) true iff, for any world w which comprises σ , (Q) is true _{w} . It was this conception of what makes a sentence true that we were using when we concluded on Donnellan's behalf that, if x in fact invented the zip, the state of affairs of *x's being F* alone made the sentence 'Julius is F' true. This is a perfectly legitimate way to use the notion of what makes a sentence true. What is not legitimate is to use the notion one way, and then slip into using it in the other way. Just such an illegitimate transition was made in the argument which we speculatively attributed to Donnellan. For that argument started from the claim that what makes 'Julius is F' true is *x's being F*—here relying upon the modal conception of what makes the sentence true—and concluded that the sentence is about x , conveys information about the existence of the state of affairs of *x's being F*, and hence, that someone who knew only the reference-fixing definition could not understand it.

When I use the notion of what makes a sentence true, I shall tie the notion to the sentence's content. But there is an ineliminable modal element in the notion of what makes a sentence true. For what can it mean to say that any one of a set of states of affairs renders a sentence true, other than that, if any one of them obtains, the sentence will be true, and that, if any one of them *had* obtained, the sentence *would have been* true? If this is so, to hold that the state of affairs of *y's being the inventor of the zip and being F*, for example, could serve to make the sentence 'Julius is F' true, would appear to commit one to the view that, had y invented the zip and been F, the sentence would have been true. But is this not inconsistent

³⁰ The problem presented by (S)—the fact that it is contingent—in no way depends upon the possibility that things might have existed which do not in fact exist, so there is no reason to quantify over possible but non-actual objects. ' $x, y, z . . .$ ' can here simply be a list of the objects that exist.

with the fact that the sentence 'Julius is F' is not true with respect to that situation?

I accept the counterfactual claim, but there is no inconsistency in doing so. The point only shows that 'true _{w} ' must not be glossed as 'if w were actual, . . . would be true'.³¹ But this is no matter; the notion 'true _{w} ' is purely internal to the semantic theory, and needs no independent explanation. The counterfactual claim is true, because, had y invented the zip, y would have been the referent of 'Julius', and if he had also been F, then the referent of 'Julius' would have been F, and nothing more is required for the truth of the sentence. Notice that this is not a trivial counterfactual claim about the truth of a sentence identified merely as a sequence of expression types. On the contrary, in the circumstances mentioned, the sentence 'Julius is F' would have been true *as a sentence of English*. The hypothesis that y invented the zip no more involves the hypothesis of a semantical change in English than the hypothesis of its having rained on a day which was in fact dry does. The hypothesis that the name 'Julius' refers to y would involve a semantical change in English only if the reference-fixing definition (D) had established a semantical connection between the name 'Julius' and some particular person other than y . But we have seen that this is not the case. Even if someone did invent the zip, (D) does not introduce a semantical connection between 'Julius' and that person. Neither (D), nor the clauses (2) and (4) based upon it, are existentially committing, and knowledge of them cannot constitute knowledge of a relation between 'Julius' and some item; one cannot know of the existence of a relation between two things, not even a semantical relation, without knowing that those things exist. (D) instituted a semantical connection between a name and a description as fixing its reference, and that connection is preserved unchanged under the supposition that y is the inventor of the zip. Exactly the same theory of meaning serves to describe the language which would be spoken had y invented the zip, as describes the language which is actually spoken.

We are now in a position to reply to the argument of

³¹ This is why it is vital to distinguish 'true'(s, w)', formed by the same process that gives us 'bald'(x, w)', and 'true _{w} (s)'. See n. 16.

Kripke's against the Fregean status of names introduced by description which I mentioned earlier.³² The argument rested on the claim that, even when the reference of 'a' is fixed by the description ϕ , one could not be taken to have said that the ϕ is F by uttering 'a is F' since the statement that a is F, and the statement that the ϕ is F have different modal properties. We can now see that this argument depends upon just the connection between content and proposition which I have challenged. It does not follow from the fact that one who utters the sentence 'a is F' says that the ϕ is F, that the sentence 'a is F' and 'The ϕ is F' will embed inside modal operators in the same way. We can regard the theorem:

'a is F' is true iff the ϕ is F

as showing the sense or content of the sentence 'a is F' without arriving at the result that:

'Possibly (a is F)' is true iff Possibly (the ϕ is F).

While we have agreed that absolute truth conditions must be derivable from true conditions, there is no reason for supposing that true_w conditions must be derivable from truth conditions, or content, but this is what Kripke's argument requires.

IV

The relationship which holds between what a sentence says and the proposition with which it is associated allows for the possibility that there be a wholly uninformative sentence which is nevertheless contingent. A sentence with this character could express *a priori* knowledge without engendering any paradox, since, in knowing it to be true, one thereby knows nothing about the world. To make this possibility clear, let us return to the sentences:

John is as tall as John,

John is as tall as himself.

³² See p. 182 above.

If we ignore for a moment the worry that John might not exist (a worry which corresponds to nothing in the case of worlds), we can know both sentences to be true *a priori*. Now, the second sentence presents us with no problem, for we can think of it as ascribing to John a property possessed by every factual and possible object. But if we thought of the first sentence as ascribing to John the property of being the same height as John, as we are entitled to do, our *a priori* knowledge might give us pause, for this is certainly not a property which is possessed by every actual and possible object. We might worry: how can we know, *a priori*, without looking, that John has this property? How can we know without investigation that John is not one of the objects which lack this property?

No one would really find this perplexing. While it might appear on a casual inspection that one cannot know *a priori* that an object possesses a property not possessed by every object, with these examples in mind, we will readily amend the principle to hold that one cannot know *a priori* that an object possesses a property not possessed by every object, unless the specification of that property requires reference to that object. The same thing goes for properties of worlds. We can know *a priori* that the actual world possesses certain properties not possessed by every world so long as the specification of those properties requires reference to the actual world, or to what is actually the case.

The parallel with the sentences about John's height is also useful because it permits a very sharp characterization of the position adopted by Donnellan and those he was following. One holds a position parallel to Donnellan's in relation to the sentences about John's height if one maintains the view that, if a property is possessed by an object which is not possessed by every object, then it must hold in virtue of the existence of some state of affairs which can only be known to exist *a posteriori*—'by looking'. Presumably, the relevant state of affairs must be that of John's being a particular height—say 6 feet. So, one concludes that the sentence 'John is as tall as John' really says that John is 6 feet tall, that 'as tall as John' refers to the height of 6 feet, and that someone who does not know what height John is, and

hence does not know to what height 'as tall as John' refers. does not really understand the sentence.

Sentences constructed with the operator 'Actually' provide very clear illustration of the way the possibility I have described can be exploited to yield harmless contingent *a priori* truths. Such sentences also help to free our minds from the idea that the puzzle about the contingent *a priori* has any particular connection with the theory of reference. I take 'Actually' to be an operator which yields a sentence true with respect to a world iff the sentence to which it applies is true *simpliciter*, and, correspondingly, that an object satisfies_w'Actually' A iff it satisfies_{w*}A.³³ The sentences Q, and 'Actually (Q)' will be epistemically equivalent, but will embed differently under modal operators, and are hence associated with different propositions. For example, though the sentence:

If you had painted the post with this paint, it would have been red,

said of a green post, might be true, the sentence:

If you had painted the post with this paint, it would have been actually red,

said of the same post, would not be true.

Now, consider a sentence of the form:

$$(x) (F(x) \supset \text{Actually} (F(x)))$$

(e.g. 'If anything is red it is actually red'). This statement is clearly contingent; there are worlds with respect to which it is not true, viz. any world in which there are things which are not red in the actual world which are red. Nevertheless, it can clearly be known *a priori*. So, there is a property which we can know *a priori* to be possessed by the actual world, but which is not possessed by every world. Is this disturbing? Not at all. For consider what property it is:

³³ It is not important if this does not provide an exact description of the behaviour of the 'actually' operator of English; as with descriptive names, all that matters for the puzzle about the contingent *a priori* is that an operator with these properties can coherently be introduced into the language. For a treatment of a modal logic containing 'Actually' see J. Crossley and L. Humberstone, 'The Logic of "Actually"', *Reports on Mathematical Logic* 8 (1977), pp. 11-29; and M. K. Davies, *Truth, Quantification and Modality* (Ph.D. diss., Oxford, 1976).

$$\lambda w [(x) (F' (x, w) \supset F' (x, w^*))].$$

It is hardly surprising that we know *a priori* that the actual world possesses this property:

$$(x) (F' (x, w^*) \supset F' (x, w^*)).$$

This is an example of a contingent *a priori* truth which cannot be dealt with along Donnellan's lines; it can hardly be maintained that one who knows the meaning of 'red', the quantifiers, and the 'actually' operator, does not understand these sentences, nor is it easy to see what bit of *a posteriori* knowledge would be required to really understand them.

This brings us very close to a solution to the puzzle presented by (S). The property (S) demands of an arbitrary possible world is:

$$\lambda w [(\exists x) (\phi' (x, w)) \supset [a] \phi' (a, w)].$$

Given the way the reference of 'a' (i.e. 'Julius') is fixed, this is the property:

$$\lambda w [(\exists x) (\phi' (x, w)) \supset (Ix) (\phi' (x, w^*); \phi' (x, w))].$$

This is certainly not possessed by every world, but the fact that we can know *a priori* that it is possessed by the actual world is once again quite unsurprising, for it amounts to the knowledge that:

$$(\exists x) (\phi' (x, w^*)) \supset (Ix) (\phi' (x, w^*); \phi' (x, w^*))$$

and knowing this, we know absolutely nothing about the actual world.

To summarize the position we have arrived at, we can now see that perplexity occasioned by the contingent *a priori* status of (S) stems from a confusion between two different notions of contingency, which usually, but do not always, coincide. A sentence may be either superficially contingent or deeply contingent. A sentence (Q) is superficially contingent iff $\Diamond (Q)$ is true, or, equivalently, there exists a world *w* such that it is not the case that (Q) is true_w. It is only in this sense that (S) is contingent. But it was wrongly assumed that (S), if contingent, was contingent in a different sense, captured by Kripke's phrase: 'dependent upon some

contingent feature of the actual world'. We have the idea of a state of affairs, or a set of states of affairs, determined by the content of a sentence as capable of rendering it true, so that one who understands the sentence and knows it to be true, thereby knows that such a verifying state of affairs obtains. A deeply contingent statement is one for which there is no guarantee that there exists a verifying state of affairs. If a deeply contingent statement is true, there will exist some state of affairs of which we can say both that had it not existed the sentence would not have been true, and that it might not have existed.³⁴ The truth of the sentence will thus depend upon some contingent feature of reality.

(S) is not deeply contingent. A conditional sentence is rendered true by anything which falsifies the antecedent or verifies the consequent. Sub-(S), the consequent, is made true by the obtaining of any member of the set of states of affairs:

{x's being the inventor of the zip and inventing the zip; y's being the inventor of the zip and inventing the zip; z's being the inventor of the zip and inventing the zip . . . }

which reduces to the set:

{x's being the inventor of the zip; y's being the inventor of the zip; z's being the inventor of the zip . . . }.

The antecedent is falsified by the absence of every member of this set. Hence, (S) is made true by the absence of every member of this set or by the presence of any member of this set. Given that this is so, although (S) is true, and superficially contingent, it is not deeply contingent—there is no contingent feature of reality on which its truth depends.

(S) demands nothing of the actual world, which is why knowing it to be true *a priori* constitutes no paradox. Nevertheless, we have seen how it is possible for a sentence which demands nothing of the actual world to be false with respect to some possible world. This can arise when the condition which it demands of an arbitrary possible world is formulated

³⁴ Notice that in formulating the notion of deep contingency, one uses the notion 'true' (s, w) and not 'true_w(s)'.

in terms of what is actually the case. Thus a sentence can be superficially contingent without being deeply contingent, and with this comes the possibility of a perfectly innocent, if rather uninteresting, example of the contingent *a priori*.³⁵

³⁵ The distinction between deep and superficial modal properties also has application to the case of the necessary *a posteriori*, but I must resist the temptation to elaborate. The distinction is adopted, and provided with a fascinating background which enables many further consequences to be drawn, in 'Two Notions of Necessity' by M. K. Davies and L. Humberstone, in *Philosophical Studies* 38 (1980), pp. 1-30.

Pronouns

1 Introduction

A very natural, preliminary classification of the uses of pronouns would include the following three categories:

(i) Pronouns used to make a reference to an object (or objects) present in the shared perceptual environment, or rendered salient in some way. The sentence (1),

(1) He's up early

said of a man passing in the street, or (2),

(2) I'm glad he's left

said of a man who has just walked out of the room, exemplify this use.

(ii) Pronouns intended to be understood as being co-referential with a referring expression occurring elsewhere in the sentence. One of the readings of the sentence (3) results from such a use of the pronoun *his*.

(3) John loves his mother.

(iii) Pronouns which have quantifier expressions as antecedents, and are used in such a way as to be strictly analogous

From *Linguistic Inquiry* 11, 2 (Spring 1980), 337–62. Reprinted by permission of the MIT Press. About the origin of this article. I published in the *Canadian Journal of Philosophy* a long paper 'Pronouns, Quantifiers and Relative Clauses (I)' [reprinted here in Chapter 4. Ed.] addressed to a philosophical audience, on the semantics of pronouns. Since my ideas bore upon recent work by linguists, it was suggested to me that it might be of interest if I presented them in a form and place more accessible to linguists, and 'Pronouns' is an attempt to do this. Although it is intended to be self-contained, and includes much that is not in the original paper (especially sections 4 and 5, which are devoted to a discussion of Lasnik's views on co-reference), I hope that those who are interested will be encouraged to consult the original paper, where many matters merely raised here are dealt with in detail. I am very grateful to Deirdre Wilson and Andrew Radford for helpful comments.

to the bound variables of the logician. The pronoun in the sentence (4) appears to be used in this way.

(4) Every man loves his mother.

I have two main points which I want to make about pronouns. First, there is a fourth category, which I call 'E-type pronouns', the members of which are very frequently confused with the members of category (iii), but which in fact are semantically quite different. E-type pronouns also have quantifier expressions as antecedents, but they are not bound by those quantifiers. For example, the pronoun in the sentence (5):

(5) Few MPs came to the party but they had a good time

is an E-type and not a bound pronoun. I shall discuss these pronouns in section 2.

My second main point concerns the semantics of pronouns in category (iii)—'bound pronouns', as I shall call them. Linguists tend to regard the semantics of bound pronouns (or of bound variables) as a mystery clearly understood by logicians, and to leave matters there. But we cannot afford to be so incurious. It is a very striking fact about pronouns in natural languages that they have this use, in addition to their other uses, and we must wonder whether this is an accident, or whether there is some underlying semantic principle which accounts for these apparently disparate uses in a unified way. In fact there is this very striking connection between pronouns in categories (ii) and (iii): whenever we substitute a singular term for a quantifier binding a pronoun, we arrive at a sentence in which the pronoun can be interpreted as co-referential with that singular term. This surely suggests that some common principle underlies the use of pronouns in categories (ii) and (iii)—that the capacity we have to understand sentences like (3) is, at the very least, *connected with* the capacity we have to understand sentences like (4). If we look at matters in this way, we see that the relationship between pronouns in categories (ii) and (iii) is a version of a problem which is frequently encountered in semantics, for there are many devices which occur, apparently univocally, in both singular and quantified

sentences. The semantic problem posed by those dual occurrences can be solved quite generally if we provide a semantic account of quantified sentences which proceeds by way of a semantic account of their singular instances. If we adopt such an account, which is motivated quite independently of any consideration of pronouns, we have only to explain the semantic significance of pronouns in category (ii) and nothing whatever has to be said, in addition, about pronouns in category (iii)—they simply look after themselves. I shall attempt to show this in section 3.

The ideas which I shall advance in section 3 place me in direct opposition to an approach to pronouns originally advanced by Lasnik¹ and received with favour by some other linguists.² One way of explaining Lasnik's main idea is to say that he proposes an incorporation of pronouns in category (ii) into category (i). If we regard an object's having been mentioned in a previous conversation, or having been mentioned previously in the conversation, as a way of its being salient for purposes of reference, as we must, why should we not regard being mentioned elsewhere in the same sentence as a limiting case of this mode of salience? If we do so, there is no good reason for distinguishing pronouns in category (ii) from those in category (i). (Since the reference of pronouns in category (i) is determined by what may loosely be called 'pragmatic' factors, I shall call Lasnik's theory 'the pragmatic theory of co-reference'.) At first sight, Lasnik's proposal has the appeal of simplicity, but on further reflection, we can see that its price is too high. For to assimilate pronouns in category (ii) to those in category (i) is to preclude the recognition of any connection whatever between pronouns in the unified category (i) + (ii) and those in category (iii)—that is, it forces us to regard as a complete accident that the same expression shows up in both (3) and (4). I shall attempt to explain this point, and the distinction

¹ H. Lasnik, 'Remarks on Coreference', *Linguistic Analysis* 2 (1976), pp. 1-22.

² See, e.g., Bresnan, 'A Realistic Transformational Grammar', in M. Halle, J. Bresnan, and G. A. Miller (eds), *Linguistic Theory and Psychological Reality* (Cambridge, Mass.: MIT Press, 1978); Chomsky 'Conditions on Rules of Grammar', *Linguistic Analysis* 2 (1976), pp. 303-51. Lyons puts forward a doctrine similar to Lasnik's in Lyons, *Semantics*, 2 (Cambridge: Cambridge University Press, 1977), sect. 15.3.

between pragmatics and grammar which it forces upon us, in section 4. In the remaining section I shall advance some other criticism of Lasnik's approach to pronouns.

2 The Existence of E-Type Pronouns

I shall begin my attempt to demonstrate the existence of a fourth category of pronouns by considering sentences containing plural quantifiers, since the distinction between a bound and an E-type interpretation of a pronoun comes out most clearly when the pronoun has a plural quantifier as antecedent. Consider the two sentences (6) and (7):

- (6) Few congressmen admire only the people they know.
 (7) Few congressmen admire Kennedy, and they are very junior.

At first sight the relationship between the quantifier phrase *few congressmen* and the pronoun *they* in these two sentences appears to be the same, but on closer inspection we can see that it is really quite different. In (6) the pronoun is bound by the quantifier phrase, while in (7) it cannot be. If the pronoun in (7) is to be bound by the quantifier phrase, *few congressmen*, then its scope must extend to the second conjunct, and the sentence would be interpreted as meaning that few congressmen both admire Kennedy and are very junior. But this is not the interpretation naturally placed upon (7). First, (7) entails, as its supposed paraphrase does not, that few congressmen admire Kennedy, period. Second, (7) entails, as its supposed paraphrase does not, that *all* the congressmen who admire Kennedy are very junior.

Similarly, in the sentence (8), the quantifier phrase *some sheep* does not bind the pronoun *them*.

- (8) John owns some sheep and Harry vaccinates them in the spring.

If it did bind the pronoun, the sentence would be equivalent to (9), in which both pronouns are bound.

- (9) Some sheep are such that John owns them and Harry vaccinates them in the spring.

But (9) is not equivalent to (8). (8) entails, as (9) does not, that Harry vaccinates *all* the sheep which John owns.

A useful test of whether or not a pronoun is bound by an antecedent quantifier is to replace the antecedent with the quantifier expression *No*, and see whether the result makes sense. Where a pronoun is bound by a quantifier, as in (10),

(10) Few men despise those who stand up to them

it is part of a complex predicate (e.g. '() despise those who stand up to them'), which is affirmed to be satisfied in the case of many girls, some sheep, or, in the example, few men, and which can be affirmed to be satisfied in the case of no girls, no sheep, or no men. Thus we have:

(11) No men despise those who stand up to them.

When we apply this test to pairs of examples we have considered, we find confirmation of the semantic difference which we have noticed.

(12) No congressmen admire only the people they know.

(13) *No congressmen admire Kennedy, and they are very junior.

(14) No sheep are such that John owns them and Harry vaccinates them in the spring.

(15) *John owns no sheep, and Harry vaccinates them in the spring.

When a pronoun is bound by a quantified expression, it does not make sense to ask to what it refers (on that occasion of use). If you have any doubt of this, you should be able to convince yourself of its truth by attempting to answer the question in the case of sentences like (4), (6), (11), (12), and (14). You will get into hopeless muddles, which have been analysed in detail by Geach in his book *Reference and Generality*.³ How the semantic functioning of bound pronouns is to be accounted for is something I shall discuss in section 3, but it ought to be clear at even this early stage

³ P. T. Geach, *Reference and Generality* (Ithaca, NY: Cornell University Press, 1962).

that a pronoun bound by a quantifier does not refer to anything. However, just looking at the examples we have so far considered, it appears that the pronouns we have shown not to be bound by their quantifier antecedents *are* interpreted as referring to something. If someone asked, about (8), 'Harry vaccinates *them*? Which sheep?', the natural (and, in my view, rule-governed) answer is: 'The sheep which John owns, of course.' Equally, the question 'They are junior? Who do you mean?' would be answered 'Why, the congressmen that admire Kennedy.' It looks as though the role of the pronoun in these sentences is that of referring to the object(s), if any, which *verify* the antecedent quantifier-containing clause.⁴ If this is the role of these E-type pronouns, we explain why the truth of the clause containing them requires that *all* the relevant objects satisfy the predicate, and we explain why these pronouns cannot have a *No* quantifier as antecedent.

It is clear from the examples I have given that whether or not a pronoun is interpreted as bound by a quantifier phrase depends upon the grammatical relation in which it stands to that quantifier phrase. All the examples so far have involved coordinate structures, but consideration of a wide range of examples containing plural quantifiers supports the generalization that a pronoun will be interpreted as bound by a quantifier phrase only if it precedes and *c-commands* the pronoun.⁵ Using this rule, one can construct endless examples of E-type pronouns:

(16) After Harry bought some sheep, Harry vaccinated them.

(17) *After Harry bought no sheep, Harry vaccinated them.

(18) If many men come to the ball, Mary will dance with them.

⁴ By 'objects, if any, which verify the antecedent quantifier-containing clause' I mean those objects, if any, which satisfy the predicate in the antecedent clause and thereby make that clause true.

⁵ A constituent A *c-commands* a constituent B if and only if B is dominated by the first branching node which dominates A. In 'Pronouns, Quantifiers and Relative Clauses (I)' [reprinted here in Chapter 4. Ed.] I used Klima's term 'in construction with' for the converse of this relation.

- (19) *If no men come to the ball, Mary will dance with them.

When we come to consider *singular* quantifiers, it appears that the precede and *c*-command relationship marks a similar semantic distinction. For example, it is clear that the scope of the quantifier in the sentence (20):

- (20) Just one man drank champagne and he was ill

can extend only to the end of the first clause; if it is interpreted as extending to the whole sentence, to bind the pronoun *he*, we get the quite different proposition: 'Just one man both drank champagne and was ill'. The point also shows up very clearly with *every*-sentences. When the quantifier *every* precedes and *c*-commands a singular pronoun, as in (4), the pronoun is interpreted as bound, but when it does not stand in this relation to the pronoun, the result is unacceptable:

- (21) *Every congressman came to the party, and he had a marvellous time.⁶

If it is the role of pronouns not *c*-commanded by their quantifier antecedents to refer to the object(s) which verify the antecedent clause, the deviance of (21) is explained, since in the antecedent clause there are asserted to be a plurality of such objects. (21) is certainly improved by pluralizing the pronoun:

- (22) ?Every congressman came to the party, and they had a marvellous time.

It is also possible to show that the precede and *c*-command configuration determines the difference between bound and E-type pronouns in the case of the simple existential quantifiers, 'some man', 'a boy', 'there is a girl', etc. If we look back on the arguments which we have used to show that a given pronoun is not bound by a quantifier antecedent, we shall see that they are basically of two kinds. One kind of argument is a scope argument; we show that in order to give a certain sentence the correct interpretation, the scope of the

⁶ (21) is very similar to an example noticed by Chomsky, op. cit.: '(i) Every soldier is armed, but will he shoot?'

quantifier cannot extend to the clause containing the pronoun—we show that the quantifier must have 'narrow scope'. If we use 'Q + CN' to represent a given quantifier plus common noun, 'A()' to represent the context of the quantifier expression in the first clause, and 'B()' to represent the context of the pronoun in the second clause, then the conjunctive sentences on which we have been concentrating can be represented schematically as follows:

A(Q + CN) and B(it/them, etc.).

To represent the pronoun in the second clause as bound requires understanding the whole sentence as having the form:

$Q_{CN}x(A(x) \& B(x))$

in which the quantifier phrase, $Q_{CN}x$, e.g. 'For some man *x*', 'For all girls *y*', has 'wide scope'. And in the case of certain quantifiers (*few*, *most*, *just one*, *three*, etc.), we are able to show that this interpretation is incorrect, since the original sentence entails, as the 'wide scope' interpretation does not, the simple:

$Q_{CN}x(A(x))$.

The second kind of argument that a pronoun is not bound depends upon its being interpreted in such a way that the truth of the clause containing it requires that *all* of the objects of a certain class satisfy the predicate of that clause. This again is a feature of the interpretation of the sentence which does not obtain if the pronoun is taken to be bound by the quantifier. I explained this feature by suggesting that the pronoun has a referential role, similar to that of the phrase 'the congressman who came' etc.

With slight modification, both of these kinds of argument can be constructed for the case of the simple existential quantifiers. A scope argument cannot focus upon conjunctive sentences, since the wide scope:

Some man *x* (A(*x*) and B(*x*))

does entail:

Some man *x* (A(*x*)).

However, a clear difference between wide and narrow scope interpretations of the existential quantifiers shows up in conditional sentences. Thus, in the sentence (23):

(23) If a man enters his room, he will trip the switch

we must give the *a man* quantifier narrow scope (leaving *he* unbound); we are not saying that there is a man such that if *he* comes, he will trip the switch. Equally, in the sentence (24):

(24) If there is a man in the garden, John will tell him to leave

the *there is a man* quantifier must have narrow scope; we are not saying that there is a man such that, if *he* is in the garden, John will tell him to leave.

The second kind of argument can apply even in cases in which the scope argument cannot apply, but we must first ask what is the feature of the use of a *singular* pronoun which corresponds to the requirement made by a plural pronoun that *all* the members of a certain class satisfy the predicate. If it is the role of E-type pronouns in general to refer to the object or objects which verify the antecedent clause, and if an E-type pronoun is singular, then we would predict that the use of that pronoun will convey the implication that there is *just one* object verifying the antecedent clause—an implication which is not carried by the use of the existential quantifiers themselves. It seems to me that this is exactly what we find. When a pronoun is in a clause coordinate with the clause containing the quantifier, as in (25),

(25) Socrates owned a dog and it bit Socrates

there is a clear implication that Socrates owned just one dog. It is precisely because of the implication carried by such pronouns that it is not acceptable to report the non-emptiness of the class of Welsh doctors in London by saying (26):

(26) There is a doctor in London and he is Welsh.

Notice that no such implication is carried by the use of a pronoun which is *c-commanded* by its antecedent:

(27) Socrates owns a dog which bites its tail.

This point about uniqueness may be felt to be slender evidence, though the counter-examples that have been suggested to me normally ignore the temporal parameter implicit in the tense of the verb.⁷ In a great many cases the implication is clear, and since the presence or absence of this implication appears to depend upon the precede and *c-command* configuration presumed to be relevant on quite independent grounds, the hypothesis that that configuration determines the difference between bound and E-type pronouns seems to be sustained even in the case of simple existential quantifiers.

The principle I have stated effectively restricts the scope of a quantifier to those elements which it precedes and *c-commands*. However, there are quantifiers in English which are almost always given wide scope, and the principles must be qualified to exclude them. The two most important examples are *a certain* and *any*. Thus, in the sentences (28) and (29), the quantifiers are given wide scope, and the pronouns are bound and not E-type.

(28) If a certain friend of mine comes, he will tell the police.

(29) If any man loves Mozart, he admires Bach.

I mention these exceptions not only for the sake of accuracy, but also as a partial explanation of why it has taken so long for the important *grammatical* distinction between bound and E-type pronouns to be noticed. If these 'wide-scope-seeking' quantifiers are not excluded, it is very difficult to see a pattern in the jumble of examples.

⁷ e.g. the sentence (i) has been given to me by Geach as a counter-example to my claim that such pronouns carry an implication of uniqueness:

(i) Socrates kicked a dog and it bit him and then Socrates kicked another dog and it did not bite him.

However, the tense in the verb effectively introduces an initial existential quantifier *There was a time such that . . .*, and my claim is that the truth of the sentence requires that there be a time such that Socrates kicked only one dog *at that time*, not that Socrates only kicked one dog *ever*. However, the interpretation of *a*-expressions is unclear, and we may be forced to recognize that they are sometimes used as equivalent to *any* and sometimes to *a certain*. (See below.) This affects only the scope, and not the substance, of my claim.

Finally, what is the importance of the distinction between bound and E-type pronouns—of the addition of another category of pronouns to the list? Not, it must be admitted, very great. Pronouns are often used as referring expressions, and it is not particularly surprising that some of them should have their reference fixed by a description recoverable from the antecedent, quantifier-containing, clause. The point only assumes importance in the context of certain current views. Certain logically-minded philosophers have been so impressed by the undoubted analogies which exist between some pronouns and the bound variables of quantification theory that they have enthusiastically adopted the hypothesis that *all* natural language pronouns with quantifier antecedents are bound by those antecedents.⁸ Other researchers, concentrating upon sentences containing E-type pronouns, have naturally been impressed by the idea that such pronouns are referring expressions, and have, in a contrary excess of enthusiasm, attempted to regard bound pronouns as referring expressions, but without any success. There has therefore arisen the idea that there are two approaches to the unified subject of ‘anaphora’—the bound variable, and the referential—between which we have to choose.⁹ The one merit of taking seriously the argument of this section is that this profitless debate should end. There is not a single class of pronouns for which we must find a unitary explanation. There are two kinds of pronouns, which are sharply distinguished by their grammatical position, and which function in quite different ways.¹⁰

⁸ The most prominent example is Geach; see esp. the papers collected in sect. 3 of *Logic Matters* (Oxford: Blackwell, 1972).

⁹ See, e.g., Stenning, ‘Anaphora as an Approach to Pragmatics’, in Halle, Bresnan, and Miller (eds), op. cit.

¹⁰ I am by no means the first to have noticed that some pronouns with quantifier antecedents are not bound by them. Apart from the paper of Chomsky’s cited in fn. 2, reference should be made to Karttunen, ‘Pronouns and Variables’, in R. I. Binnick, A. Davison, G. Green, and J. Morgan (eds), *Papers from the Fifth Regional Meeting of the Chicago Linguistic Society* (Chicago, Ill.: University of Chicago, 1969) and R. Jackendoff *Semantic Interpretation in Generative Grammar* (Cambridge, Mass.: MIT Press, 1972). However, these works contain isolated examples rather than theory, and give no clue of the extent of the phenomenon. In reply to an example of Jackendoff’s, Janet Fodor (*Semantics: Theories of Meaning in Generative Grammar* (New York: Crowell, 1977)) suggests that the thesis that all pronouns with quantifier antecedents are bound might be sustained

3 The Semantics of Bound Pronouns

With E-type pronouns out of the way, we can raise the question of the proper treatment of bound pronouns. Any such treatment must take account of two fundamental points. The first point is very well known. We cannot give the same account of the pronoun in the sentence (4):

(4) Every man loves his mother

as we might be inclined to give to the pronoun in the sentence (3),

(3) John loves his mother

namely that of referring to whatever its antecedent refers to, since the antecedent expression *every man* is not an expression referring to anything.¹¹ I mentioned the second point in my introduction; it is equally important but it has not received the attention it deserves. We cannot give utterly unconnected explanations of the roles of the pronouns in the sentences (3) and (4); it is simply not credible that the speaker’s capacity to understand the sentences ‘John loves his mother’, ‘Harry loves himself’, ‘Susan admires those who are nice to her’, etc., is in no way connected with his understanding of the sentences ‘No man loves his mother’, ‘Every man loves himself’, ‘Just one girl admires those who are nice to her’, etc. Putting these points together, we are confronted with the following problem: we must provide an account which is adequate to deal with pronouns in both our categories (ii) and (iii), but we cannot *directly* apply the most obvious account of pronouns in category (ii) to pronouns in category (iii).

This may seem to be an insoluble problem, but in fact it is a version of a problem with which the semantics of quantified sentences make us very familiar. For there are many devices which appear in both singular and quantified sentences in such a way that, though the two uses are clearly

by ‘relaxing the usual constraints on the binding of variables’ (p. 192)—an intriguing proposal which I wish I could understand.

¹¹ Even if you can persuade yourself that *every man* refers to every man, *his* still cannot be co-referential with its antecedent, on pain of generating the incorrect reading: every man loves the mother of every man.

connected, the most obvious account of their occurrence in singular sentences does not immediately apply to their occurrence in quantified sentences. The most familiar examples are provided by the sentential operators, *and*, *or*, *if*, etc. When these expressions join singular sentences, as in (30),

(30) If John is in love, John is happy

they can be regarded as truth-functions—that is to say, expressions which map pairs of truth values onto truth values. If they joined only singular sentences—expressions which have a truth value—then all we would need to say about them is that they yield a true sentence when the constituent sentences have such-and-such a combination of truth values. But, in the sentence (31)

(31) If any man is in love, he is happy

the constituents which *if* joins are not sentences with a truth value at all. (31) is true iff¹² any man satisfies ‘if () is in love he is happy’, but here *if* joins the parts of a complex predicate, and since the parts do not have a truth value, we do not yet appear to have any account of this role. A position parallel to the one which certain contemporary linguists adopt in the case of pronouns would then be to cast around for another account to deal with the connective *if* in this role—a position on which *if* (and all the other connectives) are deemed to be ambiguous. But Frege, and subsequent logicians, were not content with this approach; they searched for a way to unify the two roles.

One very natural, though not the only, way of doing this was adopted by Frege. Frege’s approach involves no modification whatever of the truth-functional account of the role of the connectives; this is left to stand as the single account of their semantic contribution. The novelty comes in the account which Frege offered of the notion of *satisfaction*. When a predicate is complex, a Fregean explains the notion of an object’s satisfying it in terms of the *truth* of a sentence which results when a singular term referring to that object is substituted in, or coupled with, the predicate. Thus,

¹² i.e. ‘if and only if’.

a Fregean does not define the conditions under which an object satisfies the predicate (32):

(32) If () is in love, he is happy

directly, in terms of the satisfaction conditions of the two parts, but rather says this: an object *x* satisfies the predicate (32), iff the *sentence* (33):

(33) If (β) is in love, he is happy

is true when we interpret ‘ β ’ as referring to *x*. Now, relative to that interpretation of ‘ β ’, *if* once again joins expressions which may be assigned a truth value, and its role as a truth-function can be invoked.

I do not say that this is the only way to provide a unitary account of the role of the sentential operators, but it is certainly a very natural way. And it is completely general. If we adopt this Fregean explanation of what it is for an object to satisfy a complex predicate, we have only to explain the significance of a device as it occurs in singular sentences, and its occurrence in quantified sentences simply takes care of itself. In particular, this observation applies to pronouns; if we adopt a Fregean account of satisfaction, we have only to give an account of the pronoun-antecedent construction as it occurs in singular sentences—no further explanation need be given of pronouns with quantifier antecedents.

As we have seen, a natural explanation of the role of pronouns with singular antecedents is in terms of *co-reference*—the pronoun refers to whatever the antecedent refers to. This account will secure the result that there is a reading of the sentence ‘John loves his mother’ on which the reference of *his* is the same as that of *John*, and hence that the whole sentence is true iff John loves John’s mother. If we put this obvious account together with the general Fregean explanation of satisfaction—an explanation which is independently needed to provide a unification of the roles of other devices which occur in both singular and quantified sentences—then we have an explanation of the role of the bound pronoun in (4):

(4) Every man loves his mother.

Such a sentence is true iff every man satisfies the complex predicate '() loves his mother'. By the Fregean explanation, we know that an arbitrary object, x , satisfies that predicate iff, taking ' β ' as referring to x , the sentence ' (β) loves his mother' is true. Now we can apply the co-reference rule quite properly, and learn that such a sentence is true iff the referent of ' β ' loves the mother of the referent of ' β ', and hence iff x loves the mother of x . Hence, x satisfies the complex predicate iff x loves x 's mother, and so the whole sentence is true iff for every man x , x loves x 's mother.

This explanation of the functioning of bound pronouns presupposes the following: that whenever we replace a quantifier which binds a pronoun with a singular term, in order to form a relevant substitution instance, the resulting sentence will be one which admits of a reading on which the pronoun is co-referential with that singular term. (I am ignoring the problem of number-agreement.) But this is exactly what we find. We remarked earlier that a quantifier can bind a pronoun only if it precedes and c -commands that pronoun, and it is sufficient (though not necessary) for a pronoun to be interpretable as co-referential with a singular antecedent that it be preceded and c -commanded by that antecedent. The fact that this is so constitutes a powerful argument, not for the need to unify the two roles, which I take to be indisputable, but for a unification along Fregean lines. The fact that there is this correspondence between sentences containing bound pronouns and singular sentences containing pronouns admitting of a co-referential interpretation can no more be regarded as an accident than the fact that pronouns are used in both singular and quantified sentences in the first place.

When we come to evaluate substitution instances of the form ' (β) is A ' we must use information about the context in which the original quantified sentence (e.g. 'Everything is A ') is uttered in order to evaluate any ambiguous or context-dependent words which occur in A . In particular, we presume that anyone who utters a token of the sentence type (4):

(4) Every man loves his mother

intends to use *his* either as a device making an independent

reference to some salient object or as a device for registering co-reference. In the first case, we evaluate the truth of instances of the form ' (β) loves his mother' by taking the reference of *his* to be constant and determined by the appropriate contextual factors. In the second case we evaluate it as before.

To summarize: while it is quite correct to observe that one cannot deal with bound pronouns by directly applying a co-reference rule, this point should not lead us to the desperate conclusion that pronouns with singular and quantified antecedents are semantically unrelated. If we employ a Fregean explanation of the notion of satisfaction of a complex predicate (and some such explanation is independently needed to deal with other devices which show up in both singular and quantified sentences), we find that a co-referential explanation of pronouns is all the explanation we need.

With these considerations in mind, let us look at the main ideas behind a theory of pronouns which is currently popular among linguists. I think that we can show that it is incompatible with this, or any other, attempt to see a unitary semantical phenomenon in pronouns with singular and quantified antecedents, and hence, must be wrong.

4 Consequences for Bound Pronouns of the Pragmatic Theory of Co-reference.

Lasnik begins his 1976¹³ article by taking account of the existence of pronouns in category (i), and then goes on to question whether any additional account needs to be given of pronouns in the supposed category (ii). Pronouns in category (i) involve a reference to an object which is salient in some way. Since one of the ways in which an object can be salient is by having been mentioned in a previous conversation, it would appear to be possible to regard reference to an object mentioned elsewhere in the same sentence or clause as a limiting case of the exploitation of this kind of salience. What we might rather loosely call 'pragmatic' factors would seem to ensure that one of the uses of the sentence 'John loves his mother' is to say that John loves the mother of John.

¹³ Lasnik, *op. cit.*

The flavour of this, at first sight very plausible, approach can be gathered from Lasnik's discussion of the sentence (34):

(34) After John talked to Mary, they left the room.

Within the system I have proposed . . . (34) is not problematic . . . that is, no co-reference rule is needed to explain (34) because there is nothing to explain. *They* in (34) can be used to refer to any group of entities; under many discourse situations however, John and Mary are the only likely candidates.¹⁴

Lasnik's main thesis is that 'even sentence-internal cases of co-reference are not produced by any rule'.¹⁵ According to Lasnik, the only rule of the language which concerns the interpretation of pronouns is a rule of *Non-co-reference* which, for example, prohibits any two noun phrases related as *he* and *John* are related in (35) from being co-referential, unless the second is a pronoun.

(35) He is happy when John is in love.

I will not discuss Lasnik's Non-co-reference rule until the next section; what concerns me here is the view that all the rules of the language tell us about a sentence like (3):

(3) John loves his mother

is that the pronoun *can be* referential with *John*; there is no special rule which secures an interpretation on which it is.

Implicit in this statement of Lasnik's position, and throughout his paper, is the distinction between those facts about the interpretation of an utterance which are explained by reference to *the rules of the language*, and those facts which are explained by 'pragmatic' factors. Although the distinction is difficult to make precise, it is impossible to deny, since one must admit that there are facts about the interpretation of a sentence which are in no way determined by a linguistic rule. For example, there is no linguistic rule which determines that a *he* or a *that man* refers to *x* rather than *y* in the vicinity, or that it refers to someone who

¹⁴ In this and subsequent quotations, I have altered the numbering of examples to conform to the ordering in this article.

¹⁵ Lasnik, *op. cit.*, p. 9.

has just left rather than someone who has recently been mentioned.

Chomsky has built this distinction into his current framework.

Let us say that the grammar contains a system of rules that associate a derivation with a representation of LF (read 'logical form' but for the present without assuming additional properties of this concept). I will understand LF to incorporate whatever features of sentence structure (1) enter directly into the semantic interpretation of sentences and (2) are strictly determined by properties of sentence grammar. The extension of this concept remains to be determined. Assume further that there is a system of rules that associates logical form and the products of other cognitive faculties with another system of representation SR (read 'semantic representation'). Representations in SR, which may involve beliefs, expectations and so on, in addition to properties of LF determined by grammatical rule, should suffice to determine role in inference, conditions of appropriate use etc. (Some would argue that LF alone should suffice, but I leave that an open empirical question.)¹⁶

For his part, Chomsky has stated elsewhere his viewpoint on this empirical issue.

Given the logical forms generated by sentence grammar, further rules may apply. Pronouns not yet assigned antecedents may be taken to refer to entities designated elsewhere in the sentence, though this is never necessary, and is not permitted under certain conditions . . . These further rules of reference determination may involve discourse properties as well, in some manner; and they interact with considerations relating to situation, communicative intention and the like.¹⁷

One important, and traditional constraint upon the domain of grammar proper is that it should deal only with matters that are 'sentence-internal'. Chomsky considers the discourse (36):

(36) Some of the men left today. The others will
leave later

and argues, using this constraint: 'The rule assigning an interpretation to *the others* however is not a rule of sentence grammar at all, as (36) indicates.'¹⁸ He continues:

¹⁶ Chomsky, *op. cit.*, pp. 305-6.

¹⁷ Chomsky, *Reflections on Language* (London: Temple Smith, 1976).

¹⁸ Chomsky, 'Conditions on Rules of Grammar', p. 323.

Returning to the basic theory outlined earlier, the rule of reciprocal interpretation, and DR [Chomsky's rule of Disjoint Reference] relate derivations . . . to LF, while the rule assigning an interpretation to *the others* belongs to an entirely different component of the system of cognitive structures relating to LF and other factors to a full semantic representation. It might be quite appropriate to assign this rule to a theory of performance (pragmatics) rather than to the theory of grammar.

Chomsky then goes on to apply this point in the case of anaphoric pronouns:

In [15] I pointed out that the rules of anaphora associating *he* with *John* in such sentences as (37) appear to violate otherwise valid conditions, a problem for the theory presented there:

- (37)a. John thought that he would win.
 b. John thought that Bill liked him.

Others have reiterated this point, arguing that it undermines the theory outlined. But my observation was simply an error. The rule of anaphora involved in the (normal but not obligatory) interpretation of (37) should in principle be exempt from the conditions of sentence-grammar, since it is no rule of sentence grammar at all. Cf. Lasnik.¹⁹

Chomsky's notion of *rule of sentence grammar* comprises both syntactic and semantic (or interpretive) rules, and it is clear from his paper that when Lasnik maintains that 'even sentence internal cases of co-reference are not produced by any rule', he is using 'rule' in pretty much this sense. As his criticism of Jackendoff's interpretive rule of co-reference makes clear, Lasnik's thesis is not simply that there is no *syntactic* rule (e.g. a pronominalization transformation) underlying cases of co-reference. I stress this, because I do not wish to be taken to be committed to the existence of a pronominalization transformation by opposing Lasnik's basic thesis.

Although some line must be drawn between matters belonging to grammar (widely understood) and matters belonging to pragmatics, it is not at all clear where the line should be regarded as falling. For example, while it may be clear that *syntactic* processes are 'sentence-internal', it is not

¹⁹ '15' refers to Chomsky, 'Conditions on Transformations', in S. Anderson and P. Kiparsky (eds), *A Festschrift for Morris Halle* (New York: Holt, Rinehart & Winston, 1973); and 'Lasnik' to Lasnik, *op. cit.*

at all clear to me why *semantic* rules (rules mapping surface structures onto their logical forms, in Chomsky's current framework) can concern only single sentences taken one by one, rather than sequences of such sentences. The thesis that semantic rules are sentence-internal constitutes a substantial empirical hypothesis for which empirical reasons should be given. In the light of these and other similar questions, one might despair of being able to assess Lasnik's treatment of pronouns without a tremendous amount of preliminary, and fundamental, work. However, the argument I wish to advance against Lasnik's position will exploit only the most unchallengeable property of the distinction between what belongs to grammar and what belongs to pragmatics. For it seems indisputable that if certain truth-relevant features of the interpretation of any utterance of a sentence type are held to depend upon the context in which that utterance is made, then it will not make any sense to enquire into the truth value of the sentence type, considered independently of a particular context of utterance. Since Lasnik holds that the reference of all (non-bound) pronouns is determined by pragmatic factors—'considerations relating to situation, communicative intention and the like'—he must hold that it does not make any sense to enquire into the truth value of the sentence 'John loves his mother' considered independently of any particular context of utterance; this is so even when the interpretations of the expressions 'John', 'loves', and 'mother' are given. This is the only property of a pragmatic explanation of co-reference on which I need to rely. Relying on it, I want to show that Lasnik's pragmatic theory of co-reference precludes any explanation of the connection between pronouns in his unified category (i) + (ii) and pronouns in category (iii), and thus treats it as an accident that the same expression is used in sentences like (3) and (4).

The best way to see the difficulty for the pragmatic theory of co-reference is to attempt to apply the natural, and perfectly general, method of unification which I have ascribed to Frege. We know that 'Every man loves his mother' is true iff every man satisfies '() loves his mother', and, using the Fregean notion of satisfaction, we know that an arbitrary object *x* satisfies this predicate iff, interpreting ' β ' as referring

to x , the sentence ' (β) loves his mother' is true. But at this point we are stopped, for it does not make sense, on Lasnik's view, to enquire into the truth value of a sentence of the form ' (β) loves his mother' independently of information about a particular context of utterance, and this is so even when the interpretation of the name ' β ' is fixed. It is, of course, true that the original quantified sentence will be uttered in a context, and facts about that context can be used to determine the interpretation of context-dependent or ambiguous expressions in the substitution instances. But these facts will enable us to settle upon a preferred interpretation of a given expression which is constant in all the substitution instances; there is no way these facts can determine a different referent for the pronoun in each substitution instance. We could attempt to say that x satisfies the predicate iff there is *some* possible context of utterance in which the sentence type ' (β) loves his mother' could be truly uttered, but then we will certainly get the truth conditions of the quantified sentence wrong. For if there is a context of utterance in which Harry is salient, and in which I can refer by the pronoun *his* to Harry, then, if John loves Harry's mother, but not his own, I can truly utter the sentence 'John loves his mother', so that John will satisfy the complex predicate despite the fact that he is a counter-example to the claim made by the quantified sentence.

Lasnik himself makes no attempt to explain the connection between the pronouns in his unified category (i) + (ii) and bound pronouns; he reserves bound pronouns for an appendix to the main article in which he states that 'the relationship [between antecedent quantifier and bound pronoun] should be characterized as that holding between a quantifier and the variable it binds',²⁰ and he leaves matters there. It should now be clear that this feature of Lasnik's treatment is not an accidental defect of the presentation of his ideas, but an immediate consequence of those ideas themselves. Strictly, I have only considered the unavailability of the Fregean way of discerning a connection to one who holds a pragmatic theory of co-reference, but the point holds quite generally, since the alternative (Tarskian) mode of

²⁰ Lasnik, op. cit., p. 18.

discerning the connection between devices in singular and quantified sentences is even less compatible with Lasnik's views.²¹ The point should be obvious. For it seems clear that there is no common semantic principle explaining the behaviour of pronouns in categories (i) and (iii), and this is obviously unaffected by the inclusion of members of category (ii) in category (i).

It is the price of being able to recognize the obvious semantic connection between pronouns with singular and quantified antecedents that we distinguish semantically between pronouns used as devices of co-reference, and pronouns whose reference is secured in some other way, e.g. deictically. But this is not a complication of the same kind as we have just pointed to in Lasnik's treatment, and which therefore must be thrown into the balance and weighed against it. In Lasnik's case, the complexity results from a failure to discern a connection between two obviously connected capacities. The connection can be shown empirically by demonstrating the speakers' capacity to understand new sentences—those which contain, as bound pronouns, expressions which had not explicitly figured in that role before—a capacity which presumably relies upon their familiarity with sentences in which the expression has a singular term as antecedent. But to distinguish between the functioning of pronouns in categories (i) and (ii) is not to bifurcate a single capacity in the same way. Let us agree that to understand a pronoun as referring to an object mentioned in a previous conversation is to interpret the pronoun in a way which is not specifically secured by any rule of the language—it is simply a manifestation of one speaker's

²¹ I give an account of the Tarskian approach to quantified sentences in 'Pronouns, Quantifiers and Relative Clauses' (see above). If Frege can be said to take the role of a device in singular sentences as primitive, Tarski does the opposite by taking the role of devices in forming complex predicates as primitive. The Tarskian account of the role of pronouns would then be in terms of the impact they have upon whether or not a sequence of objects satisfies the predicate in which they occur: the same element of the sequence would have to be assigned to the pronoun as is assigned to this or that other position in the predicate. I say that this is not compatible with Lasnik's views because the Tarskian cannot even prevent the complex predicate '() loves his mother' from being attached to a referring expression (like *John*), thus providing a source of a kind not recognized by Lasnik for the sentence 'John loves his mother'.

general capacity to *make sense of* the acts (including the linguistic acts) of others. Now, I am suggesting that when the previous reference is within the same sentence as the pronoun (and subject to certain further conditions) the co-referential interpretation of the pronoun is secured, as one interpretation of the sentence, by a linguistic rule. Is this to fail to see a connection between connected capacities? Not at all; we are not obliged to postulate a different mechanism of understanding in the two cases, as though in one case a book labelled *Rules of the English Language* is consulted, while in the other case, it is the book labelled *How to Make Sense of One's Fellow Men*. It is just that we describe a propensity to interpret what speakers say as being in accordance with a rule of the language only under certain conditions; one of these conditions is when the interpretation of other utterances obliges us to ascribe semantic properties to sentence types considered independently of context.

Here we touch upon a point of general interest, for it becomes clear that we can use the interpretation of quantified sentences as a general guide in drawing the line between grammar and pragmatics. In order to illustrate this, let me go back to E-type pronouns. Given what I said in section 2, it might be tempting to hold that E-type pronouns are referring expressions whose reference is fixed by pragmatic factors. On this view, all that the grammar of the language tells us about the pronoun in the sentence (38):

(38) John owns some donkeys and feeds them at night

is that it can refer to any group of entities salient in the context; one very likely (but, from the point of view of the semantics of the language, in no way privileged) group will be the donkeys which John owns.²² However, an objection can be made to this pragmatic theory which is exactly parallel to the objection I have just made to Lasnik's pragmatic theory of co-reference. For we also have to consider E-type pronouns in sentences like (39):

²² I think that this is the proposal which B. H. Partee has in mind when she suggests in 'Bound Variables and Other Anaphoras' (unpublished mimeo, 1978) that E-type pronouns involve 'pragmatic' uses of pronouns. However, the reasoning in her sect. 2 is another argument to the same conclusion as I try to establish in the present section.

(39) Every villager owns some donkeys and feeds them at night.

Now, it plainly does not make sense to ask which group of donkeys the pronoun in (39) refers to. Once the proper name *John* has been supplanted by the quantifier *Every villager*, there is no determinate answer to such a question. Hence, since it is not referring to anything in (39), a pragmatic account of E-type pronouns in sentences like (38) would leave this pronoun unaccounted for. But, once again, it is obviously wrong to see no connection between the use of the pronouns in (38) and (39), and to cast around for a new account of the pronoun in (39)—whatever that could be. Here is yet another facet of the general problem presented by devices showing up in both singular and quantified sentences, and it is susceptible of the same, Fregean, solution. Provided we give an account of the E-type pronoun in a sentence like (38), the pronoun in (39)—of which (38) is a substitution instance—will take care of itself. However, once again, the Fregean treatment presupposes that there is an interpretation of the pronoun in (38) on which its reference is determined by linguistic rule, and not by 'considerations relating to situation, communicative intention, and the like'. As in the previous case, the intention of the person who utters the quantified sentences is germane to the interpretation of that utterance, for we must know whether or not he uttered the pronoun as governed by the hypothesized rule rather than with the intention of referring to some salient group of donkeys. But if he did so, it is the rule which determines the reference of the pronoun in the relevant substitution instances; *their* reference cannot be determined by pragmatic factors, since pragmatic factors cannot determine the reference of a pronoun in a sentence whose interpretation we are considering independently of any particular context of utterance.

5 Other Difficulties in the Pragmatic Theory of Co-reference

Lasnik's thesis is that the grammar of English does not oblige us to draw any distinction between the uses of pronouns in categories (i) and (ii). Clearly we can make the objection

to such a thesis that there appear to be quite delicate syntactic restrictions upon when a pronoun can be used with the intention that it be understood to be co-referential with a given singular term—restrictions which have no parallel in the case of pronouns which are intended to be understood as making an independent reference. The restrictions concern the case in which the pronoun precedes the term with which it is intended to be co-referential; in general, such a use is felicitous only if the pronoun does not *c*-command the term.²³ For example, the pronoun in the sentence (40):

(40) He is happy when Oscar is in love

cannot felicitously be used with the intention that it be understood as co-referential with the term *Oscar*. The restriction is not pragmatic; (40) confronts the hearer with the same task of deciding upon a referent for the pronoun, and one highly salient candidate is Oscar. Examples like this make it appear that there are grammatical rules of the language which specifically regulate the occurrence of pronouns in category (ii), and hence that the grammar of the language obliges us to distinguish between pronouns in categories (i) and (ii). In this final section, I want to see whether this impression is correct, and thus to see whether the very general argument of section 4 can be reinforced by an argument resting upon details of English grammar. (Simply to have some notation, let me use the symbols *he*→, ←*she*, ←*them*, etc., to represent occurrences of pronouns which are intended to be understood as being co-referential with some succeeding or preceding singular term, and *he*↑, *them*↑, etc., to represent occurrences of pronouns which are intended to be understood as making references of all other kinds.)

Of course, Lasnik is aware of the apparent distinction the grammar of language draws between ←*he* and *he*↑, and he attempts to render the facts consistent with his unification of the categories to which ←*he* and *he*↑ belong in the

²³ Actually, Lasnik uses the notion *kommand* defined as follows: 'A kommands B if the minimal cyclic node dominating A also dominates B.' Whether we should formulate the relevant restrictions using the notion of *c*-command or of *kommand* is not a matter on which I am competent to pronounce, and is anyway not relevant to the disagreement I go on to express. I have retained the notion of *c*-command to avoid irrelevant complication.

following way. It is true, he says, that on the co-referential interpretation (40) infringes a rule of grammar, but the rule is not one which is specifically concerned with the possibilities of pronominal co-reference. The rule which (40) infringes is more general, since it prohibits co-reference between *any* two noun phrases occupying the position which *he* and *John* respectively occupy in (40), unless the latter position is occupied by a pronoun. Lasnik observes that we are inclined to take the two *Oscars* in the sentence (41)

(41) Oscar is happy when Oscar is in love

as referring to different people, and his idea is to bring this observation under the same rule which explains the infelicity of (40). The final formulation which he offers of the relevant rule of *Non-co-reference* is (42):

(42) If NP₁ precedes and *c*-commands NP₂ and NP₂ is not a pronoun, then NP₁ and NP₂ are disjoint in reference.²⁴

Now if the infelicity of (40) is indeed explained by this rule, then no argument against Lasnik's position can be based upon it. (42) is a rule of grammar which, in so far as it mentions pronouns at all, can be regarded as dealing with a unitary category, to which ←*he* and *he*↑ both belong.

Before proceeding, we must take note of an extraordinary feature of the rule which Lasnik formulates and defends: the notion of *disjoint reference* which it uses is a purely extensional one. Two NPs are co-referential in the sense of the rule (i.e. are not disjoint in reference) iff, as a matter of fact, they refer to the same thing. Hence, the rule predicts that a simple sentence of the form NP₁ — Verb — NP₂ is *ungrammatical* if it so happens that NP₁ and NP₂ refer to the same thing. (When I use the term *ungrammatical* in the context of Lasnik's views, I mean simply 'infringes a rule of sentence grammar'. The rule may of course be a semantic one.) This is not an error in formulation; Lasnik explicitly considers the objection that his rule makes all true identity statements ungrammatical and responds to it not by holding it to be based on misunderstanding, but by espousing an indefensible metalinguistic

²⁴ See fn. 23.

theory of identity statements.²⁵ But the absurdities to which a purely extensional rule of Non-co-reference gives rise are not limited to identity statements, and so no theory of identity statements, however eccentric, can repair the damage. For example, Lasnik must hold that the simple sentence (43):

(43) This man is the same height of Stalin

(said, perhaps, as part of an investigation into the man's identity) is not true, but rather *ungrammatical* if *this man* does in fact refer to Stalin. But even beyond this, Lasnik's theory predicts a difference in our reactions to the sentences (44) and (45) when *his* and *this man* are both being used to refer deictically to Stalin.

(44) Stalin signed this man's papers.

(45) Stalin signed *his* papers.

According to Lasnik's rule, (44) is ungrammatical while (45) is not.

It is not necessary for Lasnik to embrace these views. It is necessary that his Non-co-reference rule should not be specifically concerned with the co-reference possibilities of pronouns, and hence that it include *Oscar-Oscar* pairs in addition to *Oscar-his* pairs. But it is *not* necessary that it embrace *Cicero-Tully* pairs as well. Lasnik does say at one point²⁶ that 'it is possible that the notion of coreference in this rule and elsewhere in this discussion should be replaced with that of intended coreference', and such a change certainly seems indicated. Actually formulating the required notion is not easy, since one who sincerely utters an identity statement does intend that the two terms be co-referential. However, I shall suppose that it can be done—that there is some notion which can apply both to a pair of expressions one of which is a proper name and the other of which is a pronoun which has that name as antecedent, and also to two occurrences of the same proper name. For I want to challenge the heart of Lasnik's position that there is a common source to the unacceptability of (40) and (41) when this relation of 'intended co-reference' holds between the two NPs in both sentences.

²⁵ Lasnik, *op. cit.*, fn. 7.

²⁶ Lasnik, *op. cit.*, p. 6, fn. 5.

Consider a very simple example of a sentence which, when the two names are intended to be taken as co-referential, infringes Lasnik's rule:

(46) Oscar loves Oscar's mother.

It seems to me doubtful that this sentence is ungrammatical at all. Here, oddly enough, Chomsky seems to agree. Considering the two sentences 'John is here, but will he shoot' and 'If John is here then he will shoot', he writes:²⁷

In the latter case, and perhaps the former as well, substitution of *John* for *he* seems to me to impose disjoint reference. It seems that the rule applying here is not a rule of sentence-grammar, but is rather a rule assigning a higher degree of preference to disjoint interpretation the closer the grammatical connexion.

One reason for maintaining that the principle is not one of sentence grammar is that it also appears to apply to sequences of sentences; as Chomsky notes, substitution of *John* for *he* in (47) is difficult, if co-reference is intended.

(47) John is here. Will he shoot?

Another, perhaps better, reason is the fact that it is easy enough to find contexts in which repetition of the name has some point, and in which the implication of disjointness of reference is cancelled. For example, a logic teacher might say to a student:

(48) Look, fathead. If everyone loves Oscar's mother, then certainly Oscar must love Oscar's mother.

Or again, someone might reasonably say:

(49) I know what John and Bill have in common. John thinks that Bill is terrific and Bill thinks that Bill is terrific.²⁸

²⁷ Chomsky, 'Conditions on Rules of Grammar', fn. 37.

²⁸ To deem sentences like 'John thinks that John is ill' ungrammatical is to claim the existence of a significant limitation upon the expressive power of English. How are we to state John's possession of a 'non-self-conscious' belief that he may have about someone who is in fact himself, but whom he does not know to be himself?

Or again,

- (50) Who loves Oscar's mother? I know Oscar loves Oscar's mother, but does anyone else?

When the conversational concern is with those who satisfy the predicate '() loves Oscar's mother', there is some point in abandoning the familiar and less prolix way of saying that Oscar loves his mother which pronouns make possible, and uttering (46) instead. Whether or not we go all the way with Chomsky, and deny that the infelicity of (46) is grammatical in origin, it cannot be denied that the appropriate conversational setting vastly increases the acceptability of such sentences. However, we will find that nothing can be done to increase the acceptability of sentences which infringe the conditions upon when a pronoun can pick up its reference from an NP elsewhere in the sentence, and this constitutes fairly strong evidence that the explanation of the unacceptability of the sentences (40) and (41) is different.

It might appear that the right way to test the hypothesis that there is a single phenomenon underlying the deviance of (40) and (41) is to compare the discourses (51) and (52):

- (51) Everyone has finally realized that Oscar is incompetent.
Even Oscar has realized that Oscar is incompetent.
- (52) Everyone has finally realized that Oscar is incompetent.
Even he has finally realized that Oscar is incompetent.

However, the sentence which is supposed to provide the appropriate conversational setting unfortunately also provides an alternative antecedent for the pronoun, and the result is quite tolerable. (I shall return to this kind of discourse below.) To test the hypothesis we need a suitable conversational setting which does not provide an antecedent for the pronoun. For example:

- (53) Everyone here admires someone on the committee.
Joan admires Susan, Mary admires Jane, and Oscar admires Oscar.

- (54) Everyone here admires someone on the committee.
Joan admires Susan, Mary admires Jane, and he admires Oscar.

Although to my ear (53) is tolerable, (54) is quite impossible, yet Lasnik's grammar cannot distinguish between them. Again, consider:

- (55) Everyone eventually realizes that someone dear to them is incompetent. For example, Mary has realized that Fred is incompetent, Susan has realized that her daughter is incompetent, and Oscar has realized that Oscar is incompetent.
- (56) Everyone eventually realizes that someone dear to them is incompetent. For example, Mary has realized that Fred is incompetent, Susan has realized that her daughter is incompetent, and he has realized that Oscar is incompetent.

These examples show that the unacceptability of a sentence like (40) is not due to its infringing a general rule prohibiting 'intended co-reference' between any two NPs related as the NPs are in (40). We get no closer to the truth by restricting Lasnik's Non-co-reference rule to pronominal co-reference, though that would undermine Lasnik's position. For a discourse like (52) shows that a pronoun can both precede and c-command an NP with which it is intended to be a co-referential—*so long as it does not pick up its reference from that NP*. To formulate the rule correctly, we need this idea of one term's *picking up its reference from* another. So far, we have considered rules which use a *symmetrical* relation between NPs, whether it is the purely extensional notion of 'referring to the same thing', or the notion of 'intended to be understood as referring to the same thing', which we supposed Lasnik could formulate.²⁹ But it seems that (40) infringes a rule which states when a term, normally a pronoun, can pick up its reference from another term—and this is an asymmetrical relation between the terms. To avoid confusion with the symmetrical relations, let us use the expression *t is referentially dependent on t'* to mean that *t* is

²⁹ A relation R is symmetrical iff, whenever *a* is R to *b*, *b* must also be R to *a*.

to be understood by being taken to have the same reference as *t'*. Now, this relation of referential dependence is quite different from, though it entails, that of intended co-reference; while two occurrences of the proper name *Oscar* or two occurrences of the pronoun *you* may be intended to be co-referential, neither occurrence is referentially dependent on the other. In our discourse (52), while the pronoun is intended to be co-referential with the second occurrence of the name *Oscar* (since it is referentially dependent upon the first occurrence, which is itself intended to be co-referential with the second), it is referentially dependent upon the first occurrence of the name, and not the second.³⁰

Using this notion, we can formulate the rule which (40) infringes as follows:

- (57) A term can be referentially dependent upon an NP if it does not precede and *c*-command that NP.

The precise formulation of the rule does not matter.³¹ The important point is that a principle of grammar must be explicitly concerned with the relation of referential dependence between pronouns and antecedents. The grammar to which (57) belongs will have to provide a list of expressions which can be referentially dependent upon other expressions—it will include third person pronouns as well as pronominal epithets, but will exclude first and second person pronouns as well as proper names. It will have to explain the semantic significance of the relation of referential dependence, and will therefore have to recognize as one among

³⁰ It seems clear that we must be able to speak of a pronoun's being referentially dependent upon one rather than another occurrence of a term. If we compare the discourses (i) and (ii),

(i) What do you mean, Oscar loves no one? He loves Oscar

(ii) What do you mean, no one loves Oscar? He loves Oscar

it seems clear that the second is less acceptable than the first, which can only be explained by allowing that the pronoun is picking up its reference from the previous rather than the subsequent occurrence of the name, and that (as is well known) it is easier for a pronoun in a subject position to pick up its reference from a term in subject position than from one in object position.

³¹ In particular, I make no allowances for the fact that a pronoun which is to be referentially dependent upon a prior occurrence of a term which stands in a certain grammatical relation to it must be a reflexive rather than an ordinary pronoun, and I have consistently ignored the need for number and gender agreement.

other interpretations of a sentence like (3), that it is true iff John loves John's mother. Since it will also have to recognize that pronouns can be used to make independent reference to salient objects, the grammar must itself distinguish between $\leftarrow he$ and $he\uparrow$, and will treat a sentence like (3) as ambiguous—in the sense that interpretations can be provided for it which exploit different rules of the grammar.

So far, I have been concerned to show that Lasnik's grammar, which combines a pragmatic theory of the reference of all pronouns with a Non-co-reference rule, is not adequate. I have not yet examined the main argument he offers for his position. The structure of his argument is this: any grammar which incorporates such a rule as (57) must also incorporate a Non-co-reference rule like (42), but the latter rule will suffice, on its own, to reject all the ungrammatical sentences. Lasnik states his argument succinctly in response to a suggestion of Wasow. He considers the sentence (58),

- (58) I told him that John was a jerk

which he claims is ungrammatical when *him* is co-referential with *John*, and says:³²

True, his [Wasow's] anaphora rule will not apply in (58). But Wasow explicitly stipulates that 'the failure of two NP's to be related anaphorically does not entail that they have distinct referents'. Here again, no provision is made for disallowing 'accidental' co-reference. I argued earlier that regardless of whether there is a pronominalization transformation or an interpretive co-referenced rule, a non-co-reference rule is required to account for the ungrammaticality of such examples as (58). In fact, I concluded that both types of co-reference devices can be dispensed with . . . and that the non-co-reference rule cannot be dispensed with.

We have already seen considerable deficiency in Lasnik's claim that a Non-co-reference rule *suffices* to account for the grammatical sentences, but we are now in a position to see what is wrong with the argument that is supposed to establish the *need* for a Non-co-reference rule.

I understand Lasnik's idea to be this. Any grammar with a co-reference rule like (57) will have to acknowledge the existence of pronouns used to make an independent reference,

³² Lasnik, *op. cit.*, p. 13.

and hence will have to allow that a sentence like (58) might involve such a use of the pronoun. And such a pronoun can be used to refer to anything. Thus, unless it is explicitly prohibited by a Non-co-referential rule, one possible interpretation of (58) will be that on which *him* refers to John, i.e. is co-referential with *John*. (This is what Lasnik means by 'accidental co-reference'.) But, according to Lasnik, such a use of the pronoun will result in an ungrammatical sentence.

This argument seems to me to rest upon a confusion between the three notions associated with the term *co-reference* which we have been at pains to distinguish. The sentence (58) is ungrammatical if, and only if, the pronoun is intended to be *referentially dependent* upon the later occurrence of the name *John*. Since our rule (57), and Wasow's parallel principle, expressly prohibit such dependence, there is no possibility of such an interpretation of the sentence being declared acceptable. Now, alternatively, the pronoun in (58) may be used to make a reference to John, but one which is independent of the occurrence of *John* in (58). There are two cases to consider, but neither of them results in an ungrammatical sentence, which must then be excluded by a Non-co-reference rule. In the first case, suggested by the notion of 'accidental co-reference', the pronoun is used to make a reference, possibly deictic, to a salient individual who turns out to be John. In this case, *him* and *John* in (58) are co-referential in merely the extensional sense, and the resulting utterance is in no way deviant. Another possibility is that the pronoun stands in the relation of intended-co-reference with *John*, although not referentially dependent upon it. This would be the result if the pronoun was referentially dependent upon some prior occurrence of the name *John* which itself stands in the relation of intended-co-reference with the occurrence of the name *John* in the sentence. Here again, strict ungrammaticality does not result, as can be seen from the discourse (similar to (52))

- (59) What do you mean John loves no one?
He loves John

Strict ungrammaticality is produced when and only when the

pronoun is intended to be referentially dependent upon that occurrence of the name which it precedes and commands. And (57) blocks any such sentence.

I have discussed Lasnik's argument in order to illustrate the importance of keeping in mind the difference between the three notions of *co-reference* that we have distinguished. To give one final example of the necessity for clarity on this matter, let me show how a major problem in the literature on anaphora dissolves when one attempts to formulate it in the vocabulary I have proposed. I shall let Lasnik introduce the subject.³³

I turn now to a complex phenomenon hinted at in Jackendoff and Postal and discussed at length in Wasow. Wasow points out that there is no reading of (60) on which the three italicized NPs can all be understood as co-referential.

(60) The woman *he* loved told *him* that *John* was a jerk.

However, it appears that independently, *he* can be co-referential to *him* and *he* can be co-referential to *John*.

(61) a. The woman *he* loved told us that *John* was a jerk.

b. The woman *he* loved told *him* that we were all jerks.

Lasnik goes on to say that a grammar needs a Non-co-reference rule to exclude the impossible reading of (60) without it, the reading would be allowed 'because of the transitivity of co-reference'.³⁴

Let us formulate this supposed problem using the notion of referential dependence. We start from the fact that there is no reading of (60) on which both *he* and *him* are referentially dependent upon *John*—presumably because there is no reading at all on which *him* is referentially dependent on *John*, as is predicted by (57). (61a) shows us that the first pronoun can be referentially dependent on *John* and (61b) shows us that the second pronoun can be referentially dependent upon the first pronoun or indeed, correctly construed in the first way, (61b) can be represented as (62).

(62) The woman *he** loved told *him that *we were all jerks.

Construed in the other way, it can be represented as (63):

- (63) The woman he → loved told him¹ that we were all jerks.

Presumably what matters for the supposed conundrum is that there exists the reading we have represented as (62), for then we have it that the first pronoun in (60) can be referentially dependent upon *John*, and that the second pronoun can be referentially dependent upon the first pronoun. Does it follow that the second pronoun can be referentially dependent upon *John*? Certainly not. Examples (61a, b) did not establish the possibility of the *simultaneous* dependence of the first pronoun on *John*, and the second pronoun upon the first. On the contrary, when we suppose the second pronoun to be referentially dependent upon the first, we suppose that the first pronoun is making an independent reference, and, thus construed, it cannot be referentially dependent upon anything. Therefore, when we formulate the problem with the notion of referential dependence, and when we make the natural assumption that if *t* is referentially dependent upon *t'*, then *t'* cannot be referentially dependent upon anything, the problem simply disappears.³⁵

³⁵ It has been suggested by a reader that my notion '*t* is referentially dependent upon *t'*' is equivalent to Chomsky's notion '*t* is an anaphor whose antecedent is *t'*'. This cannot be so for two reasons. First, Chomsky endorses Lasnik's pragmatic theory of the reference of (supposedly co-referential) pronouns, and, on that theory, it is not clear that it makes sense to speak of a pronoun's picking up its reference from one rather than another occurrence of a singular term, both of which serve to render their referent salient. Second, and more importantly, Chomsky speaks of quantifiers as antecedents of bound variables, when the relation of referential dependence is out of the question.

Things Without the Mind— A Commentary upon Chapter Two of Strawson's *Individuals*

What is the connection between the idea of an objective world and the idea of a spatial world? If someone has a conception of a world, something whose existence and operations are independent of his experience of it, must he thereby conceive of a system of spatial relations in which both he and the phenomena he experiences have a place? This question can be put another way. We can imagine a series of judgements 'Warm now', 'Buzzing now', made by a subject in response to changes in his sensory state, which have no objective significance at all. But we can imagine a similar series of judgements, prompted by the same changes in the subject's sensory state, which do have such a significance: 'Now it's warm', 'Now there's a buzzing sound'—comments upon a changing world. What is involved in this change of significance? In particular, if 'Now it's warm' is interpreted as a report on the world prompted by experience, must it be tantamount to: 'Now it's warm *here*'?

The connection between space and objectivity lies so deep in our conceptual scheme that many philosophers pass from 'objective' to 'outer' without even noticing the question they beg. The subjective being regarded as what is 'in the mind', the objective becomes what is 'without the mind', and then it is easy to say with Hobbes that if we have a conception of a thing without the mind, we have a conception of space.¹

In the second chapter of *Individuals*, Strawson probes this. From Zak van Straaten (ed.), *Philosophical Subjects: Essays Presented to P. F. Strawson* (Oxford: Clarendon Press, 1980). Reprinted by permission. I am grateful to Crispin Wright, David Pears, and John McDowell for reading an earlier draft of this paper and offering helpful comments.

¹ Hobbes, *Elements of Philosophy*, II, 7, ii.

conceptual connection with great subtlety and imagination. He makes us aware of the potential gap between 'objective' and 'outer'—a gap which he then attempts to bridge with a fascinating argument. Strawson illustrates the argument by imagining the situation of a being whose experience is wholly auditory. He argues that the concepts of an objective world, crucially the idea of existence unperceived, would not have any application in the experience of such a being unless that experience provides him with at least some analogue of space. Such an analogue can be provided in a purely auditory experience if each experience of a particular auditory phenomenon is accompanied by the experience of a master-sound—a constant sound whose variations in pitch enable the subject to give substance to the idea that he is moving. It is true that Strawson claims only that the master-sound provides an analogy of space. But it would be wrong to be misled by this, and by the fact that he labels the auditory universe a 'No-Space world', into thinking that Strawson is unsympathetic to the Kantian thesis that space is a necessary condition for objective experience. On the contrary, the chapter contains an argument for a slightly weakened version of that thesis.

Only part of the interest of the chapter lies here—in seeing how much is involved as an indispensable accompaniment to the idea of an objective world. The chapter is bold in another way. For if there is a frame of mind in which it is surprising how *much* is involved before the idea of objectivity can take root, there is an equally familiar frame of mind in which the surprise lies rather in how *little* a subject seems to need to think objectively. Certainly it is no part of Strawson's intention to derogate from the reality of things, but is our conception of our own world not just a little shaken by the thought that there could be a wholly auditory universe—by seeing how simply a tissue of reality can be woven out of regularities in experience?

I shall consider both these aspects of the chapter. I shall begin by considering whether Strawson successfully defends the Kantian thesis. After trying to show that the main line of argument is not successful, I go on to explore another line of argument also to be found in the chapter. In the final two sections, I elaborate two different reasons for doubting

whether a subject whose experience was wholly auditory could be regarded as having a conception of an independent reality.

Before proceeding, there is one important preliminary. What sense does the phrase 'objective world' bear in the questions Strawson took to define his investigation, and which therefore define ours? Strawson is explicitly not concerned with an idea of objectivity which rests upon interpersonal agreement. For the purpose of focusing upon a manageable section of our overall conceptual scheme rather than because it represents a genuine conceptual possibility, Strawson pretends that his subject makes no allowance for the existence of other observers. (Throughout, then, 'existence unperceived' is, effectively, 'existence unperceived by me'.) The idea which *does* concern Strawson is the idea of an experience's being *of* something distinct from it, and therefore the idea of something which is capable of existing independently of any experience of it.

Now Strawson dubs his investigation the elaboration of the conditions for a 'non-solipsistic consciousness'. But there is implicit in his notion of objectivity a restriction which precludes the application of the title 'theory of the objective world' to views which their proponents did not believe were solipsistic. The proponents of these views have a yet more general notion of objectivity in mind, according to which a theory has objective significance if it comprises propositions which, if true at a given time, are not true in virtue of the state of the subject at that time. This differs from Strawson's notion, for there may be no question of the subject's *experiencing* the reality which is constituted by the truths that do not belong in his biography.

For example, the subject's experience may be regular in a way which allows him to express various conditional or counterfactual propositions about what would be experienced were such-and-such else to be experienced. These propositions, if true at all, are not true in virtue of any actual occurrence in the subject's biography, and may be taken² as the

² They have been so taken by phenomenologists. See, e.g., C. I. Lewis, *An Analysis of Knowledge and Valuation* (La Salle, Ill.: Open Court, 1946), pp. 226-30, and *Mind and the World Order* (New York: Dover, 1956), pp. 135-9.

basis of a claim that an objective reality is thereby allowed for. However this may be, it is not an objective reality in the sense which concerns Strawson:

. . . I shall mean by a non-solipsistic consciousness, the consciousness of a being who has a use for the distinction between himself and his states on the one hand, and something not himself, or a state of himself, of which he has experience on the other.³

For whatever it is that makes these purportedly 'objective' propositions true, it is not something of which the subject can be said to have experience.

I

Intertwined in the chapters are several ideas on which a defence of the Kantian thesis might rest, but the main line of argument is one in which the need for space arises from the requirement that the subject of experience (hereafter 'Hero') be able to reidentify the objects of his experience. As Strawson summarizes the results of this chapter in a later one:

. . . we had to introduce, in auditory terms, an analogue of space in order to make room for the idea of reidentifiable particulars . . .⁴

A defence of the Kantian thesis on these lines must have two distinguishable stages; it has to be shown that the idea of reidentifiable objects is implicit in the idea of objectivity, and it has to be shown that criteria of reidentification, with the attendant distinction between qualitative and numerical identity, can only be framed in a spatial (or quasi-spatial)⁵ world.

On the first stage of the argument, Strawson has this to say:

. . . to have a conceptual scheme in which a distinction is made between oneself or one's states and auditory items which are not states of oneself, is to have a conceptual scheme in which the existence of auditory items is logically independent of the existence of one's states or of

³ P. F. Strawson, *Individuals* (London: Methuen, 1959), p. 69: my italics.

⁴ Strawson, *op. cit.*, p. 118.

⁵ I shall drop this qualification in what follows; except where the context indicates otherwise, I shall use 'spatial' in the weak sense of 'spatial or quasi-spatial'.

oneself. Thus it is to have a conceptual scheme in which it is logically possible that such items should exist whether or not they were being observed, and hence should continue to exist through an interval during which they were not being observed. So it seems that it must be the case that there could be reidentifiable particulars in a purely auditory world if the conditions of a non-solipsistic consciousness could be fulfilled for such a world. Now it might further be said that it makes no sense to say that there logically could be reidentifiable particulars in a purely auditory world, unless criteria for reidentification can be framed or devised in purely auditory terms. And if this is correct, as it seems to be, we have the conclusion that the conditions of a non-solipsistic consciousness can be satisfied in such a world only if we can describe in purely auditory terms criteria for reidentification of sound particulars.⁶

The second stage of the argument is not filled out in any great detail, but perhaps it is obvious how it would run. Hero has to be able to distinguish among later experiences of qualitatively indistinguishable phenomena those that are, and those that are not, later stages of the same phenomenon that he experienced earlier. This can only be done by taking into account the *relations* in which the phenomena stand; more specifically, relations which do not hold in virtue of the intrinsic non-relational character of the things related. And perhaps this may be seen as an abstract formal description of spatial relations.

Ingenious though this argument is, there is room for considerable doubt about its cogency, especially when it is interpreted in such a way that Strawson's auditory world provides an illustration of it. In that world, qualitatively identical sounds may be distinguished by their 'location at' (by being heard with) different pitch-levels of the master-sound. Given that this is the case, one may well feel that no genuine distinction between qualitative and numerical identity has been provided. Since different pitch-levels of the master-sound are qualitatively distinguishable, auditory presentations of numerically distinct sounds are never qualitatively indistinguishable, so long as a sufficiently inclusive view is taken of what to apply this concept to. But in the present context, a weaker point will suffice. If Strawson's Hero uses a genuine criterion of numerical identity, then space is not necessary

⁶ Strawson, *op. cit.*, pp. 72-3.

for him to do so. Since his criterion does not make any real use of the dimensionality provided by the continuous variation in the pitch-levels of the master-sound we may suppose with equal legitimacy that criteria of reidentification can be framed in an auditory universe in which an *unordered* series of master-sounds plays the same distinguishing role as the ordered series of pitch-levels of Strawson's single master-sound.⁷

The upshot of the weaker point is this: if the requirement deduced in the first stage of the argument can be satisfied in Strawson's auditory universe, the second stage of the argument cannot be completed. This difficulty seems to stem from the quite unparalleled role played by the master-sound in generating the space of the auditory world. Since change of position is not logically tied to a change in the subject's relations to the occupants in the space, but rather is a change in one particular aspect of his experience, there is no reason why that feature of experience to which it is tied should mimic the dimensionality of space.

In addition to wrecking the argument, the master-sound is the source of extreme disanalogies with our own system of spatial relations. The space of Strawson's auditory universe is an absolute space, and not a framework constituted by the spatial relations of its occupants.⁸ The fact that 'same place' in Strawson's auditory universe is not dependent upon 'same thing' is not merely a curiosity; it lays the whole scheme open to the most straightforward phenomenalist reduction, since 'God save the Queen is now playing at position *L*' is apparently equivalent to 'If master-sound of pitch-level *l* were heard, God save the Queen would be heard.' But, as the

⁷ I assume here and throughout that the existence of an intrinsic ordering between 'places' is essential to a space or a quasi-space. This certainly seems to be the framework within which Strawson was working: 'We want the analogy of Space . . . to provide for something like the idea of absence and presence—but not just of absence and presence in the most utterly general sense these words could bear, but absence or presence in a sense which would allow us to speak of something being to a greater or lesser degree removed from, or separated from, the point at which we are.'

⁸ It is perhaps surprising that Strawson should invent such a space in chapter 2, having emphasized in chapter 1 the mutual identification-dependence in our world, of things and places—a dependence which arises, on one side, from the fact that places are not intrinsically perceptible.

Carnap of the *Aufbau* learned to his cost, in a universe where space is constituted by the spatial relations between things, there can be no phenomenal characterization of being at a particular position, since things can move and change.

Taking the master-sound illustration seriously, we are prevented from completing the intended defence of the Kantian thesis, or are committed to regarding as spatial a scheme of thought so radically unlike our own as seriously to undermine the interest of the conclusion, were we able satisfactorily to defend it. Moved by these connected considerations, we must surely suspect that it is the illustration that is at fault, and not the argument which it is intended to illustrate. And further reflection appears to show the master-sound to be unnecessary even if an illustration of the argument is sought in a wholly auditory experience. For it may be held⁹ that a one-dimensional space which is genuinely parallel to our own may be constructed out of an auditory experience (or, indeed, out of any experience) provided that experience exhibits such order and regularity that its course can be seen as simultaneously due to the way the world is laid out and to the subject's continuous motion through it. Provided the subject's experience is sufficiently regular to enable him to establish short-term generalizations of the form:

An experience of kind *k* will intervene between any experience of kind *k'* and any experience of kind *k''*

it might appear that he can distil from its changing course a more or less detailed map of his world, with an object of kind *K* between (in a 'travel-based' sense of this word) objects of kind *K'* and *K''*. Possessed of such a map, the subject can make empirical sense of the distinction between a change in his position and a change in the world, and being able effectively to apply this distinction, he can make a revision of, and additions to, his map. That the resulting theory of the world has an interlocking, holistic character, as the subject simultaneously solves for the way the world is laid out and for the route he has taken through it, is no

⁹ It has been held by Jonathan Bennett, in his *Kant's Analytic* (Cambridge: Cambridge University Press, 1966), p. 37.

objection since it might reasonably be taken to mirror a feature in our own scheme. The holistic character of the resulting theory certainly renders it immune to any simple phenomenalist reduction.¹⁰

More important for our immediate purpose, this way of introducing a spatial order into an auditory world seems to provide a much better illustration of the themes of Strawson's argument. Distinct but qualitatively indistinguishable sounds can now be distinguished by their positions in the 'travel-based' ordering, and, no matter how wide a segment of experience is taken, distinct sounds can present genuinely indistinguishable appearances. Furthermore, the second stage of the argument is not undercut since we run no risk of dimensionless parallels. To construct a travel-based space is necessarily to construct an *ordering* of the objects or phenomena whose relations constitute the space.

In view of these considerations, let us suppose that Strawson would accept this as a better illustration of his argument. Eliminating the master-sound enables us to concentrate upon the nub of the argument—the connection between objectivity and reidentification. Here, too, I believe there is reason for scepticism.

My first objection, which does not run very deep, concerns the role which the concept of reidentification plays in the argument. The theories with which Strawson is concerned incorporate the idea that the subject has experience of phenomena which are independent of his experience of them. From this general acknowledgement of the independence of the world, Strawson is surely entitled to extract the corollary that the temporal dimensions of an experienced phenomenon may diverge from the temporal dimensions of any experience of it. Further it seems reasonable to suppose that Hero must be able to understand one particular application of this general idea, namely the possibility that the experienced phenomena should continue throughout a gap in his experience of them. Strawson expresses this idea in the language we use to talk about the persistence of material bodies, so that Hero is to think that he has experience of auditory *items* which continue to *exist* while

¹⁰ An illustration of this point will be developed later; pp. 266–8.

unobserved and which may therefore be *reidentified*. But it is not clear that the concept of identity need be involved here at all, still less that it need be involved in just the way it is involved in our scheme of three-dimensional bodies.

In the first place, it seems that this particular application of the possibility of temporal divergence between phenomenon and experience can be captured with the concept of *continuity*, as it occurs in the thought that it may continue to rain after one falls asleep, or that it may rain continuously between the time one falls asleep and the time one awakes. It would appear that the idea of its raining continuously is prior to, and independent of, the idea of a *single rainstorm*, and that it is possible to enrich what Strawson calls a 'feature-placing' language with an operator having the force of 'continuously' without disturbing its ontological simplicity—without introducing quantification over, and reidentification of, particulars. If this is so, a theory of an objective world can be couched in a feature-placing language, and the concept of identity does not belong in Strawson's argument at all.¹¹

Aside from this, there is a worry which arises even if we suppose that our Hero does express the idea of continuous sound by using the concept of identity. Let us suppose that Hero registers the independence of the world by allowing for the possibility of there being later, unheard parts of the same sound of which he has heard an earlier part, and therefore for the possibility of his hearing still later parts of the same sound of which, before an interruption, he has heard an earlier part. This is still not to think of an auditory *item* which *persists through time*, but rather of an auditory *process* which *is extended in time*. If the concept of reidentification is to be used in connection with processes, it must be understood that it is being used in a different sense from that which it has in connection with things. We reidentify a process

¹¹ In view of Strawson's defence of the coherence of a feature-placing scheme of thought later on in the book (pp. 202–13), it is perhaps surprising that he gives the impression in this chapter that he has demonstrated that any coherent scheme of thought about the objective world must involve the idea of reidentifiable particulars. Strawson does occasionally appear to restrict his question to schemes of thought involving particulars (as in: 'What are the most general stable conditions of knowledge of objective particulars?'—p. 62) but to defend the formulation of the argument by taking this restriction seriously is to do so at the cost of significantly reducing its interest.

when we hold that an occurrence encountered at one time is *part of* the same process as an occurrence encountered at another, but it is a distinctive (and some have thought incoherent) feature of our conceptual scheme of material bodies that we suppose an object to be both present *as a whole* on one occasion, and literally identical with an object present as a whole on another.

Now, a concept of reidentification can be used in connection with processes, and Strawson's detailed discussion of the auditory universe makes it clear that this is the concept which he intends.¹² Nevertheless, in the absence of any explanation or qualification, the use of the concept of reidentification which was originally introduced in the description of our conceptual scheme of material bodies carried with it the suggestion of a greater parallel between that scheme and the scheme of the auditory universe than is strictly warranted—a parallel which can only be purchased at the cost of ignoring the considerable difference between things and processes.

Identity is, indeed, a double irrelevance, since it appears not only that continuity can be registered without it, but also that identity can be recognized in the absence of continuity—at least where processes are concerned. We can quite intelligibly hold that later occurrences are parts of the same extended game as earlier occurrences to which they are not linked by any continuous series of game stages. This point has been raised in objection to Strawson's argument, for it seems to open the possibility of thinking reidentifyingly without thinking objectively.¹³ The obvious reply is that the criteria of reidentification mentioned in the argument must be restricted to those which require continuity; but if this reply is made the concept of identity once again drops out of the argument, in favour of that of continuity.

The reason why I said that this objection was not very

¹² '... identified as *part of* the same particular M as that of which the previously heard instance of A was a *part* . . .' (p. 70); 'There is a clear criterion for distinguishing the case of hearing a *later part* of a particular unitary sound-sequence of which the *earlier part* has been heard previously . . .' (p. 77) etc. etc. (My italics.)

¹³ See Don Locke, 'Strawson's Auditory Universe', *Philosophical Review* 70 (1961), pp. 518–32.

deep was that it seems simply to invite a restatement of the argument in terms of the concept of continuity. Hero must be able to make sense of the idea that, after his perception has ceased, there should be ϕ -ing stages continuous with the ϕ -ing stages which he perceived, and therefore that some such stage may be encountered later. But not just any later ϕ -ing stage is continuous with a given earlier one. So, Hero must have a way of drawing a distinction between those later ϕ -ing stages that are, and those that are not, continuous with a given earlier stage. Is this distinction not exactly parallel with our distinction between qualitative and numerical identity, and will it not similarly presuppose space?

I do not believe this argument is successful, for it seems to beg the question against a No-Space world in a subtle but decisive way.

In a spatial world there is no absolute notion of (temporal) continuity; we can only speak of spatio-temporal continuity. Now, in order to affirm upon the basis of a later perception of ϕ -ing that the ϕ -ing one experienced at t did continue (did have later stages) one has to be sure, not merely that the later ϕ -ing is continuous with *some* ϕ -ing in existence at time t , but also that it is continuous with the particular ϕ -ing experienced. For, in a spatial world, and possibly only in a spatial world, there can be distinct but simultaneous instances of the same universal. Thus, to be sure that the toy I saw at time t survived until time t' , it is not enough to be sure (i) that the indistinguishable toy I see at t' was in existence at time t , one must in addition be sure (ii) that the route by which it arrived at its position at t' was one which started from its occupation at t of the position in which I saw a toy.

Compare this with a very much simpler way of thinking—one which does not admit of the possibility of distinct but simultaneous instances of the same universal. The objectivity of ϕ -ing is recognized in this scheme (that is to say, it is intelligible that it be ϕ -ing when no ϕ -ing is perceived) but if, at any time, ϕ -ing is perceived, then that is all the ϕ -ing that the universe affords. Justice is done to this conception of reality by utterances of the unrestricted form 'It's now ϕ -ing'. Now, if Hero thinks in these terms, he will

certainly make sense of the idea that the ϕ -ing that he is perceiving might continue, but by this he means no more than that it may ϕ for all times between the time at which his experience of ϕ -ing ceases, and some later time. And while in the spatial scheme he would have two questions to ask upon encountering ϕ -ing after a gap, now he has just one: 'Was this ϕ -ing going on continuously between the time I ceased, and the time I began, to experience ϕ -ing?', for by this he means: 'Was it ϕ -ing for all times between the time I ceased, and the time I began, to experience ϕ -ing?' In this crude way of thinking there is no parallel to the second question of the spatial scheme, no analogue to the distinction between qualitative and numerical identity, and therefore no need for a criterion employing quasi-spatial considerations to assist Hero in drawing it.

If this is correct, then the space Strawson extracted out of the concept of objectivity is the space he smuggled into it, by limiting his attention to those theories of the objective that allow for distinct but simultaneous instances of the same universal. It is not surprising if such theories can be shown to be implicitly spatial; it is precisely for this reason that we were prepared to allow that the second stage of the argument might be completed. Perhaps there is some hidden incoherence in the crude and limited way of thinking, but that has to be shown, and if it can be shown the Kantian thesis can be established directly and Strawson's argument becomes an unnecessary circuit.

It is true that there is a distinction which someone thinking in these very crude terms must understand: namely, between the case where it is, and the case where it is not uninterruptedly ϕ -ing during a gap in his ϕ -experience. For in this sense, it remains true that 'not just any later ϕ -ing stage is continuous with a given earlier one'. But space does not appear to be involved in this distinction in any obvious way, and if it is involved in some unobvious way, this also has to be shown. It certainly cannot be shown by gesturing towards a distinction in which space is involved—the distinction between qualitative and numerical identity—but to which the required distinction is not remotely analogous.

To defend the Kantian thesis, the idea of space must be

shown to be implicitly involved in the very idea of existence unperceived, even as it is embedded in such a purported scheme as this. It is possible to find in Strawson's chapter materials for another line of defence of the Kantian thesis—a line of defence which would have just this effect.

II

Strawson suggests that thinking of an auditory experience as experience of an objective world confronts our Hero with the problem of 'making sense of' the idea of sounds existing unperceived. He maintains that 'the most familiar and easily understood sense in which there exists sounds which I do not now hear is this: that there are places at which those sounds are audible but at which I am not now stationed.'¹⁴ Space is clearly one way in which this difficulty can be resolved. Various other ways in which we 'make sense of' the idea of unheard sounds are mentioned, namely those which bring into play the idea of one sound *drowning* another, and the idea of deafness, but Strawson maintains that Hero cannot make use of them.

This is obviously a sketch of a line of argument rather than the argument itself, and in the text it is woven together with the argument we have just considered in a way which makes it difficult to disentangle, but I think that it is interesting and distinct, and I shall try to elaborate it. What, then, is the problem, and why should space be thought indispensable to its solution?

Hero must be able to understand the hypothesis, even if, in fact, he never believes it to be the case, that the phenomena of which he has experience should occur unperceived. Now, the idea of unperceived existence, or rather the idea of existence now perceived, now unperceived, is not an idea that can stand on its own, stand without any surrounding theory. How is it possible that phenomena *of the very same kind* as those of which he has experience should occur in the absence of any experience? Such phenomena are evidently *perceptible*; why should they not be perceived? To answer this question, some rudimentary theory, or form of a theory of perception

¹⁴ Strawson, *op. cit.*, p. 74.

is required. This is the indispensable surrounding for the idea of existence unperceived, and so, of existence perceived. (It is not to be thought that the idea of existence unperceived is an additional hurdle to be surmounted after the idea of existence perceived has been understood; the two ideas are sides of a single idea: the idea of an objective world.)

The same point can be put in other words. We might pretend for a moment that we are tracing the development, in a child's thought, of an utterance 'It's ϕ -ing', originally tied to a recurring pattern of his experience—a cry with which experiences of a certain kind are greeted.¹⁵ For an utterance like 'It's ϕ -ing', originating in this way, to become an assertion about an objective world, it must loosen its tie with experience, so that it makes sense to suppose that it is true even when no experience occurs. But, although it must loosen its tie with experience, the tie must not be severed; that which is potentially true in the absence of any experience must be the very same statement as may, on occasion, be affirmed upon the basis of experience. There must be no question of allowing for 'It's ϕ -ing' to be true in the absence of experience by introducing a new sufficient condition for its truth, unconnected with its existing basis. This would merely produce ambiguity, so that what is required would not yet have been accomplished—sense has not been made of the idea of the very same state of affairs that is on occasion experienced obtaining in the absence of experience. Now, we can detach 'It's ϕ -ing' from experience, without pulling the concept apart, only if that in virtue of which 'It's ϕ -ing' is true is connected with experience by some condition which is sometimes, but not always, satisfied. The proposition 'It's ϕ -ing' will then be understood to entail that, if that condition is satisfied, it may be perceived to be true. In the formulation of the condition there lies a theory, or the form of a theory, of perception.

Provided that he is capable of telling whether or not this condition is satisfied, such a connection with experience allows Hero to give empirical content to the supposition that it is now ϕ -ing, irrespective of whether he currently

¹⁵ I do not mean to suggest that this is the way all concepts of the objective world originate; far from it. See pt. III below.

perceives that it is ϕ -ing. If it is true that it is now ϕ -ing, then it must be the case that if the condition is satisfied, he will perceive it to be ϕ .¹⁶

This, then, is what it is to 'make sense of' the idea of existence unperceived. And the requirement that Hero have a conception of the world of sufficient complexity to enable him to understand why what is perceivable should sometimes be, and sometimes not be, perceived will surely rule out some purported theories of the world of excessive simplicity. But have we any reason for thinking that it will rule out all the spatial theories; that 'the most familiar and easily understood sense' in which there exist unperceived phenomena is the only sense?

There are two kinds of explanation of why a perceptible phenomenon may not be perceived, if we exclude those that obviously rely upon spatial notions, such as the observer's being in the wrong position, or having the wrong orientation, or there being something in the way. There are those that cite deficiencies in the perceiver, such as that he is inattentive, unreceptive in the proper modality, unconscious, or asleep. And there are those which cite the absence of factors in the world which are causally necessary for perception, as the absence of light is cited to explain why we cannot see a table. For reasons which I try to explain in the next section, explanations of this second kind do not represent a very promising avenue for exploration, if we are trying to find a non-spatial way of making sense of existence unperceived. Very briefly: we can make sense of the idea of a *material* object or substance existing in the absence of conditions causally necessary for its perception, but we cannot do the same for sensory objects; a rainbow cannot exist in the darkness, even if, were there to be light, a rainbow would be visible. And it appears that, if our Hero is to think of his

¹⁶ Do not say: empirical sense has not been given to the supposition that *it is ϕ -ing unperceived*—i.e. that it is ϕ -ing and the opportunity to establish whether or not it is ϕ -ing is not taken up. All that can be required is that empirical content be given to the hypothesis that *it is ϕ -ing*, and in such a way that it may be the case that it is ϕ -ing unperceived. The supposition that *it is ϕ -ing unperceived* is one whose intelligibility is consequential upon such a way of giving sense to the simple supposition that it is ϕ -ing, but it is obviously not in its turn a supposition that needs to be given empirical content in the sense of conditions under which one can conclusively establish that it obtains.

experience as of a world, and this world is not to be a spatial world, it will be a world composed of phenomena analogous to our sounds, smells, and rainbows, rather than to our material substances.

But there does not appear to be any parallel reason why Hero should not make sense of the idea of unperceived sounds (or, more generally, phenomena) by thinking in terms of some block of unreceptivity in himself. 'Perhaps', he thinks, 'there are sounds which I do not now hear, because I am unreceptive; if I was to become receptive, I would be able to hear them.'

Let us suppose that this abstract form of a theory is filled out in the following way. Let us suppose that Hero's prior experience had been of the unceasing auditory sequence 'tick tock tick tock . . .', but that upon one occasion the sequence in experience had been 'tick tock tock tick . . .' Why should Hero not use ordinary canons of scientific inference (let us not enquire too closely into what they are) to hypothesize that there was a tick which he did not hear, and understand this hypothesis in turn by supposing that he must have been unreceptive?

The objection Strawson would make to the coherence of this scheme of thought must be gathered from the following extract drawn from the passage in which he discusses various ways in which we make sense of the idea of unperceived sounds:

Alternatively, they turn upon such an idea as that of failing sensory powers. But why do we think of our powers failing rather than the world fading? This choice cannot be used to explain a conception it presupposes.¹⁷

I am not at all sure what objection Strawson has in mind here, but perhaps it is this. 'If one asks oneself why, in any particular case, one supposes that one's sensory apparatus is defective, it is clear that such a judgement cannot rest only upon internal features of one's experience (e.g. hearing nothing, or hearing things fainter and fainter), since it cannot be logically ruled out that there should be nothing to hear, or that what there is to hear is getting fainter and fainter. A judgement that one's sensory apparatus is defective must

¹⁷ Strawson, *op. cit.*, p. 74.

rest upon a view that this or that thing is there to be heard. Since the view that one's apparatus is defective ('this choice') must rest upon ('presuppose') a view as to what objectivity exists, it cannot be used to give the indispensable surrounding ('explain a conception') which the idea of what objectively exists was discovered to require.'

If this is the objection, then it appears wrong in principle. It is right to insist that all of the elements of the theory of an objective world should be present, but wrong to insist that they be independently intelligible. It is true that the idea of a perceptual breakdown presupposes the idea of an objective reality, and that, upon the envisaged scheme at least, the idea of an objective reality presupposes the idea of a perceptual breakdown (or lack of receptivity). The ideas form a circle, and any theory constructed with their aid will have a holistic character as a result. Propositions about how the world is will be derivable from propositions about the course of Hero's experience only when they are taken together with propositions about when he was, and when he was not, receptive, while propositions of this latter kind will in their turn depend both upon propositions about what Hero is (or is not) experiencing, and also upon propositions about what there is to be experienced. Hero must see the course of his experience as simultaneously determined by the way the world is and his changing receptivity to it; each is connected to experience, but only as modified by the other. All this is correct. What is not correct is that there is anything objectionable in principle in such an arrangement.

The best possible reason against objecting to such a structure in Hero's theory is that it can also be discerned in a spatial theory. In the modified version of a spatial scheme in the auditory universe, Hero can tell that he has changed position by the changing course of his experience, but only when this is taken together with a map of a fairly stable world. But that map, in its turn, can only have been established, and must constantly be revised, by Hero's adopting views as to where and when he is moving. (In the somewhat Cartesian setting of the auditory universe, in the absence of other subjects to whom Hero and his movements can be objects of perception, all it can mean for him to be at a position is for

him to perceive what is audible at that position.) The parallel between these two kinds of theories is not complete, since, in the place of an absolute notion of receptivity—present or absent at a time—the spatial theory effectively employs a relativized notion: receptive to (= located at) this or that position. But this greater complexity does not prevent its central concepts from having that interlocking character which would appear to expose it to Strawson's criticism, if that criticism was well founded.

In fact, a stronger point might be suggested, namely that it is not merely permissible, but positively necessary, for that condition which is to account for the presence or absence of perception to be connected *a priori* with, and therefore, known to be satisfied only upon the basis of, propositions about the way the world is. For it was precisely this feature of the revised spatial theory of the auditory world which secured for it an immunity to that simple phenomenalist reduction which threatened the master-sound theory. (In the master-sound theory, the subject's changing position is definitionally tied to a change in one phenomenally identifiable aspect of his experience.)¹⁸

Although each person in a large circle of people can be sitting upon the knees of the person behind him, this is not a feat which only two or three people can manage. Perhaps the objection is not that there is a circle in a theory relying upon deafness or unreceptivity to give sense to the idea of existence unperceived, but that the circle is too small. For, while it is true in the spatial theory that deciding whether or not one has moved (and therefore deciding whether a change in one's experience signals objective change at some given place) requires taking as given certain propositions about the way the world is, these are not the very propositions about the world whose truth one is required to establish, but rather propositions about how it is with adjacent places. (This reveals another part of the structure of the theory: the subject can only move continuously through space.) Of course doubt might be raised about the condition of these adjacent places, which could be resolved in the same way provided weight is shifted on to knowledge

¹⁸ I shall discuss the significance of this irreducibility below; see pp. 288–9.

of still other places; an indefinite series of such challenges could bring us back to the place from which we started. However, the theory, though interlocking, has enough structure to get off the ground; one who holds it can meet a challenge parallel to the challenge in Strawson's rhetorical question:

But why do we think our position changing rather than the world changing?

Compare this with the scheme using receptivity. Hero is supposed to be able to make sense of the idea of its now ϕ -ing unperceived by relying upon the concept of receptivity, and in particular, by supposing that, if he were now to become receptive, he would perceive ϕ -ing. However, this only gives content to the idea that a sound now exists unperceived if there is some criterion of Hero's now becoming receptive, other than his perceiving ϕ -ing. But what could it be?

Equally, Hero was supposed to be able to understand the hypothesis that there was an unperceived tick by using the supposition that he was unreceptive. But the past regularity in his experience cannot be regarded by Hero as conclusively establishing the hypothesis that there was an unperceived tick. If it does, this would not be because Hero had made the uniformity of nature a logical truth—no one could do that. It would simply be that he had established a new, and independent, sufficient condition for the statement that there is a tick, and therefore failed to give sense to one and the same state of affairs existing both perceived and unperceived. But, if the inductive considerations fail conclusively to establish the hypothesis of an unperceived tick, then Hero must be able to distinguish, at least in thought, between the case in which the regularity was perpetuated, and the case in which it was not. Here we should like Hero to be able to appeal to the counterfactual conditional 'If I had been receptive, I would have/would not have heard a tick'. But such a conditional is quite vacuous if the only possible conception that he can have of his being receptive at that time is simply that of being able to hear what is there to be heard.

Here, surely, are the materials for a possible line of defence of the Kantian thesis—a line of defence which rests upon the

idea that only a spatial theory can satisfy the demand that the factor accounting for the presence or absence of perception of perceptible phenomena should be at once *a priori* connected with the propositions about the world, and yet subject to significant empirical control. I shall not now pursue this line any further; to do so would involve the consideration of a variety of alternative schemes¹⁹ in a detail which is not perhaps commensurate with their interest. The principle of the argument should be clear, and if it is clear, we have perhaps derived such illumination of the role of space in *our* thinking as it is in the power of this, or any, defence of the Kantian thesis to afford. After all, as Strawson himself emphasizes, this is the object of the exercise.

Instead, in the remaining two sections, I want to turn to what must be a brief, and I fear rather dogmatic, consideration of the question of whether a coherent theory of an objective world can be constructed upon the basis of an experience that is wholly auditory, even when that experience exhibits whatever degree of order and connectedness is necessary for the subject to apply 'travel-based' spatial notions to it. There does not appear to be any concept parallel to that of matter or material substance which can be framed in the auditory universe. The first question I want to ask is: Can there be a world without substance?

III

It seems possible to draw a distinction between two kinds of properties which objects may have, though a complete elaboration and defence of this distinction would be a very difficult task. Into the first kind fall those properties which are dispositions to affect sensitive beings with certain experiences—these we might call *sensory properties*, or, in deference to a long-standing tradition in philosophy, *secondary properties*. For an object to have such a property is for it to be such that, if certain sensitive beings were

¹⁹ A spatial scheme is not the only scheme to employ a relativized receptivity condition, with the possibilities of additional empirical control which that provides: we can make sense, perhaps, of the idea of being ϕ -receptive, where receptivity is relativized to a universal. And there are other possibilities.

suitably situated, they would be affected with certain experiences, though this property may, in its turn, be identified with what we should normally regard as the ground of the disposition. However, in the first instance, a sensory property is a dispositional property.

It is not necessary for our immediate purpose to have any other characterization of primary properties than as non-sensory properties of objects. So defined, the class is extremely heterogeneous. What is important, though, is that the properties constitutive of the idea of material substance as *space-occupying stuff* should be acknowledged to be primary. These include properties of bodies immediately consequential upon the idea of space-occupation—position, shape, size, motion; properties applicable to a body in virtue of the primary properties of its spatial parts; and properties definable when these properties are combined with the idea of force (e.g. mass, weight, hardness). The way these properties relate to experience is quite different from the way sensory properties relate to it. To grasp these primary properties, one must master a set of interconnected principles which make up an elementary theory—of primitive mechanics—into which these properties fit, and which alone gives them sense. One must grasp the idea of a unitary spatial framework in which both oneself and the bodies of which one has experience have a place, and through which they move continuously. One must learn of the conservation of matter in different shapes, of the identity of matter perceived from different points of view and through different modalities, and of the persistence of matter through gaps in observation. One must learn how bodies compete for the occupancy of positions in space, and of the resistance one body may afford to the motion of another. And so on.

To say that these primary properties of matter are theoretical is not to explain or to mystify, but to highlight an analogy between the way our grasp of them rests upon implicit knowledge of a set of interconnected principles in which they are employed, and the way our understanding of such a property as electric charge rests upon explicit knowledge of a set of propositions more familiarly regarded as a theory. Certainly, to deny that these primary properties are *sensory*

is not at all to deny that they are *sensible* or *observable*, for we are obviously able, after the appropriate training, to perceive the shape, motion, and hardness of things. The point is rather that it is not possible to distil the concept of hardness solely out of the experiences produced by deformation of the skin which is brought into contact with a hard object, for it is not possible to distil out of such an experience the theory into which the concept fits. It is no more possible to have a purely sensory concept of hardness than it is to have a purely kinaesthetic conception of what it is for one's legs to be crossed, or to have a purely muscular conception of the motion of one's body, or to master the concept of electricity solely by learning to recognize electric shocks. And, though this is less obvious, it does not appear to be possible to regard the conception of the shape of a material thing—with all the propositions about its characteristic behaviour and interaction with other bodies which that implies—as the same as whatever shape concepts might be grounded in the colour mosaic thought to be given in immediate visual experience. This would certainly seem to be suggested if we can demonstrate, as I believe we can, that the blind are capable of a perfectly adequate mastery of shape concepts, and of spatial concepts generally, for no single *sensory* property can be defined in relation to different senses.²⁰ Berkeley²¹ and Mill²² both saw this point but, constrained by a theory of concept formation that would not allow for the formation of ideas of any other than sensory properties, they concluded that these spatial concepts were sensory concepts, but related to the sense of touch as heat is. This is possible only by supposing that the concept of solidity is a sensory concept, as we have seen that it is not, and further, by supposing that the concept of the motion of the subject, and of the parts of his body, are concepts of kinaesthetic experience, which they are not.²³ I wish to stress my divergence from Berkeley and

²⁰ The issue of the spatial concepts of the blind surfaces again in pt. IV below.

²¹ Berkeley, *An Essay Towards a New Theory of Vision*.

²² J. S. Mill, *An Examination of Sir William Hamilton's Philosophy* (London: Longmans, 1872), pp. 270–313.

²³ Further, any attempt to explain spatial concepts in this way runs into the objection that it only provides 'serial' and not 'simultaneous' spatial concepts. For this distinction, and the significance of the objection, see pt. IV below. For

Mill on this point in order to guard against a serious misunderstanding. The distinction we shall presently find between Hero's conception of his world and our conception of the material world, does not at all reside in whatever difference is to be found between auditory, and tactuo-kinaesthetic experience. On the contrary, it resides in the difference between a conception of the world which is directly and exclusively, and a conception of the world which is neither directly nor exclusively, woven out of materials given in experience.

I do not take myself to be saying anything new in drawing the primary/secondary quality distinction in this way, since it is almost exactly the way Thomas Reid explains the distinction in his *An Inquiry into the Human Mind*. Consider, for example, the following remarks on hardness:

When the parts of the body adhere so firmly that it cannot easily be made to change its figure, we call it *hard*; when its parts are easily displaced we call it *soft*. This is the notion which all mankind have of hardness and softness: they are neither sensations nor like any sensation . . .²⁴

. . . hardness is a quality of which we have as clear and distinct a conception as of anything else whatsoever. The cohesion of the parts of a body with more or less force is perfectly understood, though its cause is not: we know what it is, as well as how it affects the touch. It is therefore a quality of a quite different order from those secondary qualities we have already taken notice of, whereof we know no more naturally, than that they are adapted to raise certain sensations in us.²⁵

In the words of a recent commentator:

Reid tries to show that our concepts of certain primary qualities are bound up in an elementary theory of bodies, a natively given primitive mechanics. Since the primitive mechanics gives us various ways of telling when and where the primary qualities are present, then we can develop and apply these concepts independently of our having all the corresponding sensations.²⁶

Mill's wrestling with the issue of simultaneity, and his conclusion ('The idea of Space is at bottom one of time') see *ibid.*, p. 278–83.

²⁴ *An Inquiry into the Human Mind* (ed. T. J. Duggan) (Chicago, Ill.: University of Chicago Press, 1970), p. 61.

²⁵ *Ibid.*, p. 69.

²⁶ Norman Daniels, *Thomas Reid's Enquiry* (New York: Burt Franklin, 1974). P. xiv; see also chap. 4.

With such an excellent and detailed discussion to appeal to, I feel easier in giving nothing but a very brief sketch of the distinction.²⁷

All it can amount to for something to be red is that it be such that, if looked at in the normal conditions, it will appear red. This formulation embodies what we might call the dispositional route from subjective experience to objective property, a route with which we are already familiar.²⁸ Philosophers have tried to provide a different account of what it is for a colour ascription to be true which does not so much involve a different route, as an attempt to make the most direct possible leap from subjective experience to objective property. They have tried to make sense of the idea of a property of redness which is both an abiding property of the object, both perceived and unperceived, and yet 'exactly as we experience redness to be'. By concentrating upon one's experience of colour, one is supposed thereby to know what it is for an object to have this property: 'This', one is to say, referring neither to the experience nor to any primary property of the thing, 'this, just as it is, can exist in the absence of any observer'.²⁹

But the leap gets us nowhere, for it inevitably involves an attempt to make sense of an exemplification of a property of *experience* in the absence of any experience. Wittgenstein once imagined a world in which there were places which affected everyone painfully, so that pains were located at places in the way we locate smells. Suppose this fantasy came true. Would it then make sense to give a non-dispositional

²⁷ This way of drawing the distinction also echoes recent work; on the inter-connection between the primary properties and the idea of space see A. M. Quinton 'Matter and Space', *Mind* 73 (1964); on the idea of secondary qualities as dispositions to affect us with experiences see J. Bennett, *Locke, Berkeley, Hume* (Oxford: Clarendon Press, 1971), chap. 4. I differ from Bennett in not making the dispositional character of the secondary qualities a matter of the meaning of sentences ascribing secondary qualities, but relying instead upon the obscurer notion of that in which their truth consists. It seems decisive against any dispositional account of the meaning of such a term as 'red' that the only way to characterize the experience red objects produce in us is as such.

²⁸ See above, sect. II.

²⁹ See, e.g., J. L. Mackie, *Problems from Locke* (Oxford: Clarendon Press, 1976), chap. 1. Mackie regards such a concept as intelligible, and used by the common man, though in fact Mackie himself does not believe there is any scientific use for it. In these views he follows Locke.

account of what it is for there to be a pain at such and such a spot; to suppose a 'pain as we feel it' existing in the absence of any observer? What can the latter form of words mean save that something awful is going on there, and how can that be, when there is no one who is hurt? To modify a dictum of Wittgenstein, conceiving of a pain which no one feels upon the model of a pain which one does feel is none too easy a thing to do.

We may ask a philosopher who claims to find intelligible the idea of an objective property extracted from our experiences of colour in this direct way, whether or not such a colour property can characterize an object in the dark. He can hardly say 'Yes', since it would be quite obscure how a 'colour-as-we-see-it' can exist when we cannot see it, and how our experiences of colour would enable us to form a conception of such a state of affairs. Further, it would have to be explained in what the difference between such an objective colour property, and the dispositional property, consists. Observing the results or switching on the light merely tests for the dispositional property; what could show whether or not objects did in fact retain these other colour properties in the dark? To maintain, on the other hand, that such colour properties cannot be true of objects in an unlit cellar seems to undermine the status of the property to being an objective property of a body, since it seems to depend for its existence upon the conditions necessary for the human perception of it. Further, the concept is *said* to be different from a dispositional property but it is difficult to see in what a grasp of the supposed residue would consist. Presumably, it is conceivable that objects which are not really 'red-as-we-see-them' should appear red to us; indeed, this appears to be the situation Locke supposed actually to obtain. But what one conceives, when one conceives that objects which appear red to us are, in addition, really red, or are, in addition, not really red, and how one might manifest, either verbally or behaviourally, these supposed conceptions, is quite opaque.³⁰

³⁰ These remarks are equally directed against those 'hard nosed' philosophers who wish to maintain that 'science has shown that objects are not really red'. Such a position would equally require the *intelligibility* of a non-dispositional concept directly fashioned from experience, which I am trying to deny.

The idea that objects lose their colour properties in the dark suggests a diagnosis of the position—it suggests that it arises because philosophers are being led astray by their imaginations. We must suspect that the philosophers who claim to find intelligible an objective but non-dispositional colour property try to conceive of an object's possessing such a property in the absence of any observer by imagining a red object which no one sees—a feat of the imagination which is impeded if part of the imagined story is that the object exists in a pitch-black cellar. Now, if the conceivability of an object's having such properties in the absence of any observer comes to this, then Berkeley's arguments against taking those imaginings as at their face value is decisive. What, after all, is being imagined but experiencing a red object unseen by anyone *else*?³¹

In our world, auditory properties are sensory properties, and auditory phenomena are sensory phenomena, and since they are imagined to rest upon the same basis in experience, they cannot be different for Hero, who inhabits a purely auditory universe. For both Hero and ourselves, the truth of a proposition to the effect that there is a sound at such-and-such a position must consist in this: if someone was to go to that position, he would have certain auditory experiences, or rather, to bring out the force of the conditional, if some-

³¹ This argument is found both in Berkeley's *Principles of Human Knowledge*, sect. 23, and in the first of the *Three Dialogues Between Hylas and Philonous*. In general I take Berkeley to have been quite correct to argue against Locke's contention that we can form an idea of a world existing independently of observers out of the ideas of primary properties which Locke provided, since they remained sensory concepts. Cf. M. R. Ayers, 'Introduction' to *Berkeley's Philosophical Works* (London: Dent, 1975) '... the only case of conceiving a thing's intrinsic properties that Locke can suggest... is conceiving of its "primary properties", the ideas of which, as he himself holds, are acquired through sense. Thus the real issue between them [Locke and Berkeley] is whether we get, or could possibly get a sensory concept through which we could conceive of reality in a sense-independent way' (p. xiii). Reid saw that Berkeley was right about this, but thought that, rather than reject the material world, we should reject the poverty of the mechanisms of concept formation which the empiricists recognized: 'The very existence of our conceptions of extension, figure, and motion, since they are neither ideas of sensation nor reflection overturns the whole ideal system by which the material world hath been tried and condemned...' (op. cit., p. 79). Actually, Berkeley's idealism doesn't follow immediately, even within the empiricist framework, since there is the 'dispositional route' from experience to objective property which the phenomenologists later explored.

one was to go to that position, he would thereby be caused to have certain auditory experiences.³² But there is a difference, for we have, and Hero does not have, the resources to make sense of the idea of the persisting categorical basis or ground of that disposition, in the object, or at the place, to which it is ascribed. Unlike Hero, we have the concept of substance, of space-occupying matter, for we have the concepts of the primary properties of matter.

In order to make this difference clear, we must make another difference clear. Michael Dummett has drawn attention to the fact that, for every proposition which is true, we like to think that there is something that makes it true, and further, that we do not like to think that conditional propositions can be *barely true*, that is to say, we do not regard that in virtue of which a conditional proposition is true as specifiable only by a repetition of that proposition. Now, there are two different kinds of non-conditional statements in whose truth the truth of a conditional may be taken to consist, and therefore two different kinds of ground for a disposition. The first kind of ground is simply the generalization whose truth we normally regard as evidence for the conditional. Hero's propositions about the world, true in the first instance in virtue of the truth of a conditional, need not be regarded as barely true, for they can be regarded as having a ground of this kind; the truth of the proposition 'There is now a ϕ -sound at p' can be regarded as true in virtue of the truth of the generalization 'Whenever in the recent past I have gone to p, I have had ϕ -experiences', or perhaps more complicated generalizations from which the conditional may be derived.

There is another kind of ground for a disposition, namely a relatively abiding property of the object to which the disposition is ascribed, capable of being characterized independently of the disposition, and therefore capable of providing, when taken together with the antecedent of the conditional (and perhaps certain other conditions understood as normal)

³² Since going to the position must produce the experiences if the place is to have the dispositional property, it is possible for Hero to make some sort of distinction between veridical and hallucinatory perception, the latter consisting of experiences which are not causally dependent upon his position.

a causal explanation of the occurrence which the consequent of the conditional reports. It is a ground of this kind which we suppose a disposition like fragility has in the arrangement and binding of molecules; more relevantly, we suppose that any disposition of a place to affect us with certain experiences has a ground of this kind in the occupation of that place by a matter of a certain sort. Now, Hero can have no idea of such a ground for the dispositions which places have to affect him in a certain way. He has no resources, or at any rate has not obviously been provided with resources, for forming the idea of any property of the world that is not a disposition of the world to affect him in a certain way.

It is important to appreciate the difference between these two kinds of ground for a conditional if we are to gain a proper understanding of phenomenalism. Sir Isaiah Berlin has objected to phenomenalism that it reduces categorical existence to the truth of subjunctive conditionals which are not, in their turn, grounded in anything else.³³ Dummett has replied, on behalf of the phenomenalist, that, provided he gives up bivalence for statements concerning material objects in remote parts of the world, there is no reason why he should suppose that any subjunctive conditionals are barely true, since he can maintain that propositions about the explored material world are true in virtue of observed regularities in our experience. But it is clear that Dummett is not offering what Berlin was missing, namely a ground of the second kind for these subjunctive conditionals—a relatively abiding property of an object or place which, together with a subject's presence, could be used to explain his experiences. Whether or not it is something that would or should worry the phenomenalist, Berlin certainly put his finger upon a deep conceptual prejudice of ours that is offended by dispositional properties without categorical grounds of the second kind. The sense of disquiet which we feel at the idea of two glasses which are exactly alike in all that is abiding, yet different in that if one is struck it will emit middle C, and if the other is struck it will not, is not at all diminished by citing the generalization upon which the proposition

³³ I. Berlin, 'Empirical Propositions and Entailment Statements', *Mind* 59 (1950).

may be asserted, namely, that whenever in the past one has been struck, it has emitted middle C, and whenever in the past the other has been struck, it has not. And this prejudice is equally offended by the idea of two places alike in what occupies them between visits, yet of which one is such that if one goes to it, one will have certain experiences, and the other is not.

This is just the situation that Hero must accept in his world; places have powers that cannot be identified with anything continuously occupying them, so that going to a place is just a basic, causally relevant factor in the explanation of the course of his experience. However, I do not want at this point to examine whether there is more to our resistance to such an idea than mere prejudice, but rather turn to this question: if this is the situation, can we continue to suppose that Hero has a coherent theory which incorporates the idea that he has experience of an objective world?

The notion of objectivity arises as a result of conceiving a situation in which a subject has experience as involving a duality: on the one hand, there is *that of which there is an experience* (part of the world) and, on the other, there is *the experience of it* (an event in the subject's biography). We have been exploring the consequences of this duality, especially the consequence that, though the temporal dimensions of these two elements overlap, they need not coincide. And, if the situation does comprise these two elements, they are not unconnected—they are not two distinct states of affairs existing simultaneously by accident or as the result of a pre-established harmony. Thus unconnected, the one could not be regarded as an experience *of* the other, as a way of gaining knowledge of it, and thereby, of the world of which it is a part.

Now, can these features be recapitulated in Hero's scheme—in a scheme where that in virtue of which the 'objective proposition' is true can only be generalizations about the past course of Hero's experiences? The answer is surely 'No'. We do not have two states of affairs existing simultaneously and related causally. All that exists at the position is Hero and his experiences. The only cause of Hero's having those experiences on going to that position *is* his going to that position. That which makes the 'objective proposition' true

cannot be cited as a cause. If it is regarded as a barely true dispositional property of a position, then it is ineliminably characterized in terms which logically connect it to the event that is to be explained. It is no improvement to consider the 'objective proposition' not as barely true, but as true in virtue of the past regularity in Hero's experience, for the regularity can hardly be regarded as something which causes those events which perpetuate it, nor can it sensibly be regarded as something Hero *experiences*.

In fact, without ideas corresponding to our ideas of the primary properties of matter, Hero cannot make sense of the same thing existing both experienced and unexperienced, for he cannot recognize any unexperienced *existence* at all. It may hold good of a place when no one occupies it that *if* one was to go there, one would be affected with such-and-such experiences, but *that* patently does not report any contemporaneous existence, nor, when we inquire into its ground, do we discover any contemporaneous existence there either; all we have are generalizations. Hero is forced to think as the phenomenalist would have us always think, and the remark Berlin made about the phenomenalist's scheme applies equally well to Hero's:

. . . what troubles the plain man is the thought that if the hypotheticals are unfulfilled, if no observers were in fact observing, then, if the phenomenalist analysis is in fact correct, there was . . . nothing at all.³⁴

We can think of sounds as perceptible phenomena, phenomena that are independent of us, and that can exist unperceived, because we have the resources for thinking of the abiding stuff in whose changes the truth of the proposition that there is a sound can be regarded as consisting. A fly is moving its wings; this is an event which we perceive, partly in an auditory way, and which is naïvely regarded as the ground of the proposition that, if one goes into a certain room one will hear a buzzing. (More knowledgeable thinkers would locate the ground in the movement of the air molecules which

³⁴ Berlin, *op. cit.* This passage makes it clear that Berlin was lamenting the absence of a contemporaneously existing categorical ground for the hypothetical propositions.

these events produce.) And the event of a fly's moving its wings is not even in the first instance a sensory phenomenon; it is an event consisting in space-occupying objects, possessed of qualities characterized independently of observers, moving in relation to one another. The sensory phenomena we typically recognize are, in fact, properties of things or stuffs—persisting space-occupying substances—in whose primary qualities, or primary quality changes, the disposition to produce experiences may be regarded as grounded. It is these substances which we perceive as coloured, or as making a sound.

There is a passage in the chapter in which Strawson seems to be expressing similar thoughts:

It helps us to think of one particular M being drowned or submerged by the stridencies which intervene . . . ; and thus to think that they were there to be heard, would have been heard but for these stridencies. But now we have only to think of the reasons, the evidence, we have for thinking something like this in real life—the visible but inaudible scrapings of the street violinist as the street band marches by, and then we lose interest in the suggested criterion for the case of the purely auditory world.³⁵

We find here the idea that what enables us to think of sounds as being drowned out, and in this way, existing though unperceived, is the knowledge that their categorical basis—the scrapings—continues; this is, on the whole, the point I have been making. I would want to guard against the misunderstanding that might be involved in thinking of the scrapings as particularly *visible*, as though the difficulty could be surmounted by providing Hero with a richer network of correlations between his experiences. But the main idea is surely there, together with the critical equation upon which it rests: 'was there to be heard/would have been heard'. I disagree with Strawson only in that I go further, and see the difficulty he points to as arising for Hero in *any* attempt he makes at the idea of an unheard sound. Exactly the same background is necessary for making sense of the idea of a sound's existing unheard at a place; the place must be occupied by an object characterized in other than sensory terms, and in whose states and doings

³⁵ Strawson, *op. cit.*, p. 71.

(scrapings) the existence of an unheard sound may be taken to consist.³⁶

This is the first respect in which I believe that Strawson does not provide the subject of auditory experience with a coherent conception of external reality—the conception is one constructed exclusively out of sensory concepts. However, even if it is now clear that such a conception of an independent reality is not possible, it may be less clear why the conception by a subject of auditory experience of his world must take this form. After all, I have stressed that our ideas of matter are independent of any particular kind of experience, for example, tactual experience. So why can there be no analogue to the idea of matter in an auditory universe?

The issues raised by this question are enormous, and I should not be able to deal with them in this paper, even if I knew how. If the hypothetical theory is to follow ours at all closely, sounds would have to *occupy* space, and not merely be located in it, so that the notions of force and impenetrability would somehow have to have a place, and we may well wonder whether we can make sense of this without providing Hero with an impenetrable body and allowing him to be an agent in, and manipulator of, his world. But perhaps this is the wrong line to pursue. Perhaps we should explore the possibility of a theory more closely analogous to the field theory of some physicists, or even a physical theory working on principles quite different from any we have knowledge of. Unsurprisingly, I cannot myself conceive of such a theory nor do I know how one might set about demonstrating its

³⁶ It might be worth pointing out that considerations similar to those of the present section bear upon the question of how Hero conceives of himself. Presumably Hero must be able to make sense of his *existing unperceiving*—located in space although asleep. In the first instance, Hero can understand such location in terms of dispositions: 'If I had been awake, I would have had experiences as of position *p*.' But if Hero's being at a position is to be cited in the explanation of Hero's experiences, when he does perceive the world, this disposition must have a ground which can be independently characterized, and this can presumably be found only in the occupation of that position by something which is identical with Hero, and, since it exists unperceived, something which is characterized by primary properties.

Strawson made a half-hearted attempt to provide Hero with an audible body (pp. 84–5) but, since it does not have primary properties, it does not meet the need here indicated.

possibility, nor whether a search for such a demonstration is even coherent. What I hope to have pointed out is the need for some physical theory over and above the ideas which Strawson attributed to his subject; not that such a supplementation cannot be provided. Though my conclusion is limited, I believe that it is worth drawing. For it is extremely tempting, upon first reading Wittgenstein's fantasy about locatable pains, to think: 'How simple! This is all that separates the inner from the outer—this is all that is required to make an object, and hence a world, out of experience.' Liberating and appealing though this thought may be, it does not appear to be correct.

IV

I shall raise the second doubt I have about Strawson's auditory 'universe' rather obliquely, taking as my starting-point a debate about the spatial concepts of the blind. For some centuries now, philosophers and psychologists have disagreed over whether and to what extent the spatial concepts of the blind are similar to those of the sighted. There have been those who have maintained that the blind do not, strictly speaking, have genuine spatial concepts at all; as Lotze puts it:

... the space of a blind man may not be so much what we mean by space, as an artificial system of conceptions of movement time and effort ...³⁷

This position is also taken by Platner:

In reality, it is time that serves, for the man born blind, as space. Remoteness and proximity only mean to him the time, more or less long, and the number, more or less, of intermediaries which he needs in passing from one tactual impression to another.³⁸

Occupying a middle position are those who agree that the distance concept of the blind is essentially a concept of the time and bodily movements necessary to traverse the distance.

³⁷ H. Lotze, *Metaphysic*, Vol. II (Oxford: Clarendon Press, 1887), pp. 272–3.

³⁸ E. Platner, *Philosophische Aphorismen* (1793), Vol. i, sect. 765, p. 439. Quoted in J. S. Mill, *op. cit.*, pp. 283–4. The most extreme and dogmatic version of this position is found in von Senden's book *Space and Sight* (London: Methuen, 1960).

but who deny that in this fact there is any ground of difference from the sighted, since their spatial concepts have exactly the same feature. This is Mill's position.³⁹ And although I do not know of any text in which Poincaré specifically addressed himself to the question of the blind, he is committed by his general views on the concepts of space to occupying this position:

To localize an object simply means to represent to oneself the movements that would be necessary to reach it. It is not a question of representing the movements themselves in space, but solely of representing to oneself the muscular sensations which accompany these movements and which do not presuppose the existence of space.⁴⁰

The other group opposed to the idea that the spatial concepts of the blind are *toto caelo* different from those of the sighted accept that the only genuine spatial concepts are those that are instanced in an array of simultaneously existing objects, and thus that are paradigmatically applicable to a simultaneously presented array, but deny that such concepts are inaccessible to the blind. According to this last group, while it is true that the blind must receive information about the spatial arrangement of the world successively, it is possible for them to organize the information they receive into a form in which genuine spatial concepts are used, or to which they may be applied. This point has been made with particular reference to the haptic perception of an object like a chair, too large to be encompassed by the hand, but presumably the point would also apply to the blind man's conception of the room or city in which he lives. Revesz puts the point about haptic perception like this:

... But even when the details have been touched, the total form is not yet given. The parts which are touched must be finally unified in a total form, in a complete impression. This synthesis presupposes a specific constructive process which we see in the visual sphere only occasionally ... Thinking and fantasy exert their effects together with intuition. The parts of a figure grasped haptically become fixated abstractly.⁴¹

³⁹ Mill, *op. cit.*, pp. 274–86.

⁴⁰ *The Value of Science* (New York: Dover, 1958), p. 47.

⁴¹ Revesz, *The Human Hand* (London: Routledge & Kegan Paul, 1958), p. 26.

It is difficult not to think of this synthesis in terms of the formation of an image, and this is the way in which Pierre Villey, a Montesquieu scholar who was himself blind, put the point:

The image which a blind man receives by touch rids itself very easily of the characteristics which constitute the modalities peculiar to tactual sensation . . . The residue which it keeps, if it does not contain colouring which is absolutely foreign to tactile nerves, and if it be less rich than the contents of the visual image, may frequently not include any element which is not in the visual image, and may coincide very nearly with that.

He acknowledges that his tactual perception of the chair is successive while visual perception is simultaneous, but he goes on:

But if, an hour after feeling it, I search in my consciousness for the memory of the vanished chair . . . I do not reconstruct it by means of fragmentary and successive images. It appears immediately and as a whole in its essential parts . . . There is no procession, even rapid, of representations . . . I couldn't tell in what order the parts were perceived by me . . . What is the residue of this work? The limit towards which they tend, and which they appear to reach, is simply form.⁴²

I do not mean to engage in this fascinating dispute now, but to get a little closer to my objective by extracting from it the distinction between two different kinds of spatial concepts which it highlights. On the one hand, we have what I shall call *serial* spatial concepts—concepts explained in terms of the succession or sequence of the subject's perceptions, and any muscular or kinaesthetic sensations accompanying these changes, whether they arise from the movement of the whole, or merely part, of the subject's body. I shall call these concepts 'spatial', but in view of the scepticism that is to follow, neither this term, nor the corresponding term 'travel-based' ought to be taken too seriously. For, as Poincaré said, the 'movements' can be characterized in terms which do not presuppose the existence of space.

Distinguished from these are what I shall call *simultaneous* spatial concepts, a notion which is much more difficult to make precise. Perhaps we may characterize them as relational

⁴² Pierre Villey, *The World of the Blind* (London: Simpkin, Marshall, Hamilton, Kent, 1922), p. 183.

concepts the situation for whose most direct application is one in which the elements related by them are simultaneously presented or perceived.⁴³ Relative to this characterization, the dispute about blind men is a dispute about whether someone who has no capacity to make the most direct application of a concept *to the world* might nevertheless possess it, and how this possession might be manifested.

However exactly the distinction between these two kinds of spatial concepts is to be drawn, that there is such a distinction seems fairly clear. It seems fairly clear, that is to say, that there are two quite different ways in which, for example, the fact that three objects *a*, *b*, and *c* lie, in that order, upon a straight line might be established. Someone might be able to tell that the line connecting the objects was straight by means of the kind of bodily movement necessary to pass from one to another, and that *b* lay between *a* and *c* by means of the temporal relation between the experience of *a*, *b*, and *c*. On the other hand, someone who was able to see, might be able simply to *see* that such an arrangement existed. Equally, it seems fairly clear that we can identify, in these different ways of detecting spatial facts, the application of different kinds of spatial concepts, which have different presuppositions and which sustain different kinds of reasoning. Someone who had information given in, or stored with the use of, concepts of one kind, or in a form to which concepts of one kind would be directly applicable, would find certain problems easier, and certain problems harder, to solve than one who relied on concepts of the other kind.⁴⁴

Armed with this distinction, let us return to the auditory universe, and ask what kind of spatial concepts its inhabitant has been provided with. So far, in fact, Hero has a theory of

⁴³ In view of the discussion of the concepts of primary properties at the beginning of the preceding part, this characterization must in no way be taken to suggest that such concepts can be *extracted from* just an experience in which distinct elements are simultaneously presented.

⁴⁴ In fact, the distinction between serial and simultaneous spatial concepts is implicit in much of the psychological literature upon spatial perception and behaviour, especially since Tolman argued for the use of the notion of a *cognitive map* in psychological explanation. For an explicit use of the distinction see F. N. Shemyakin, 'Orientation in Space', in B. G. Ananyev *et al.* (eds.), *Psychological Science in the U.S.S.R.*, Vol. I (Washington, D.C.: Office of Technical Services, 1962), pp. 186-225.

perhaps excessive simplicity, since it concerns a space of just one dimension, in which distance is only measurable upon an ordinary scale. It uses just one primitive spatial concept—'x is between y and z'—and as it was introduced, the concept is serial or travel-based. What it *means* to say that x is between y and z is simply that an experience of x will intervene between any experience of y which is followed by an experience of z and conversely.

If we were to provide Hero with analogues to our more complicated spatial concepts, such as 'arranged in a square', 'forming a circle', etc.—something that would be necessary if we were able to contemplate a generalization to a two-dimensional auditory universe—then we should have to provide Hero with some way of estimating the passage of time, so that a notion of distance permitting measurement upon a ratio scale would be understood in terms of the time of normal travel. (The presence of the word 'normal' is to signal that the estimate of distance, like the estimate of position, would be subject to revision in the light of considerations from elsewhere in the theory, which would therefore retain its holistic character.) The notion of a straight line could then be defined in terms of the shortest distance between two points, but the scheme would presumably be practically unworkable unless Hero could make provisional judgements of the straightness of the path he was following upon the basis of 'bodily sensations'. (In a parallel way, our holistic scheme of reidentifiable bodies and places would be practically unworkable unless we had the ability to make provisional judgements of the identity of bodies by recognizing them.) We have not been supposing that Hero has these conceptual riches, but for present purposes it would not matter if we had, since his concepts of space would remain serial.

Now, whether or not a subject in an auditory universe could have a use for simultaneous spatial concepts is a difficult question, partly overlapping with the question about the spatial concepts of the blind which I mentioned earlier. But Strawson did not suppose that the subject in the auditory universe could, and it is this that gives me the ground of my disagreement. Strawson was quite well aware of the

distinction between the different kinds of spatial concepts, and he in fact elaborates an objection to his discussion which is based upon the premiss that the subject in his auditory world would not have simultaneous spatial concepts. However exactly it is that Strawson does deal with this objection he raises to himself, it is not by denying the premiss.

The objector Strawson imagines begins by pointing out that, in visual perception, we are simultaneously presented with objects in a seen spatial array:

. . . these simultaneously presented elements . . . are simultaneously presented as being related in another respect: viz. in a respect which leads us to characterize one as being *above* or *below* or to the *left* or to the *right* of another . . .⁴⁵

He then goes on to object:

But relations between elements in respect of the auditory analogue of the spatial dimension cannot be presented simultaneously, all at once. They turn essentially upon change.⁴⁶

When Jonathan Bennett discusses this objection, he says that Strawson treats it much too tolerantly. This is wrong; there is a deep objection to the serial nature of the spatial concepts of the auditory universe, and Strawson does not treat it at all. The objection which Strawson does answer, or rather, shows that he does not need to answer, is one to the effect that serial spatial concepts are not *sufficiently analogous* to our simultaneous concepts. But, *even as Strawson presents it*, the objection has a much deeper thrust, for it calls into question the claim that a theory couched in serial spatial terms can genuinely embody the idea of an independently existing objective world:

But surely the idea of the simultaneous existence of the perceived and the unperceived is linked with this idea of the simultaneous presentation of elements, each of a definite character, but simultaneously exhibiting a system of relations over and above those which arise from the definite character of each. Surely the former idea is necessarily an extension of the latter, is just the idea of such a system of relations extending beyond the limits of observation.⁴⁷

⁴⁵ Strawson, *op. cit.*, p. 79.

⁴⁶ *Ibid.*, pp. 79–80.

⁴⁷ *Ibid.*, p. 80.

And to this objection, which expresses the doubt I have been approaching by such a circuitous route, Strawson offers no answer.⁴⁸

Certainly, any theory using simultaneous spatial concepts does genuinely embody the idea of an independently existing reality, the idea of the perceived and the unperceived existing simultaneously, and in exactly the same sense. If *a*, *b*, and *c* are envisaged to lie upon a straight line, when what is envisaged is an instance of the simultaneous concept—a concept whose most direct application lies in a presentation of the three elements *a*, *b*, and *c* together—then *a*, *b*, and *c* must thereby be conceived to exist in exactly the same way. If *a*, *b*, and *c* are believed to exist in such an arrangement when *b* is perceived and *a* and *c* are not, then *a* and *c* are conceived to exist, though not perceived, in exactly the sense in which *b*, now perceived, exists.⁴⁹

It is just this idea of the simultaneous existence of the perceived and the unperceived that we illegitimately import into the auditory universe by misinterpreting Hero's serial propositions as simultaneous, for example, by crediting him with something like a *map* of the world in which the information about succession and sequence is synthesized into a unitary framework. Tempting though this further step of interpretation may be, there is no warrant for it. The serial spatial propositions are once again conditional in form: if such-and-such an experience is had, followed by such-and-such another, then an experience of still a third kind will intervene between them. If this is the stuff of which the theory is made, how can it register the existence of anything going on unperceived? Unlike simultaneous spatial propositions, serial spatial propositions are not at a level different from, and therefore potentially explanatory of, propositions about order in experience. Strawson's Hero does not have the resources to rise above the level of the explicandum.

This sceptical point must be put with some delicacy. It will not do to say: 'the fact that I will have an experience of

⁴⁸ More accurately: he offers no answer which he does not answer himself.

⁴⁹ Someone who thinks of time *spatially* pictures the time series as a totality any member of which could be observed from a position outside the series. Precisely for this reason to think of time in this way is to think of the past and future events existing in the same way as do present events.

b between any times I have an experience of *a* and of *c* (and vice versa) does not guarantee that *a* and *c* exist now, when, for example, I am perceiving *b*.' By speaking in terms of the experience of *b* etc. the trick has already been given away, and the scepticism can only be expressed in the unwanted, and possibly incoherent form: 'Why may it not be that *a* springs into existence when I come to have experience of it?' Nor will it do to say that, while simultaneous spatial concepts could relate things existing simultaneously, serial spatial concepts could relate things existing at different times. Once again this turns the objection into a worry about objects 'springing into existence'. The objection is rather this: because serial spatial concepts do not provide us with a way of thinking about simultaneously existing objects, they are not obviously concepts of relations between (independently existing) objects at all.⁵⁰

Against the background of this scepticism, the immunity of the 'travel-based' theory to a simple phenomenalist reduction takes on a new complexion. Any proposition 'spatially' relating specified, perceivable 'objects' is reducible in a straightforward way to a proposition about the sequence of experiences; such irreducibility as there is comes only when Hero introduces expressions referring to 'places' whose identity conditions are tied to the whole network of propositions previously mentioned, but to no one taken individually. But it is hard to believe that an ontology appropriate to a theory of an objective world is introduced by Hero's supposed version of 'It's ϕ -ing at position *p*', if it is not already involved in the propositions of the form 'It's ϕ -ing between where it is ψ -ing and where it is χ -ing.'

The situation is really no different from this. There is a group of currencies each actively traded against the others in a situation of floating exchange rates. The basic propositions for describing this system will be of the form '£1 = \$1.75 at the end of . . . day's trading'. But we can imagine the description enriched by the introduction of the idea of

⁵⁰ If the line of reasoning expressed in these paragraphs is correct, those who deny simultaneous spatial concepts to the blind are committed to denying that they have a conception of an independently existing reality at all, which is surely very difficult to accept.

the value of the £, something which is reckoned to be increasing, decreasing or constant by means of some averaging of its relation to all other currencies. Just as in the 'spatial' case, there is sufficient *de facto* stability in the relations between most currencies from day-to-day to provide the background against which it makes sense to discriminate those changes in the £-\$ exchange rate which are due to the pound's falling, and those that are due to the dollar's rising. Now, a proposition to the effect that the value of the pound has declined is not reducible to any one proposition of the form 'At the end of day *d*, £1 = *n* units of X currency, and at the end of day *d* + 1, £1 = *n* - *k* units of X currency.' Such a proposition is not necessary, since X may be a currency moving down with the £, and it is not sufficient, since X may be moving up rather than the £ moving down.

If this does provide a parallel for the relation between the basic propositions of Hero's travel-based theory, and those which mention or quantify over 'positions', it is hard to take the theory's immunity to a simple phenomenalist reduction seriously. While a genuine theory of an independent reality will be thus irreducible, not every theory thus irreducible is a genuine theory of an independent reality.

It is a little surprising that Strawson does not treat the objection to his auditory universe which we have been considering with more understanding, for the point upon which it rests is one which Strawson himself emphasized in defence of the Kantian thesis:

. . . we must have a dimension other than the temporal in which to house at the present unheard sensory particulars if we are to give a satisfactory sense to their existing now unperceived . . .⁵¹

. . . we want an analogy of distance—of nearer to and farther away from—for only, at least, under this condition would we have anything like the idea of a dimension other than the temporal in which unperceived particulars could be thought of as simultaneously existing in some kind of systematic relation to each other and to perceived particulars.⁵²

The objector is simply taking this point, and insisting that, if space is to provide this system of relations, it must be a space

⁵¹ Strawson, *op. cit.*, p. 74.

⁵² *Ibid.*, p. 75.

constituted by simultaneous spatial relations; that if Hero is to think of unperceived particulars existing simultaneously with, and in relation to, perceived particulars, he must have simultaneous spatial concepts, and not those that 'turn essentially upon change'.

I may have given the impression that I disagree with all the most important points which Strawson makes in his second chapter. But this does not seem to me to be so. As important as any point I have so far discussed is something implicit in the entire procedure of discussion, something implicit in what, if he would not shrink from such a word, might be called Strawson's methodology. This is the idea that the connections between the fundamental concepts of our conceptual scheme are central objects of philosophical investigation, and that exploratory pressure may have to be put upon these connections by imagining situations radically unlike our own. (Hero must not take on a life of his own, so that speculations about him are misinterpreted as speculations 'about what would really happen in certain remote contingencies'; Hero and his 'world' are devices for 'testing and strengthening our own reflective understanding of our own conceptual structure'.) As a model of how to pursue this essentially imaginative exploration, Strawson's chapter is unsurpassed.

Understanding Demonstratives

I

It has recently been claimed that the use of demonstrative or indexical expressions like 'today', 'yesterday', 'here', 'I', 'you', 'this', etc., resists incorporation into a Fregean theory of meaning.¹ I have two reasons for attempting to show that this claim is not true. First, the reasoning seems to me to rest upon a common view of Frege's notions of sense and reference which is neither attractive nor required by the text, and second, because I believe that a Fregean approach to demonstrative expressions is essentially correct.

The argument which is supposed to show that demonstratives provide an insuperable problem for Frege runs like this. Consider the sentence:

(1) Today is fine

as uttered upon a particular day, *d*. Now, the concept expression '(ξ) is fine' has (on that occasion) a sense, but if the whole sentence is to have (on that occasion) a sense—express a Fregean thought—the expression 'today' must have (on that occasion) a sense as well as a referent, namely *d*.

From *Meaning and Understanding*, H. Parret and Jacques Bouveresse (eds.) (Berlin: W. de Gruyter, 1981). Reprinted by permission. The present paper rests on an idea of John McDowell's. Quite a few years ago, and more recently in 'On Sense and Reference of a Proper Name' (*Mind* 86 (1977), pp. 159–85), he argued that it was possible to ascribe Fregean sense to singular terms which I describe in this paper as 'Russellian'. The present paper is an attempt to apply this basic idea of McDowell's to demonstratives, though in the course of doing so I develop it in ways for which he must not be held responsible, particularly by tying sense to a way of *thinking of* a reference. Reading through Frege's works, I became convinced that the position McDowell argued for as a possibility is one to which he is in fact committed. (For similar views of McDowell's, see 'Truth-value Gaps' in *Logic, Methodology and Philosophy of Science VI* (Amsterdam: North-Holland, 1982).) In my interpretation of Frege I am much indebted to Dummett, whose writings I follow closely. I somehow seem to end up at a quite different place.

¹ J. Perry, 'Frege on Demonstratives', *Philosophical Review* 86 (1977), pp. 474–97.

Now, the expression-type, 'today' certainly has a meaning, which does not vary from occasion to occasion, which Kaplan² calls its character and Perry calls its role. But that cannot by itself provide a completing sense, if for no other reason than that a 'thought' which is a function of these unchanging senses could no more be assigned a truth value than can the sentence-type 'Today is fine'. But equally, the referent, *d*, cannot be regarded as providing a completing sense. So:

Neither the unchanging role of 'today' (its constant meaning) nor its changing value, provides us with a completing sense. A day is not a sense but a reference corresponding to indefinitely many different senses. So how do we get from the incomplete sense of '(ξ) is fine', the demonstrative 'today', and the context to a thought? This is the problem demonstratives pose for Frege.³

Obviously, if a Fregean approach to this utterance is to be sustained, the demonstrative in context must have a sense, and a different sense in different contexts. To this, Perry replies:

How can we extract from a demonstrative an appropriate completing sense? Such a sense, it seems, would have to be intimately related to the sense of a unique description of the value of the demonstrative in the context of utterance. But where does such a description come from? 'Today' seems only to get us to a day.⁴

Perry then goes on to show rather convincingly that no unique description can serve the purpose, for no thought about a day expressible with the use of a definite description true of that day is the same as the thought expressed with the use of a demonstrative; one can always take different epistemic attitudes towards them if one does not know that the day in question satisfies the description.

As far as I can make out, this is the main case against a Fregean approach to demonstrative expressions, and it rests, quite plainly, upon the view that a Fregean sense of any singular term must be either the sense of a definite description or 'intimately related' to such a sense. This assumption is

quite unwarranted, and when this is realized, the case collapses. In order to establish this, I need to explain what I take to be essential to Frege's notion of sense.

II

I am attracted by the following, very abstract, account of the interrelations between Fregean concepts of sense and reference—an account which owes considerably to Michael Dummett.⁵

The heart of a semantic theory for a language constructed on Fregean lines will be a *theory of reference*: a theory which assigns to each meaningful expression of the language something that can be regarded as that expression's reference or semantic value. Such a theory will proceed by discerning structure in the complex expressions of the language, and assigning references to those expressions upon the basis of assignments of references to their parts. A Fregean theory of reference will observe the principle of the compositionality: the reference of a complex expression is a function of the reference of its parts. Frege himself advocated a theory of reference according to which the references, or semantic values, of sentences and singular terms are truth values and objects, respectively, but neither of these choices is required by the adoption of the general conception. The only fixed point is this: an understanding of the language must be capable of being regarded as involving knowledge of the semantic values of expressions. In the case of sentences this knowledge can be regarded as more or less explicit, but for sub-sentential expressions, knowledge of their semantic values will simply be a logical construction out of the knowledge of the semantic values of the sentences in which they occur.

Before you object that one can understand a sentence without knowing its truth value, I hasten to remind you that the references of expressions can be thought of, or identified, in many different ways. One is thinking of the value True both when one thinks of it as the value True, and as the value

⁵ M. Dummett, *Frege: Philosophy of Language* (London: Duckworth, 1973), chaps. 5–7, 12.

² D. Kaplan, 'Demonstratives' (unpublished mimeo Los Angeles: UCLA, 1977).

³ Perry, *op. cit.*, p. 480. (I have changed the example.)

⁴ *Ibid.*, p. 485.

of the thought that snow is white, though one may not know that one is thinking of the same thing. Similarly, the function which is the semantic value of the concept expression '(ξ) is bald' can either be thought of as the function which yields truth given as inputs the objects . . . (here follows a list of the bald men) or as the function which yields truth given any object if and only if that object is bald. Frege's idea was that to understand an expression, one must not merely think of the reference that it is the reference, but that one must, in so thinking, think of the reference *in a particular way*. The way in which one must think of the reference of an expression in order to understand it is that expression's *sense*. No substantial, or positive theory of the notion of a way of thinking of something is presupposed by this conception of sense. If the intuitive notion needs to be supplemented, we can appeal to the general idea of an account of what makes it the case that a thought is about the object which it is about; two people will then be thinking of an object in the same way if and only if the account of what makes the one person's thought about that object is the same as the account of what makes the other person's thought about that object.⁶

Although a theory of meaning for a language must give the senses of expressions, we are not to think of the theory of sense as a separate tier, additional to and independent of the theory of reference. If sense is a way of thinking of reference, we should not expect to be given the sense of an expression save in the course of being given the reference of that expression. Rather than look for a theory quite independent of the theory of reference, we must take one formulation of the theory of reference—the formulation of the theory which identifies the references of expressions in the way in which one must identify them in order to understand the

⁶ I should explain the main point of departure from Dummett's account of Frege's views. Dummett is impressed, to my mind overly impressed, by the fact that one can understand a sentence without knowing its truth value. To take account of this, he regards sense, not as a way of thinking of reference, but as a way of *determining* reference—possibly by means which only a being with superior powers is capable of employing. To think of the sense of a singular term as a procedure for recognizing an object as the referent generates just the idea of sense as independent of the existence of a referent which is resisted in this paper.

language—and make it *serve as* a theory of sense. Thus, the clauses:

- (2) The reference of 'Hesperus' = Hesperus
- (3) The reference of 'Hesperus' = Phosphorus

are equivalent as clauses in the theory of reference, but only (2) can occur in a theory of reference which is to serve as a theory of sense, for it alone identifies the reference of the name in a way which *shows*, or *displays*, its sense. The use of the *Tractatus* metaphor to make this point is due to Dummett:

Indeed, even when Frege is purporting to give the sense of a word or symbol, what he actually *states* is what the reference said: and, for anyone who has not clearly grasped the relation between sense and reference, this fact makes his hold on the notion of sense precarious. The sense of an expression is the mode of presentation of the referent: in saying what the reference is, we have to choose a particular way of saying this . . . In a case in which we are concerned to convey, or stipulate, the sense of the expression, we shall choose that means of stating what the referent is which displays the sense: we might here borrow a famous pair of terms from the *Tractatus*, and say that, for Frege, we *say* what the referent of a word is, and thereby *show* what its sense is.⁷

As I have already said, Frege quite generally regarded the referent of a singular term as its semantic value. Therefore, on the present conception, the sense of a singular term is a way of thinking about a particular object: something that obviously could not exist if that object did not exist to be

⁷ Dummett, *op. cit.*, p. 227. This passage of Dummett's seems to me to contain the answer to those who argue, like Wallace (see 'Logical Form, Meaning, Translation', in F. Guenther and M. Guenther-Reutter (eds), *Meaning and Translation* (London: Duckworth), pp. 45–58), that Davidsonian theories of meaning are inadequate because they do not state the meanings of sentences. The similarity between a Davidsonian conception of the theory of meaning as a theory of reference truth, and a Fregean conception of a theory of sense as a theory of reference truth, should be particularly striking. Davidson lightens the ontological load, but the general idea is the same. I should explain that I am ignoring in this brief presentation the distinction between 'model-theoretic' and 'truth-theoretic' approaches to semantics—a Fregean theory of reference, with its ontological weight, should really be regarded as exemplifying the former approach. I have tried to explain the relation between these approaches in 'Semantic Structure and Logical Form', in G. Evans and J. H. McDowell (eds.), *Truth and Meaning* (Oxford: Clarendon Press, 1976). [Reprinted as Chapter 3 in this volume. Ed.]

thought about. If we take seriously Frege's metaphor of sense as a mode of presentation of reference, we shall not expect to be provided with specifications of sense save by means of specifications of reference, and therefore, if we remember Frege's equation of the reference of a singular term with its referent, we apparently discover at the heart of Frege's semantical system singular terms whose sense depends upon their having a referent—singular terms we more typically regard as Russellian than Fregean. But what makes a Fregean recognition of Russellian singular terms so much more sophisticated than Russell's own is that it allows such terms to have a sense as well as a reference. Russell himself did not grasp this possibility:

For the name itself is merely a means of pointing to the thing . . . so that, if one thing has two names, you make exactly the same assertion whichever of the two names you use, provided that they really are names and not truncated descriptions.⁸

The semantic difference between two such Russellian terms which have the same referent can be acknowledged, with all the benefits which Frege derived from that acknowledgement. The theory of reference will state their references like this:

(4) The reference of 'a' = a

(5) The reference of 'b' = b.

These clauses *show* the different senses which the two terms possess, but at the same time they could not truly be stated if the terms had no referent.⁹

Attractive though this possibility may be, various things that Frege explicitly says seem to rule it out as an interpretation of his views. Frege says in several places that empty singular terms may have a sense, and, what would be a consequential inconsistency with the conception outlined, he also says that sentences containing empty singular terms

⁸ B. Russell, *Lectures on Logical Atomism*, ed. D. Pears (London: Fontana, 1972), p. 103.

⁹ The significance of clauses like (4) and (5) as providing a formal recognition of the possibility of ascribing Fregean sense to Russellian singular terms is first elaborated in J. McDowell, 'On the Sense and Reference of a Proper Name', in *Mind* 86 (1977), pp. 159–85.

may have a sense (express a thought) even though they have no reference (no truth value). How can we attribute to Frege the view that sense is a mode of presentation of reference, is a way of thinking of reference, when he seems to say things explicitly inconsistent with it? However, before abandoning this interpretation of Frege's ideas, which does incorporate much that Frege says, and which takes into account his willingness to apply the distinction between sense and reference to linguistic expressions quite generally, we should examine carefully what Frege actually says about empty singular terms. For, though he does say that they, and sentences containing them, may have a sense, other things that he says in the same connection make it clear that this is far less the unequivocal rejection of the conception I have outlined than it might at first appear.

In the first place, it is clear that Frege regarded empty singular terms as *defective*, in the same way, and indeed for the same reason, as he regarded vague concept-expressions as defective. His picture of the functioning of atomic sentences required them to be composed of expressions of two kinds: one (or more) which signified an object, and one which signified a function which mapped objects (or *n*-tuples of objects) on to truth values. If any expression in an atomic sentence failed to refer to an entity of an appropriate kind, the possibility would be open that no further truth value would be determined for the sentence, and it is clear that Frege regarded this as a defect in a sentence of a quite fundamental kind—which he was quite right to do.¹⁰ With this picture in mind, Frege was simply prepared to insist that concept-expressions must be precise; there is no concession that vague concept-expressions may nevertheless have a sense of a kind appropriate for concept-expressions, or that a sentence which has no truth value on account of vagueness may nevertheless express a thought. He did not make quite the same uncompromising statements about empty singular terms, despite the fact that the motivation is precisely the same—indeed Frege frequently treats the two cases together—because of his willingness to regard empty singular terms as *fictional* (or *mythical*). Instead of simply saying.

¹⁰ See Dummett, *op. cit.*, p. 342–8.

'Proper names must have a reference', he says: 'Myth and fiction aside, proper names must have a reference', or 'For scientific purposes, proper names must have a reference'. Frege was well aware that language could be used in fiction, story-telling, and drama, and he appeared to be willing to regard the serious use of an empty singular term as of this kind; he says that the speaker has 'lapsed into the sphere of fiction', without knowing it. The following is one of the many passages in which he takes this line:

But if my intention is not realized, if I only think I see without really seeing, if on that account the designation 'That lime tree' is empty, then I have gone astray into the sphere of fiction without knowing it or wanting to.¹¹

So, Frege regarded serious utterances containing empty singular terms as belonging with the fictional use of language, and however much we may deplore this idea, it forces us to turn to Frege's account of fiction for our understanding of his views on empty singular terms. The most extended treatment is in the material for a book on Logic which Frege never finished dated around 1897.¹²

Names that fail to fulfil the usual role of a proper name, which is to name something, may be called mock proper names. Although the tale of William Tell is a legend and not history and the name 'William Tell' is a mock proper name, we cannot deny it a sense. But the sense of the sentence 'William Tell shot an apple off his son's head' is no more true than is that of the sentence 'William Tell did not shoot an apple off his son's head'. I do not say that this sense is false either, but I characterize it as fictitious.

Instead of speaking of 'fiction' we could speak of 'mock thoughts'. Thus, if the sense of an assertoric sentence is not true, it is either false or fictitious, and it will generally be the latter if it contains a mock proper name. (Footnote: We have an exception where a mock proper

¹¹ G. Frege, 'The Thought', p. 300 (trans. by A and M. Quinton), *Mind* 65 (1956), pp. 287-311. A similar use of the notion of fiction is made on almost all of the occasions on which Frege discusses empty singular terms. See, for example, 'On Sense and Reference' (pp. 62-3), *Grundgesetze* (p. 167), and the review of Schröder (p. 104), in *Translations from the Philosophical Writings of Gottlob Frege*, trans. and ed., P. T. Geach and M. Black (Oxford: Blackwell, 1970) and many places in *Posthumous Writings* eds. J. Hermes, F. Kambartel, F. Kaulbach (trans. P. Long and R. White) (Oxford: Blackwell, 1979), e.g., pp. 118, 122, 129-30, 191, 225, 232.

¹² 'Logic', in *Posthumous Writings*, pp. 126-51.

name occurs within a clause in indirect speech.) . . . Assertions in fiction are not to be taken seriously: they are only mock assertions. Even the thoughts are not to be taken seriously as in the sciences: they are only mock thoughts. If Schiller's *Don Carlos* were to be regarded as a piece of history, then to a large extent the drama would be false. But a work of fiction is not meant to be taken seriously in this way at all: it's all play.

The logician does not have to bother with mock thoughts, just as a physicist, who sets out to investigate thunder, will not pay any attention to stage-thunder. When we speak of thoughts in what follows we mean thoughts proper, thoughts that are either true or false.¹³

This passage makes it clear that Frege's claim that empty singular terms, and sentences containing them, have a sense, expressed briefly elsewhere, is much more complex and qualified than is usually realized. We might gloss it as follows. Yes: a sentence containing an empty singular term can 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. Frege's use of the notion of fiction wrongly directs our attention to just one case in which it is *as if* a singular term refers to something, namely when we are engaged in a pretence that it does, but there are others, and if we think of them we will perhaps speak of apparent, rather than mock or pretend, thoughts.

However indefensible Frege's idea of unwitting lapses into fiction may be, and however much his treatment of fiction depends upon a slide from 'mock assertions' to 'mock thoughts', his intention in this passage is clearly to deny that sentences containing empty singular terms really express thoughts, and is therefore one which makes it not at all impossible that he held the conception of the relations between sense and reference which I have outlined. (Indeed, the idea of 'mock', or anyway 'apparent', thoughts indicates a further direction in which Russell's conception of Russellian singular terms needs to be extended if it is to have any plausibility.)

¹³ *Posthumous Writings*, p. 130. I am grateful to Dagfinn Føllesdal for pointing this passage out to me. I have followed the translation of Peter Long and Roger White, save in retaining the traditional translation of 'Bedeutung' as 'reference'.

I want to stress that the idea of sense as a mode of presentation of reference is not *by itself* inconsistent with the quite unqualified ascription of sense to empty singular terms. Even if a Fregean went along with Russell and hived off definite descriptions for treatment as quantifiers, he might want to recognize a category of 'descriptive names'—names introduced by means of, and governed by, a 'reference-fixing' stipulation like: 'Let 'α' refer to whatever is φ'—whose sense is thereby guaranteed to be independent of whether or not it has a referent.¹⁴ But for names such as these the equation of reference with referent would have to be given up. One formally adequate possibility would be to take the reference, i.e. the semantic value, of such a name to be a set, determined by the rule:

$$(6) (x) (x \in \text{Reference of '}\alpha\text{' iff } \varphi(x))$$

with corresponding adjustments to the semantic values of concept-expressions. When nothing is φ, the name 'α' has no referent, but its reference is the empty set.¹⁵

I have tried to show that on a perfectly possible understanding of Frege's semantic theory, he recognized only Russellian singular terms—terms whose customary sense depends upon their having a referent. Although this seems to me to be the correct position, I am aware that many will regard it as highly controversial. It is therefore important to emphasize that the argument of this paper depends only upon a much weaker claim, namely: that there is at least nothing to prevent Frege recognizing Russellian singular terms: i.e. that there is no difficulty in ascribing to such terms a Fregean sense. Since this claim is the basis of my defence of Frege against Perry's attack, it is as well to work through the argument in detail.

¹⁴ The idea of a reference-fixing stipulation is Kripke's, see 'Naming and Necessity', in D. Davidson and G. Harman (eds), *Semantics of Natural Language* (Dordrecht: Reidel, 1972), pp. 253–355. I am presuming that φ is incapable of being satisfied by more than one thing. I have discussed the semantics of descriptive names in 'Reference and Contingency', *Monist* 62 (1979) [reprinted as chapter 7 in this volume. Ed.].

¹⁵ Here and elsewhere I rely upon the verbal distinction between 'referent' and 'reference' introduced by Dummett, op. cit., pp. 409 ff. Were we accustomed to Long and White's translation of 'Bedeutung' as 'meaning', the position adopted in the text would have stood out as a possibility more clearly.

The essential use to which Frege puts the ascription of sense to singular terms is to explain the differing cognitive values of the sentences $A(t)$ and $A(t')$ when t and t' refer to the same thing. (The difference between $t = t$ and $t = t'$ is just a special case of this phenomenon.) Now, to say that two sentences differ in cognitive value is to say that it is possible for anyone who understands them correctly to coherently take different epistemic attitudes towards them—to accept one sentence as true and to reject, or to be unsure about the other sentence.

Suppose now that to ascribe a Fregean sense to a singular term is to say that there is a particular way in which its referent must be thought of (as the referent) if the term is to be understood. If two co-referring Russellian singular terms have different senses, different ways of thinking of their common referent are required in order to understand them. We have linked the idea of a way of thinking of something to an account that may be offered of what makes a subject's thought about its object, and certainly no argument can be based upon this idea alone to the conclusion that senses can be grasped in the absence of a referent.

Now if the assignment of senses to singular terms t and t' is to explain the differing cognitive values of the sentences $A(t)$ and $A(t')$, it must be the case that if the singular terms t and t' have different senses the sentences will have different cognitive values—i.e. it must be possible for anyone who understands the sentences to take different epistemic attitudes towards them. And this will be so, provided the following very plausible principle is true:

- (P) If the account of what makes a subject's thought T_1 (about x to the effect that it is F) about x is different from the account of what makes his thought T_2 (about x to the effect that it is F) about x , it is possible for the subject coherently to take, at one and the same time, different epistemic attitudes towards the thoughts he entertains in T_1 and in T_2 .

At no point is it necessary for Frege to adopt any substantial theory of what form these accounts must take. In particular it is not necessary for him to suppose that ways of thinking

of objects can always be given by giving some definite description uniquely true of the object, or to make any other supposition which would lead to 'existence-independent' senses. It is not necessary, because it is not plausible to suggest that the only kind of account of what makes a subject's thought about an object which is capable of making (P) true is one which relies upon the subject's possessing a unique description of the object.

The initial 'if' in (P) can be strengthened to an 'if and only if' without loss of plausibility and if this strengthening is acceptable Frege is entitled to his equation of the sense of a singular sentence and a thought, when this is understood to be the object of propositional attitudes. An equation can be made between the senses of a singular sentence and a thought only if it is not possible for someone who has understood two singular sentences which agree in sense to take different attitudes to them, but on Frege's view this will not be possible. A difference in attitude would require a difference in the ways the subject thought of the object referred to (by the strengthening of (P)), and this would conflict with the hypothesis of an identity of sense, given that sense is a way of thinking of the referent.

Thus, we see that far from the Fregean sense of a singular term being restricted to the sense of some definite description (and therefore being 'existence-independent'), it is perfectly possible for there to be 'existence-dependent' Fregean senses—the Fregean senses of Russellian singular terms. I have in fact suggested that we should re-examine those passages in which Frege showed himself willing to ascribe sense to empty singular terms, but in any case, I know of no passage in which Frege can be construed as insisting that singular terms *must* have an existence-independent sense. In view of this we can appreciate how wrong-headed it is to consider a Fregean sense as necessarily *intermediary* between thinker and referent, as something which must, from a certain point of view, *get in the way*, or anyway render indirect what might be direct. A way of thinking of an object is no more obliged to get in the way of thinking of an object, or to render thinking of an object indirect, than is a way of dancing liable to get in the way

of dancing, or to render dancing somehow indirect.¹⁶ And, finally, we can appreciate how baseless it is to maintain that an extension of a Fregean theory to demonstrative singular terms *must* involve assigning to them the sense of, or anything like the sense of, some definite description. So Perry's argument against Frege collapses.

III

Let us return to the problem of demonstratives. We have seen that Perry's demonstration that there can be no 'completing sense' for 'today' is unsound, but something must be said about what such a completing sense might be.

'Today', as uttered on *d*, has a completing sense, if and only if there is some particular way in which one must think of the referent, *d*, in order to understand the utterance containing it. And of course there is. Even if *d* is the first day after my last lecture, I shall not have understood the utterance of (1) if I think of *d* only as the first day after my last lecture, thereby coming to believe that the utterance is true if and only if the first day after my last lecture is fine, perhaps not realizing that today is the first day after my last lecture. In order to understand (1) I must think of *d* as the current day, thereby coming to have the thought which I might express in the words: 'What the speaker said is true if and only if it is fine today.'¹⁷ Now, what makes a man's thought about a day when he thinks of it as the current day—as 'today'—is not something which it is incumbent upon Frege to explain. It is indeed a difficult question. I myself would say something like this.

To give an account of how a thought concerns an object is to explain how the subject knows which object is in question. In the case of 'today', the subject, of course, knows which day is in question, but this knowledge at least partly consists in a disposition to judge the thoughts (which depend upon

¹⁶ I have in mind here several remarks of Kaplan's, who advocates 'the semantics of direct reference . . . theories of meaning according to which certain singular terms refer directly without the mediation of a Fregean *Sinn* as meaning' (Kaplan, 'Demonstratives', p. 1). See also Kaplan, 'How to Russell a Frege Church', *Journal of Philosophy* 72 (1975), pp. 716–29.

¹⁷ It is therefore not true that 'today' only gets us to a day.

this knowledge) as true or false according to how things observably are upon that day which in no way rests upon his capacity to identify that day as meeting some antecedently given condition, but depends only upon his being alive on that day. There should be no mystery here; we can test very easily whether or not someone, in his interpretation of a sentence, is thinking of the day in the right way by seeing if he is disposed to judge the sentence as true or false according to how things observably are on that day. Similarly, I should want to place in a central position in any account of what makes a man's thought concern a particular place in the way which is required for understanding sentences containing the term 'here', a knowledge of which place is in question which at least partly consists in a disposition to judge that thought as true or false according to how things observably are at that place—a disposition which he can have *vis-à-vis* just one place in the universe in virtue of his occupying it, and which in no way depends upon his capacity to recognize that place as the unique satisfier of some description. If these accounts are on anything like the right lines, it is very easy to understand how these 'ways of thinking' are irreducible to any other, since no other way of knowing which object is in question, certainly no 'descriptive' way, can guarantee the existence of the relevant dispositions.

However, these are speculations which need to be embedded in a general theory of thought if they are to carry conviction. All that a Fregean needs from his opponent is an acknowledgement that those thoughts about a day which we typically express with the use of 'today' do involve a particular way of thinking about a day; if this is granted, he can explain how 'today', in a context, has a 'completing sense'.

Therefore, in order to understand the utterance of (1) made on *d*, one must have, on *d*, the thought which one might express in words by:

(7) What the speaker said is true if it is fine *today*.

It seems reasonable to say that such a statement is capable of showing the sense which the sentence has on that occasion. However, there might appear to be a difficulty here. A theory of reference was conceived to be a finite set of principles

from which the references of complex expressions, particularly sentences, could be derived. When we think of the principles from which (7) might be derived, we naturally think of the general statement:

(8) For all days *d*, 'today' as uttered on *d* refers to *d*.

But now, this universally quantified principle cannot be thought to show the sense of any particular use of the expression, nor does it appear to issue exclusively in theorems of the form (7), which can. After all,

(9) What the speaker said in uttering 'Today is fine' on my birthday is true if it is fine on my birthday

is equally a consequence of (8), but it apparently does not show the sense which that sentence had on that occasion.

Our interest in theories of reference and of sense is ultimately to better understand the capacity of speakers to speak and understand their language, and when we remember this interest, the present difficulty will be seen as spurious. Speakers do not literally deduce the truth conditions which sentences have from certain universally quantified principles, whose precise form we must endeavour to establish. Speakers judge the truth conditions of particular sentences, and in so doing they exercise complex and interconnected dispositions in which their understanding of the individual atoms of the language may be taken to consist. We are therefore not required to attribute to speakers the general belief that any token of 'today' refers to the day on which it is uttered—and then wonder what form that belief takes, or how they derive the right kind of judgement of truth conditions from it. We are rather ascribing to speakers a *propensity to form particular beliefs*, of particular tokens of 'today', that they refer to the day of utterance, identified in a particular way, the exercise of which yields thoughts of the form of (7). The inclusion in a theory of reference of a general principle like (8) is a gesture in the direction of identifying the relevant propensity, and it certainly requires supplementation. But this point, though it needs to be borne in mind in interpreting systematic theories of meaning.

is irrelevant to a Fregean theory of the sense and reference of 'today'. What matters for that theory is that tokens of 'today' should have a sense as well as a reference, not that the sense of all tokens of the expression should be capable of being shown in a single principle of the theory of reference.

Thus we have found no reason to depart from Frege's view:

[The thought, for example, that this tree is covered with green leaves will surely be false in six months' time. No, for it is not the same thought at all. The words, 'This tree is covered with green leaves' are not sufficient by themselves for the utterance, the time of utterance is involved as well. Without the time indication this gives we have no complete thought, i.e. no thought at all] . . . the same words, on account of the variability of language with time, take on another sense, express another thought.¹⁸

IV

In understanding the sentence 'Today is fine', said on d_1 , one can be regarded as having a Fregean thought, but is it a thought which one can have on any other day? Frege appears to have thought that it is:

If someone wants to say the same today as he expressed yesterday using the word 'today', he must replace this word with 'yesterday'.¹⁹

Frege appears to have held that to have on d_2 just the thought which one has when one thinks 'Today is fine' on d_1 , one must think 'Yesterday was fine'. Presumably this means that it is possible for someone reading yesterday's newspaper to understand sentences like:

(10) The Prime Minister is holding a cabinet meeting today

by realizing that it is true if the Prime Minister held a cabinet meeting the day before. Now, many philosophers, commenting on this passage, have concluded that Frege intended to abandon a notion of 'what is said', or 'the thought expressed' which was 'psychologically real' in the sense of being the object of propositional attitudes, and was giving expression

¹⁸ Frege, 'The Thought', pp. 309–10.

¹⁹ *Ibid.*, p. 296.

to the idea that two people would express the same thought provided that they refer to the same object (in whatever way) and say the same thing about it.²⁰ Such a conception of *what is said*, or *the thought expressed* is so wholly antagonistic to the theory of language ushered in by the distinction between sense and reference, and is otherwise so wholly absent from his work, that it seems to me to be doubtful that the passage has been correctly interpreted. It is clear, for example, that Frege would have been willing to continue the passage:

. . . he must replace this word with 'yesterday', or 'my birthday', or any other expression designating the same day?

Might Frege not have had in mind an idea of a thought the grasp of which, on a later day, requires just as specific a way of thinking of a day as does its grasp on an earlier day namely as the preceding day? Pursuing this suggestion, we discover that, far from abandoning the 'psychologically real' notion of a thought in favour of a psychologically quite uninteresting equivalence class of thoughts, Frege may well have glimpsed what results when the notion is extended to the sphere of human thinking which depends upon the position human beings have in space and time.

We must agree that, if a subject thinks on d_1 , about d_1 , to the effect that it is fine by thinking 'Today is fine', and thinks on d_2 , about d_1 , to the effect that it is fine, by thinking 'Yesterday was fine', there is some level of description at which he is thinking of the same day in different ways—the account of what makes his thoughts about d_1 in the two cases will not be entirely the same. And it is natural to think that this difference in ways of thinking can be exploited to produce the possibility of differing epistemic attitudes to the thoughts, which would then preclude their being the same thought, if thoughts are intended to be the object of propositional attitudes.

However, the natural suggestion is not correct, there is no

²⁰ See, for example, Kaplan, 'Demonstratives', p. 43. Dummett comes close to this in Frege, p. 384, although there, other expressions which can be used to express the same thought are restricted to other demonstrative expressions with the same referent.

headlong collision between Frege's suggestion that grasping the same thought on different days may require different things of us, and the fundamental criterion of difference of thoughts which rests upon the principle that it is not possible coherently to take different attitudes towards the same thought. For that principle, properly stated, precludes the possibility of coherently taking different attitudes towards the same thought *at the same time*. Consider S, who accepted the sentence 'Today is fine' when uttered on d_1 , and who rejects the sentence 'Yesterday was fine' when uttered on d_2 , perhaps because he has misremembered the weather, or because he has 'lost track of time'. Now, in order to apply the criterion of difference in this situation, we must first make a decision as to what it would be for S to have exactly the same thought on d_2 as he had when he thought on d_1 'Today is fine'. Because its application requires a prior decision on this question, the criterion for difference cannot by any means be the whole story of the identity and distinctness of thoughts, and it is powerless to upset Frege's suggestion. For, either we hold that it is possible to think again the thought entertained on d_1 or we do not. If we hold that it *is* possible, no better account than Frege's can be given of the circumstances under which it is possible. (If this is not obvious, some merits of his account will be given below.) Hence, on this alternative, to think 'Yesterday was fine' *is* to think the same thought again, and so no possibility opens up, on d_2 , of coherently assenting to the same thought as one accepted when one judged on d_1 'Today is fine', and of dissenting from the thought 'Yesterday was fine'. To hold, on the other hand, that it is not possible to have on d_2 the very same thought as one had on d_1 , while not at all a ridiculous proposal, obviously precludes use of the criterion of difference against Frege's contrary view. Some other consideration must be appealed to.

Frege's idea is that being in the same epistemic state may require different things of us at different times; the changing circumstances force us to change in order to keep hold of a constant reference and a constant thought—we must run to keep still. From this point of view, the acceptance on d_2 of 'Yesterday was fine', given an acceptance on d_1 of 'Today

is fine' can manifest the *persistence* of a belief in just the way in which acceptance of different utterances of the same sentence 'The sun sets in the West' can. Are there any considerations which can be advanced in favour of this way of looking at matters?

To answer this question, we must contrast Frege's conception with the opposing conception, according to which the thoughts associated with sentences containing temporal indexicals cannot be grasped at later times. On this atomistic conception, what Frege regards as a persistence of a belief is really a succession of different but related beliefs concerning the same time. It must of course be acknowledged that these patterns, or sequences, of beliefs are very commonly to be met with—that human beings do have a general propensity, on forming one belief in this series, later to have the other beliefs in the series, but this fact by itself does not settle the issue. Admittedly it is not clear what account can be given of this succession of belief on the atomistic conception. One belief cannot give rise to another by any *inference*, since the identity belief that would be required to underwrite the inference is not a thinkable one; no sooner does one arrive in a position to grasp the one side of the identity than one has lost the capacity to grasp the other. But one can be suspicious of the atomistic conception for other, deeper, reasons.

On the atomistic conception, whether there are later elements in the series, and whether or not they concern the same object is quite irrelevant to the subject's capacity to entertain one of the atoms. The atom must be a perfectly coherent unit of thought by itself, even if it is entertained by one who has not the least propensity to form the other members of the series. But this, Frege might well have thought, is wrong. No one can be ascribed at t a belief with the content 'It is now ψ ', for example, who does not have the propensity as time goes on to form beliefs with the content 'It was ψ just a moment ago', 'It was ψ earlier this morning', 'It was ψ this morning', 'It was ψ yesterday morning', etc., though of course this propensity can be counteracted by new evidence. Frege might be credited with the insight that a capacity to keep track of the passage of time is not an optional addition to, but a precondition of, temporal thought.

If this is so, the thought-units of the atomist are not coherent, independent, thoughts at all, but so to speak, cross-sections of a persisting belief state which exploits our ability to keep track of a moment as it recedes in time.

The metaphor of 'keeping track of something' originates in connection with another kind of thought about an object, and it provides a useful, if only partial, parallel. Suppose that one is watching a scene in which there are several similar objects moving about fairly rapidly, but no so rapidly as to prevent one's keeping track of one in particular. In such a situation, one can think about one of these objects rather than any other, but any such thought rests upon a skill we possess of keeping track of an object in a visual array over time. Our eyes and our heads move, perhaps we are also obliged to turn or move our bodies, but these changes are required to maintain contact with the same object over time. So, one's thought *at* a time is dependent upon an ability which is necessarily manifested only *over* time. One might begin the period with the belief of an object that it is valuable, and end it with a belief of the same object that it is valuable. Now, a move parallel to the one which Frege made in connection with 'today' and 'yesterday', would be to hold that one belief has persisted over time, despite the local differences which the changing circumstances have imposed upon one. And there is a parallel, opposing, atomistic move which would regard the subject as holding a *sequence* of different beliefs over the relevant period of time, altering as the subject's relation to the object altered. And the objection to the atomistic position here is the same as in the earlier case. If the atomistic position were correct, it ought to be possible to have just one of the members of the sequence no matter which others accompanied it, i.e. in the absence of any capacity to keep track of the object. But if that ability is missing, it is not possible for a subject to have a thought about an object in this kind of situation at all. Now Frege himself did not give this parallel, but he did write, after the passage just quoted: 'The case is the same with "here" and "there".' Indeed it is; our ability to think of a place as 'here' is dependent upon our general ability to keep track of places as we move about (which requires, in general,

the ability to know when we are moving), so, once again, there could not be thoughts interpretable as 'It's ψ here', if they were not entertained by a subject who had the propensity to entertain, as he moves about, thoughts expressible in the words 'It's ψ there'.

These examples suggest that we have to regard the static notion of 'having hold of an object at t ' as essentially an abstraction from the dynamic notion of 'keeping track of an object from t to t' '. And the grasp, at t , of a thought of the kind suggested by the passage from Frege, a *dynamic* Fregean thought, requires a subject to possess at t an ability to keep track of a particular object over time. It is not precluded that one should have only a momentary grasp of a dynamic Fregean thought, for it is not precluded that, after an object has engaged with one's capacity to keep track of objects of that kind, one should lose track of it, and with it, the thought. Indeed, it is an aspect of the capacity that the subject will, in general, know when this has happened. The capacities upon which certain kinds of thought rest can only be described in dynamic terms; it does not follow that any exercise of those capacities must be extended over time.

Consequently, the *way of thinking of an object* to which the general Fregean conception of sense directs us is, in the case of a dynamic Fregean thought, a *way of keeping track of an object*. This permits us to say after all that a subject on d_2 is thinking of d_1 in the same way as on d_1 , despite lower level differences, because the thought-episodes on the two days both depend upon the same exercise of a capacity to keep track of a time.²¹

²¹ Kaplan briefly raises the possibility sketched in this section, under the heading 'cognitive dynamics', but dismisses it: 'Suppose that yesterday you said, and believed it, "It is a nice day today". What does it mean to say, today, that you have retained *that* belief? . . . Is there some obvious standard adjustment to make to the character, for example, replacing *today* with *yesterday*? If so, then a person like Rip van Winkle, who loses track of time, can't retain any such beliefs. This seems strange.' I see no more strangeness in the idea that a man who loses track of time cannot retain beliefs than in the idea that a man who loses track of an object cannot retain the beliefs about it with which he began. If one has in fact lost track of time without knowing it, then one could think that one had retained one's beliefs when one has not. But, since in general thoughts associated with Russellian singular terms are such that the subject cannot infallibly know that he has one, we should not jib at denying the subject infallible knowledge of when he has the same one.

V

In discussing thoughts expressed with the use of the pronoun 'I' Frege wrote:

Now, everyone is presented to himself in a particular and primitive way in which he is presented to no one else.²²

Replacing Frege's metaphor of 'being presented with an object' with the notion of 'thinking of' which underlies it, Frege appears to be saying that each person thinks about himself in a way which is primitive and available to no one else. Since this way of thinking about oneself would be neither primitive nor available to anyone else if it exploited one's knowledge that one uniquely satisfied some description, the passage appears to provide the clearest possible evidence that Frege did not hold that all ways of thinking of objects must involve thinking of those objects as uniquely satisfying some description. It is not unreasonable to suppose, on the strength of this passage, that Frege had noticed the irreducibility of 'I'-thoughts to any other kind of thought fifty years before Castañeda made it part of the philosopher's stock in trade.²³

However, Perry holds that Frege's conception of a thought was such that to have an 'I'-thought could not possibly be to have a Fregean thought. One of his reasons is that Frege held that thoughts are 'generally accessible'.

We can see that having such beliefs *could* not consist *wholly* in believing Fregean thoughts. Consider Frege's timeless realm of generally accessible thoughts.²⁴

Since it is an immediate consequence of what Frege said about 'I'-thoughts that they are *not* 'generally accessible', Perry appears to be arguing that a Fregean approach to 'I'-thoughts must be inadequate by citing a supposed requirement upon Fregean thoughts—that they be generally accessible—which Frege appears to have shown himself free of precisely in what he says about 'I'-thoughts. Presumably

²² Frege, 'The Thought', p. 298.

²³ See H.-N. Castañeda, "'He": A Study in the Logic of Self-Consciousness', *Ratio* 8 (1966), pp. 130-57, and in many other papers.

²⁴ Perry, *op. cit.*, p. 492.

Perry would justify his line of criticism by arguing that the shareability of thoughts was such a central Fregean doctrine that nothing recognizably Fregean could exist in its absence:

Nothing could be more out of the spirit of Frege's account of sense and thought than an incommunicable, private thought.²⁵

I do not believe that this is true. It is true that Frege stresses that it is *possible* for thoughts to be grasped by more than one person; were this not so, neither communication nor disagreement would be possible. But this point requires only that thoughts are not by their very nature precluded from being grasped by more than one person, not that every single thought must be capable of being grasped by more than one person. What is absolutely fundamental to Frege's philosophy of language is that thoughts should be *objective*—that the existence of a thought should be independent of its being grasped by anyone, and hence that thoughts are to be distinguished from *ideas* or the contents of a particular consciousness. When Frege stresses that thoughts can be grasped by several people, it is usually to emphasize that it is not like an idea:

A true thought was true before it was grasped by anyone. A thought does not have to be owned by anyone. the same thought can be grasped by several people.²⁶

His most extended treatment of the nature of thoughts—'The Thought'—makes it clear that it is the inference from shareability to objectivity which is of paramount importance to Frege, rather than shareability itself. Since an unshareable thought can be perfectly objective—can exist and have a truth value independently of anyone's entertaining it—there is no clash between what Frege says about 'I'-thoughts and this, undeniably central, aspect of his philosophy. Although Frege does tend to speak without qualification of thoughts as graspable by more than one person, I do not myself see why this should be regarded as an indispensable tenet, rather than a slight overstatement, of his position. Perry certainly does not tell us why he attaches such importance to it.

²⁵ *Ibid.*, p. 474.

²⁶ *Posthumous Writings*, p. 251.

Perry makes clear that, quite independently of any question of Fregean scholarship, he thinks that fatal objections can be raised to the idea of an unshareable thought. This might seem a bit steep, coming after a criticism of Frege based upon his supposed insistence that all thoughts be 'generally accessible'. However Perry will explain that the peculiarities of 'I'-thoughts can be accommodated on his system without recognizing unshareable thoughts. On Perry's system, both you and I can grasp the thought you express when you say 'I am hot', only we grasp it *in different ways*. However, I try to show in the next section that this is just a notational variant of Frege's theory.

It is true that Perry has other reasons for saying that an 'I'-thought cannot be a Fregean thought, for he thinks that objection can be raised to a Fregean treatment of 'I'-thoughts even if Frege is allowed the 'private senses' to which he so desperately resorted. The argument here contains no surprises, for the difficulties result from the excessively wooden interpretation which Perry places on Frege's notion of 'way of being presented to oneself' as a result of trying to force it into the mould of descriptive identification.²⁷ He does point out that thoughts about times expressible with 'now' are also irreducible to description, and therefore if Frege's strategy is general:

Frege will have to have, for each time, a primitive and particular way in which it is presented to us at that time, which gives rise to thoughts accessible only at that time, and expressible at it, with 'now'. This strikes me as very implausible.²⁸

In fact we have seen that Frege does not appear to hold that such thoughts are only graspable at one time (Part IV), but even if we ignore this, I do not see that the approach can be so easily dismissed. What is so absurd about the idea that there are thoughts which one can have only because one occupies a particular position in space, or time, or because one is currently perceiving an object? This is just to say that there are ways of thinking about objects which require one to stand in a specific spatial, temporal, or causal relation to the object, and rather than deserving dismissal as implausible.

²⁷ Perry, *op. cit.*, p. 491.

²⁸ *Ibid.*

the point seems to me to be worthy of the greatest respect. And to say that the ways of thinking are primitive is to say that they are not reducible to any other, particularly not to any which exploit knowledge of a description of the object, and this too is a point which Perry should applaud, figuring as it does so extensively in his own work.

Perhaps the implausibility is supposed to lie in the consequence of there being an infinite number of distinct, primitive, and particular ways of thinking of objects—one for each time—but alarm at this idea can only rest upon a confusion. A way of thinking about an object is given by an account of what makes some thinking about that object. In the case of a particular 'I'-thought, for example, I envisage statements of the form

(11) S is thinking of S' at t because $R_1(S, S', t)$

where R_1 is an as yet unspecified relation which can only be satisfied by a triple of S , S' , and t if $S = S'$. In terms of this idea, we can make perfectly good sense of the claim that different people think of different things (i.e. themselves) *in the same way*: we do not hold that precisely the same account can be given of what makes each of their thoughts have the object that it does, but that the same *type* of account can be given—namely in terms of the relation R_1 . While it would no doubt be implausible to suppose that there are an infinite number of different *types* of account, I see no difficulty whatever in the idea of as many different *particular* accounts as there are times and persons.

VI

A Fregean thought of the kind associated with a sentence 'Today is F ' said on d can be equated with the ordered pair of the sense which 'Today' has on d and the sense of the concept-expression '(ξ) is F ', thus:

(12) (Sense on d of 'today', Sense of '(ξ) is F ')

One grasps the sense of 'today' on d if and only if one thinks of d as the current day – i.e. in virtue of one's satisfying some relational property $\lambda x(R_2(x, d))$ – so we may

equivalently equate the Fregean thought with the ordered pair:

(13) $\langle \lambda x(R_2(x, d)) \rangle$, Sense of '(ξ) is F').

One entertains the object (13) if one thinks of a day in virtue of one's satisfying the first component, to the effect that it is F. If we wished to bring out the way in which any two utterances of (1) are similar, we might, equivalently, equate the thought with the triple:

(14) $\langle d, \lambda x \lambda y(R_2(x, y)) \rangle$, Sense of '(ξ) is F');

one entertains (14) if one thinks of the first member, in virtue of oneself and the first member, satisfying the second member, to the effect that it is F. In this construction the second and third components of the sense of an utterance of (1) are always the same, though the sense which 'today' has on d can be equated with neither the first member, nor the second member, taken singly, but only with the pair. Then, when Hume thinks 'I am hot' he entertains:

(15) $\langle \text{Hume}, \lambda x \lambda y(R_1(x, y)) \rangle$, Sense of '(ξ) is hot');

and when I think that I am hot, I entertain:

(16) $\langle \text{G.E.}, \lambda x \lambda y(R_1(x, y)) \rangle$, Sense of '(ξ) is hot').

We know that Perry thinks that it is necessary to abandon the notion of a Fregean thought when dealing with sentences containing demonstratives, but with what would he replace it? He introduces a notion of thought according to which the sentence $F(t)$ uttered in context c , and the sentence $F(t')$ uttered in c' express the same thought provided the referent of t in c is the same as the referent of t' in c' . Let us call a thought of this kind a 'P-thought'; a P-thought can be identified with an equivalence class of Fregean thoughts, or alternatively with an ordered pair of an object and a sense of a concept expression ('... a thought consisting of a certain object and an incomplete sense ...').²⁹ It is the introduction of thoughts of this kind that Perry claims to be his greatest departure from Frege: 'The idea of individuating thoughts by objects or sequences of objects would be particularly out

²⁹ Cf. *ibid.*, p. 496.

of place in (Frege's) system'.³⁰ However, for Perry, when we entertain a thought, we do not just stand in a certain relation to a P-thought; we entertain a P-thought *in a certain way*. When Hume thinks 'I am F' and then thinks 'That man is F' (indicating himself in a mirror), he entertains the same P-thought, but in a different way. Perry has a positive proposal about ways of entertaining thoughts, one which links them to the 'roles' of meanings of demonstrative expressions of natural language. I shall come to this aspect of his position in a moment, but what concerns me now is to see whether, the positive characterization apart, Perry is putting forward anything other than a notational variant of Frege's position.

Perry frequently speaks as though P-thoughts are the objects of propositional attitudes; this is, indeed, what their name would suggest. This means, I take it, that a belief-ascription asserts a relation between a subject and a P-thought. If this is Perry's intention, then his position is not a notational variant of Frege's, since although Frege can say everything Perry can say (using the equivalence class to Fregean thoughts) the converse is not the case. (This is generally true when the ontology of a theory T_1 is less 'fine-grained' than the ontology of theory T_2 .) Frege, for example, can consistently describe the belief system of a subject S who understands and accepts, in context c , the sentence $F(t)$, and who understands, and neither accepts nor rejects, in context c' , the sentence $F(t')$ when t in c refers to what t' in c' refers to. S believes one Fregean thought, and neither believes nor disbelieves another Fregean thought. If belief is simply a relation to a P-thought, this situation cannot be described.³¹

If Perry is to be able to report S 's epistemic situation, his belief reports must be more complex than just a simple relational statement between a subject and a P-thought; the way in which the thought is apprehended must come in as

³⁰ *Ibid.*, p. 496.

³¹ Perry makes life too easy for himself by considering only the case where S rejects $F(t')$ in c' ; this he can, and does, describe as S 's believing the different P-thought, concerning the object, that it is not F (*op. cit.*, pp. 495-6). The importance of the case of agnosticism was noticed by Kaplan in 'Quantifying In', in D. Davidson and J. Hintikka (eds), *Words and Objections* (Dordrecht: Reidel, 1969), pp. 206-42.

well. Perry uses locutions like 'By entertaining the sense of "I", S apprehended the thought consisting of Hume and the sense of "(ξ) is hot"', and so perhaps he has in mind some such construction as:

(17) S apprehends-in-way-*w* $\langle x$, Sense of '(ξ) is F'.

But surely this is now a notational variant of Frege's approach, at best. Where Frege would write:

(18) S believes $\langle x$, *w*, Sense of '(ξ) is F'.

Perry will write:

(19) S believes-in-way-*w* $\langle x$, Sense of '(ξ) is F'.

or 'S believes, by apprehending such-and-such a sense, the thought consisting of *x* and the sense of "(ξ) is hot"'.³²

So, finally, we can come to examine Perry's positive proposals about the various ways of thinking about objects which are involved in the understanding of sentences containing demonstratives, clear in our minds that they are not opposed, but supplementary to any views of Frege's. If Perry has succeeded in making these ways of thinking clear, if he has explained the various R-relations upon which they depend, all other Fregeans have reason to be grateful. For as I have said, Frege left these ways of thinking of objects quite uncharacterized, and so the nature of the senses of these expressions is unknown, even if their existence is not. Furthermore, it has seemed that such an account must presuppose some of the profoundest philosophy. In the case of 'I' for example, one might think that an account of the relation R_1 which explicates 'self-identification' must incorporate the insights, as well as illuminate the struggles, of Descartes, Kant, and Wittgenstein, and many others. One might have expected an account of self-identification—of the way in which we know, when we think of ourselves, which object is in question—would have to relate it to our special ways of gaining knowledge of ourselves, both mental and physical, both past and present. (At the very least, Hume's realizing that he is Hume must involve appreciating the bearing of the knowledge he can gain in these special ways to the truth value of very many thoughts that Hume

is F.) In this way, the 'immunity to error through mis-identification' of these ways of gaining knowledge would be explained.³² And one might have thought that an explanation of one's capacity to grasp indefinitely many thoughts about oneself which one does not know to be true—thoughts about one's remote past, or one's future—could be provided only when the role of conceptions of personal identity in self-identification had been made clear.³³

Similarly, one might have thought that an account of demonstrative identification, which underlies the thoughts we might express in (certain uses of) sentences like 'This table is round' would have to show how thought can depend upon perception, at least in such a way that we would know what kind of perception can sustain demonstrative identification. Can one demonstratively identify a man when one sees him in a mirror, on a television, in a photograph, in an X-ray? Can one demonstratively identify a man when one hears his footsteps, when one hears him on the telephone, on the radio, on a record? Can one demonstratively identify a city when one perceives only the inside of a room located within it?³⁴

Perry's answers to these profound questions have an appealing simplicity. In the case of a 'self-conscious' thought, for example, he writes:

We accept that there is no thought only Hume can apprehend. Yet only he can know that he is Hume. It must not just be the thought that he thinks, but the way that he thinks it, that sets him apart from the rest of us. Only Hume can think a true thought, by saying to himself

I am Hume

Self-locating knowledge, then, requires not just the grasping of certain thoughts, but the grasping of them via the senses of certain sentences containing demonstratives.³⁵

³² S. Shoemaker, 'Self-Knowledge and Self-Identification', *Journal of Philosophy* 65 (1968), pp. 555–68.

³³ For two papers that help to bring out the profundity of the question, see L. Wittgenstein, 'Notes for a Lecture on Private Experience', *Philosophical Review* 77 (1968), pp. 275–320 and G. E. M. Anscombe, 'The First Person', in *Mind and Language*, S. Guttenplan (ed.) (Oxford: Clarendon Press, 1975), pp. 45–65.

³⁴ For a discussion of this latter question, see G. E. Moore, 'Some Judgments of Perception', *Philosophical Papers* (London: Routledge & Kegan Paul, 1922), pp. 220–53.

³⁵ Perry, *op. cit.*, p. 492.

By 'entertaining' the meaning or role of the demonstrative 'I', Hume thinks of himself. Similarly, by entertaining the meaning or role of the demonstrative 'here', Hume thinks of a particular place.

Simple though it is, I find that the proposal evades me. The role of a demonstrative was explained as that aspect of the meaning of the expression which was constant from occasion to occasion—presumably a (constant) function from contexts of utterance to objects. So to 'entertain the role' of the demonstrative 'I', for example, would presumably be to have this function in mind in some way. The function is determined by the rule that in any context of utterance the value of the function is the speaker in that context, so I suppose that in a derivative sense one could be said to have the function in mind, and so to be entertaining the role of the demonstrative, if one had in mind the description 'the person speaking', or 'the person who utters the token of "I"'. Since Perry insists that the meaning of 'I' is not a complete Fregean sense, and that different people entertain precisely the same meaning, perhaps we should think of this as like a description containing a free variable.

(20) The person who utters x and x is a token of 'I'.

But, leaving aside the question of how reference is achieved to a particular token—which is an aspect of the same general problem with which we are concerned—what has the idea in (20) got to do with one's capacity to think of oneself self-consciously? The problem is this. No one can give an account of the constant meaning (= role) of a demonstrative without mentioning some relational *property* (relating an object to a context of utterance) which an object must satisfy if it is to be the referent of the demonstrative in that context of utterance, but the idea of this property plays no part in an explanation of what makes a subject's *thought* about himself, or the place he occupies, or the current time.

It seems clear that we are on the wrong track, these suggestions must be as far from capturing what Perry intended by 'entertaining the role of a demonstrative' as they are from answering the questions with which we began. An alternative interpretation is suggested by Perry's remark: 'Only

Hume can think a true thought by saying to himself, 'I am Hume'."²⁶ Perhaps 'entertaining the sense of "I am Hume"' is to be understood along the lines of mentally uttering 'I am Hume', or 'saying in one's heart, "I am Hume"'. But this would be to suggest that self-conscious thought depends upon the interior exploitation of the conventional meaning of certain public linguistic devices, which is surely neither necessary nor sufficient for it. It could not be suggested that self-conscious thought would be beyond the reach of those who spoke a language which had no first person pronoun, and who had to refer to themselves with their own names, or that one's capacity to think about the place one occupies is dependent upon one's language possessing a device with the meaning of 'here'. These suggestions would surely get things exactly the wrong way round.

Perhaps Perry has no sympathy with these suggestions. Perhaps, contrary to first impressions, he intended by his use of the phrase 'entertaining the sense of "I am Hume"' not merely to label these ways of thinking, Perry has adopted a terminology according to which one can grasp the same thought in different ways, and when Perry speaks of Hume's grasping the thought 'Hume's sense of "I am Hume"' by entertaining the sense of "I am Hume", perhaps he means simply Hume's self-consciously thinking that he is Hume—however that is ultimately to be characterized.²⁷ Whether or not this last suggestion is correct, it seems clear that all good Fregeans must live in hope of a yet profounder philosophy.

²⁶ Perry, op. cit. p. 492. Emphasis added.

²⁷ This is certainly all I am able to understand by August's parallel talk of Hume's thinking of himself under the character of 'I'. The mere reading of the sentence is to be entertaining about the particular and particular way in which each person is presented to himself. (Yarnham, op. cit. p. 47). But I am forced to treat entertaining from the literal meaning of the phrase, under the idea that it is an form of entertaining that is, and further the object is not to be treated.

II

Semantic Theory and Tacit Knowledge

I

In his provocative paper, Crispin Wright threw down several challenges to philosophers like myself who have been attracted by, and supposed themselves to be participating in, the enterprise of constructing a systematic theory of meaning for a natural language. I shall have time this evening to take up only one of his challenges, which I hope is the most important.

Professor Wright notes that those who are interested in constructing a theory of meaning for a natural language insist that it should be what he calls 'structure-reflecting', as he says, all the interest of the theories or sub-theories which have been constructed lies in their capacity to exhibit the meanings of complex expressions as a function of the meanings of their parts. Wright then takes this 'structure-reflecting' requirement in one hand and examines various accounts of the nature of a theory of meaning which might justify its imposition. He considers three such accounts.

The first is this: the task of a theory of meaning is simply to enable one to state what each of the sentences of a language means. He argues, I think correctly, that if this is the task of a theory of meaning, the structure-reflecting requirement cannot be justified; indeed, and here again I agree with him, in the special case in which the language under study is included in the language in which the theory is being stated, a single axiom schema:

$$\text{True}(\varphi) \equiv \varphi$$

will serve the purpose.

From *Wittgenstein: To Follow a Rule*, S. Holtzman and C. Leich (eds) (London: Routledge & Kegan Paul, 1981). Reprinted by permission. Sections I, II, and IV of this paper were read at the conference in reply to the second section of Crispin Wright's, 'Rule-following, Objectivity and Theory of Meaning', op. cit., pp. 99-117. Section III is new material.

(Like Professor Wright, I will concentrate upon theories of meaning which yield statements of sentences' truth-conditions, since, as he says, his scepticism about theories of meaning arises equally for theories whose central notion is not that of truth but, say, warranted assertibility, or falsibility. In order to focus upon the question of structure, we can assume that no question is being raised about the empirical content of the *theorems* of the theory of meaning; the question is about the significance of their being derived from a finite set of principles (axioms) in a structure-revealing way.)

Wright argues that if the structure-reflecting requirement is to be justified, a theory of meaning must in some way or other be regarded as a theory of the competence of speakers of the language, but in what way? We can say that the theory states something which speakers of the language tacitly know, but what does this mean? Two interpretations provide the second and third accounts which Wright considers. One he finds relatively weak, and acceptable, but unable to justify the requirement; the other he finds inadequately explained and open to serious objection.

According to the weak sense of tacit knowledge, to attribute implicit knowledge of such a 'theory' is to do no more than obliquely to describe their behaviour; it is to say that they behave in just the way which someone would behave who successfully tried to suit his behaviour to an explicit statement [of the theory].¹ Let us call two theories of meaning which attribute the same meanings to sentences of a language—which agree in their theorems—*extensionally equivalent* (by analogy with the notion of extensional equivalence applied to grammars). According to Wright, to suit one's behaviour to a theory of meaning is to suit one's behaviour to its theorems. Consequently, if the behaviour of native speakers is the same as one who suits his behaviour to the explicitly formulated theory, T, it is the same as one who suits his behaviour to an explicit statement of any extensionally equivalent theory, T'. Hence native speakers tacitly know all extensionally equivalent theories, whether those theories *do* or *do not* differ structurally in nature.

¹ Wright, 'Rule-following, Objectivity and the Theory of Meaning', p. 100.

sentences or, in the case of a theory formulated with a single axiom schema, do not discern structure in their sentences at all. So, while this weak notion of tacit knowledge is perfectly clear, it does not provide one with the basis for preferring one extensionally equivalent theory to another.

We come then to the third account which Wright considers. It would be quite unfair to complain that Wright did not make this third option terribly clear, for it is one of his points that is not very clear. But in the absence of an explicit statement, we must rest content with hints. The notion of tacit knowledge is richer, and allows for the idea of 'unconscious deployment of information'. It is also capable of figuring in an *explanation* of a speaker's capacity to understand new sentences. Though this indicates the kind of direction in which Wright thinks one who seeks to justify the structure-reflecting requirement must be pushed, he expresses doubt about whether a genuine explanation could be provided by the use of the notion of tacit knowledge; to invoke tacit knowledge of a theory of meaning to explain a speaker's capacity to understand new sentences is vacuous, in the way in which explanations invoking the notion of a universal are vacuous. Secondly, he suggests that any such rich notion of tacit knowledge of a theory of meaning is only dubiously consistent with Wittgenstein's rule-following considerations:

the thesis seems to involve thinking of mastery of the language as consisting in (unconscious) equipment with the information which systematically settles the content of so far unconstructed and unconsidered sentences. Such a conception is far from patently coherent with the repudiation of the objectivity of sameness of use involved in the scepticism about investigation-independence sketched above.²

Though it is not my intention to focus on this aspect of his paper, I am unsure how Wittgenstein's considerations, at least as interpreted by Wright, can threaten this, as yet unborn, third option. Professor Wright says that Wittgenstein's rule-following considerations:

do not . . . impugn the legitimacy of at least the most basic purpose with which such a theory might be devised: that of securing a

² Ibid., p. 112.

description of the use of . . . the object language of such a kind that to be apprised of that description would be to know how to participate in the use of . . . the language.³

Someone who knows a finite theory for an infinite language is in some sense in possession of information which *settles in advance* (allows him to predict) the meanings of as yet unconstructed sentences. It is unclear to me why those who wish to argue that speakers of an infinite language tacitly know a finite theory of meaning need suppose the theory to determine the meanings of unconstructed sentences in any stronger, or more objectionable, sense than the one Wright implicitly accepts in the passage I have just quoted.

However good or bad the reasons for Wright's pessimism might be, he certainly threw down a challenge, and I want to take it up. I want to try to explain how the structure-reflecting requirement might be justified. But I must immediately mention two limitations on my attempt. I do not pretend that it is the only possible way of replying to Professor Wright's challenge, nor do I think that it would command universal assent among theorists of meaning. For example, I am fairly sure that Donald Davidson would dissent from it, since, contrary to what Wright suggests, he has conspicuously avoided reference to the psychological states of language users in his explanation of the nature of a theory of meaning. Second, I propose to do merely what I say: to defend the enterprise on which the structure-reflecting requirement is a constraint. I do not propose to defend the claim that philosophical insights and benefits accrue from taking the enterprise seriously.

Those who have followed recent debates in grammatical theory will be aware that Wright's challenge to the theorist of meaning is very similar to Quine's challenge to the grammarian:

Implicit guidance is a moot enough idea to demand some explicit methodology. If it is to make sense to say that a native was implicitly guided by one system of rules rather than another, extensionally equivalent one, this sense must link up somehow with the native's

³ Ibid., pp. 115-16.

dispositions to behave in observable ways in observable circumstances.⁴

However, in a way Wright's challenge is more radical, because it is not possible to formulate even a merely extensionally adequate grammar by the use of a single axiom schema. But this idea, that all a semantic theorist needs to say about English in English can be encapsulated in a single axiom scheme, must surely lead us to begin upon our task of meeting the challenge with a conviction that it can be met. Can it seriously be suggested that there is nothing to be said about the semantics of specific construction—of adverbs, tense, modality, intensional contexts, pronouns, quantifiers, proper names, definite descriptions, and the like? A good deal has already been said on these subjects by Frege, Russell, Davidson, Geach, Dummett, and many others, and though I can detect deficiencies in this work, they do not lead me to think that there are simply no questions of the kind these theorists are attempting to answer. Wright has criticized semantic theorists for ignoring the ideas of 'the most original philosophical thinker of the twentieth century', but it is surely equally deplorable if students of those ideas act as intellectual Luddites, dismissing the entirety of a sophisticated and developing intellectual tradition without a detailed consideration of its findings, and an alternative account of the enterprise to which the obviously compelling distinctions and observations it contains do properly belong. I do not say that Wright would himself join, or even encourage, the fanatics wrecking the machines, and I am prepared to concede that his challenge has not been squarely faced, but I should have liked to see a little more evidence that the questions he posed were 'expecting the answer "Yes"', rather than 'expecting the answer "No"'.
 II

Let us begin by considering a little elementary and finite language which contains ten names, a, b, c, \dots and ten

⁴ W. V. Quine, 'Methodological reflections on current linguistic theory', in D. Davidson and G. Harman (eds), *Semantics of Natural Language* (Dordrecht: Reidel, 1972), pp. 442-54.

monadic predicates F, G, H, \dots ; in all, the language has 100 possible sentences. I consider this case partly for simplicity, but also in order to stress that the structure-reflecting requirement has nothing whatever to do with *finiteness*. The fact that a language has an infinite number of possible sentences is a sufficient but not necessary condition of its having semantically significant structure, as our little language will illustrate. (It is unfortunate that Chomsky's writings have led people to equate the *creativity* of language use with the *unboundedness* natural languages display. Linguistic creativity is manifested in the capacity to understand new sentences, and the speaker of a finite language such as the one I have described can manifest it.) I want to consider two possible theories of meaning for this language. T_1 has 100 axioms: one for each sentence of the language. Examples would be:

Fa is true iff John is bald

Fb is true iff Harry is bald

...

Ga is true iff John is happy

Gb is true iff Harry is happy.

T_1 treats each of the sentences as unstructured. T_2 , on the other hand, has twenty-one axioms—one for each 'word' of the language, and a general, compositional one. Ten of the axioms are of the form:

a denotes John

b denotes Harry

and ten are of the form:

An object satisfies F iff it is bald

An object satisfies G iff it is happy.

The compositional axiom is:

A sentence coupling a name with a predicate is true iff the object denoted by the name satisfies the predicate.

One can derive from these twenty-one axioms a statement of the truth conditions of each of the 100 sentences of the

language—the very statements which T_1 takes as axioms. T_2 treats a sentence like Fa as structured; it discerns two distinct elements in it: the name a and the predicate F . Thus, T_1 and T_2 are extensionally equivalent in our sense, and our question is: what can be meant by saying that the practice of speakers of the language shows that one of these is to be preferred to the other, or equivalently, that they tacitly know one of these theories rather than the other?

It is tempting to answer this question by saying that T_1 is a theory tacitly known by someone who has had to receive training with, or exposure to the practice with, each one of the 100 sentences taken individually—someone who had not realized, or who could make no use of the fact, that the same expressions occur in different sentences—whereas T_2 is tacitly known by someone who has the capacity to understand *new* sentences of this simple subject-predicate form (e.g. Kc), provided that he has been exposed to the practice with a sufficiency of sentences containing the name c and the predicate K . However, though this contains the essence of the answer, it will not do as it stands, since it may reasonably be objected that T_1 also comprises a statement of what Kc means, so that someone who tacitly knew T_1 would be able to understand it.

I suggest that we construe the claim that someone tacitly knows a theory of meaning as ascribing to that person a set of dispositions—one corresponding to each of the expressions for which the theory provides a distinct axiom. In the case of T_1 , it is easy to see what these dispositions are: one tacitly knows T_1 iff one has 100 distinct dispositions, each one being a disposition to judge utterances of the relevant sentence type as having such-and-such truth conditions. It is more difficult to specify the dispositions which tacit knowledge of T_2 requires of a speaker, because they are interconnected. The only judgements which we are prepared to ascribe to speakers are judgements about the truth conditions of whole sentences—this, of course, is why we must speak of the knowledge of the axioms being tacit. However, if the subject tacitly knows T_2 , we shall regard any such judgement as the exercise of two distinct dispositions. Consequently, the dispositions which tacit knowledge of

T_2 requires can never be manifested singly. The dispositions must be interdefined, but though this makes the task of specifying the dispositions more difficult, it does not make it impossible.

For example we might say that a speaker U tacitly knows that the denotation of a is John iff he has a disposition such that:

$(\Pi\Phi) (\Pi\psi)$ if

- (i) U tacitly knows that an object satisfies Φ iff it is ψ
- (ii) U hears an utterance having the form $\Phi \widehat{a}$,
then U will judge the utterance is true iff John is ψ .

Connectedly, we say that a speaker U tacitly knows that an object satisfies F iff it is bald iff he has a disposition such that:

$(\Pi x) (\Pi\alpha)$ if

- (i) U tacitly knows that the denotation of α is x .
- (ii) U hears an utterance having the form $F \widehat{\alpha}$,
then U will judge that the utterance is true iff x is bald.

In these formulations, ' Π ' is a universal substitutional quantifier, with variables having the following substitution classes: Φ , names of predicate expressions of the (object) language; α , names of names of the (object) language; ψ , predicate expressions of our language (the metalanguage); and ' x ', proper names of our language.

Now, it is essential that the notion of a disposition used in these formulations be understood in a full-blooded sense. These statements of tacit knowledge must not be regarded as simple statements of regularity, for if they were, anyone who correctly judged the meanings of complete sentences would have a tacit knowledge of T_2 . When we ascribe to something the disposition to V in circumstances C , we are claiming that there is a state S which, when taken together with C , provides a causal explanation of all the episodes of the subject's V -ing (in C). So we make the claim that there is a common explanation to all those episodes of V -ing. Understood in this way, the ascription of tacit knowledge of T_2 does not merely report upon the regularity in the way in which the subject reacts to sentences containing a given

expression (for this regularity can be observed in the linguistic behaviour of someone for whom the sentence is unstructured). It involves the claim that there is a single state of the subject which figures in a causal explanation of why he reacts in this regular way to all the sentences containing the expression. Tacit knowledge of T_2 requires that there should be twenty such states of the subject—one corresponding to each expression of the language which the theory treats separately—such that the causal explanation of why the subject reacts in the way that he does to any sentence of the language involves two of these states, and any one of the states is involved in the explanation of the way he reacts to ten sentences containing a common element.

The difference between the ascription of tacit knowledge of T_1 and T_2 can be brought out diagrammatically, with the diagrams representing two extremely abstract and schematic psychological models of a subject's capacity to understand sentences. Tacit knowledge of T_1 and T_2 is incompletely represented in Figs. 1 and 2 respectively. Forget about the dotted lines for a moment. You will observe that in the

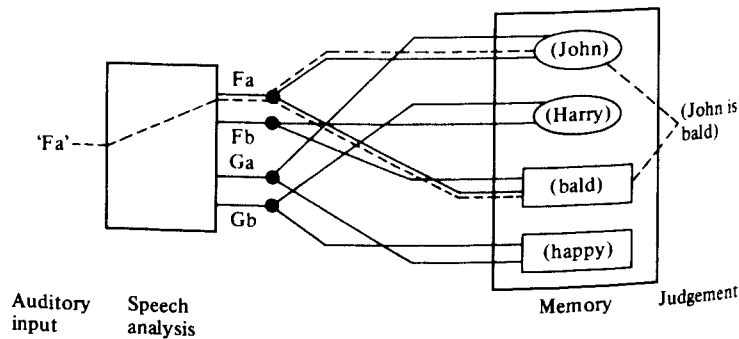


Fig. 1.

first model, there are two *independent* links between the speech-analysing device and the subject's store of knowledge of John (and in a full representation there would be ten such independent links), whereas in the second model, there is just one link between the speech-analysing device and the subject's store of knowledge of John. The dotted lines are intended to

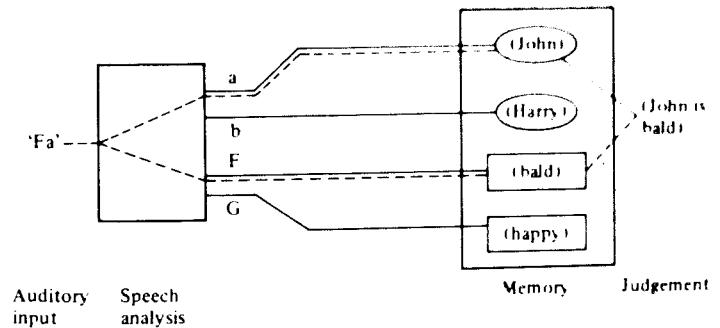


Fig. 2.

indicate what happens when the subject hears the sentence *Fa*. If you imagine dotted lines drawn to trace the consequences of the subject's hearing the sentence *Ga*, you will observe that in the former case they will not, and in the latter case they will, share a pathway with the dotted line already on the diagram. This is a representation of the fact that in the former case there is not, and in the latter case there is, a common factor which must be invoked in the explanation of the speaker's reaction to the two sentences.

It appears to me that there is a clear empirical difference between these two models of competence, and hence between tacit knowledge of T_1 and T_2 interpreted in the way I have suggested.

The decisive way to decide which model is correct is by providing a causal, presumably neurophysiologically based, explanation of comprehension. With such an explanation in hand, we can simply see whether or not there is an appeal to a common state or structure in the explanation of the subject's comprehension of each of the sentences containing the proper name *a*. However, even in the absence of such an explanation, we can have very good empirical reasons for preferring one model of competence to the other.

In the first place, we can examine the way in which the dispositions to react to sentences (the capacity to understand them) are acquired. We might find that the acquisition of the language progressed in quite definite ways, and involved a subject's acquiring the capacity to understand sentences he

had never heard before. For example, suppose a subject had progressed in his mastery of the language to the point where he understood all of the sentences which could be constructed from the vocabulary *a, b, c, d, e*, and *F, G, H, I, J*. Suppose further that he is exposed to the sentences *Ff* and *Gf* in surroundings which, or with instructions which, made it clear what they mean. Whether this exposure leads to his acquiring the capacity to understand new sentences, and which new sentences he was able to understand, would cast very considerable light upon the structure of his competence with the language, both antecedent and subsequent to the introduction of new vocabulary.

If he acquired the capacity to understand the sentences *Hf, If, Jf*, never having heard them before (and no others), this would strongly confirm the second model of his competence, for this is exactly what it predicts. On the second model, the understanding of these sentences is consequential upon the subject's possession of dispositions specific to the expressions *H, I, J*, and *f*. Further, if the second model is correct, the subject's understanding of the fragment of the language without the new name *f* showed that he possessed the first three relevant dispositions, while the exposure to just two sentences containing the name *f* can suffice for the establishment of the fourth, provided the circumstances are such that it is clear what those sentences mean. So, according to the second model, all the ingredients of understanding of some sentences are present before any of them have been heard, and they are specifically the sentences *Hf, If, Jf*.

The first model generates no predictions as to the understanding of unheard sentences. According to it, the understanding of each sentence is a separate capacity, and there is absolutely no reason why the inculcation of competence in the two sentences *Ff* and *Gf* should induce competence with any others. If it does so, this can only be accommodated on the first model by additional postulates, but why light dawns on the particular sentences *Hf, If, Jf*, rather than *Li* or *Mj*, must be left totally unexplained.

Thus, we can see more clearly what bearing the capacity to understand new sentences has upon the choice between T_1 and T_2 . The fact that someone has the capacity to understand

the unheard sentence *Hf* does not refute the ascription to him of tacit knowledge of T_1 outright—it just makes the ascription extremely implausible.

Evidence of a parallel kind can be derived from the way in which competence is lost. Suppose a subject is such that, if he loses his competence with any sentence of the form $\Phi^{\alpha}a$ (while retaining his competence with some sentences of the form Φ^{β}), he simultaneously loses his competence with every sentence of that form. This would also favour the second model of the subject's competence, since this is, once again, what that model predicts. On the first model, however, it is *inexplicable* why the loss of the capacity to understand one sentence should drag the comprehension of other sentences with it. (Evidence of this kind would be equally valuable whether the loss of competence was simply due to the subject's forgetting the meanings of words, or to brain damage.)

There is evidence of a third kind which might be used to decide between the models, since they carry with them different accounts of sentence perception. The second model requires that the subject perceive the sentence *Fa*, for example, *as* structured, that is to say, as containing the expression *a*. There is a clear difference between perceiving a sentence which does in fact contain the expression *a*, and perceiving a sentence *as* containing the expression *a*. Consequently, we can regard as relevant to the decision between the two models the various psychological tests which have been devised for identifying perceived acoustic structure, for example, the click test originally devised by Ladefoged and Broadbent.⁵

Thus, it seems to me that Crispin Wright's challenge can be met. It is possible to link tacit knowledge of one theory (rather than one of its extensionally equivalent rivals) to 'the native's dispositions to behave in observable ways in observable situations'. But to do this, one must look further than just to the dispositions to respond to, or to use, whole

⁵ P. Ladefoged and D. E. Broadbent, 'Perception of sequences in auditory events', *Quarterly Journal of Experimental Psychology* 13 (1960), pp. 162-70. See also J. A. Fodor and T. G. Bever, 'The psychological reality of linguistic segments', *Journal of Verbal Learning and Verbal Behaviour* 4 (1965), pp. 414-20.

sentences. It is possible that the scepticism which Wright expresses is due in a small way to the mistaken thought that facts of native usage which bear upon the content possessed by whole sentences are all the facts to which a theory of meaning can or need be sensitive.

Bearing in mind the interpretation of 'tacit knowledge' which I have proposed, let us briefly consider infinite languages. What would be involved in the tacit knowledge of the theory for such a language which is formulated with the use of a single axiom schema?

An axiom schema is not a theory; it is a compendious specification by their syntactic form of the sentences which do constitute the theory. In the case of an infinite language, there are an infinite number of such sentences. Someone would possess tacit knowledge of such a theory only if he possessed an infinite number of *distinct* linguistic dispositions: one corresponding to each of the sentences of the language. This we know no one can possess. I have concentrated upon the more challenging case of a finite language because it is important to stress the point that there can be compelling evidence that someone *does* not possess a battery of distinct dispositions other than the fact that no one *can* possess them, so that we may want to do for a speaker of a finite language what we are forced to do for a speaker of an infinite language.

Infinity in language results from recursiveness; syntactic and semantic rules which operate upon their own output. A standard clause for the recursive element 'and' runs like this:

A sentence of the form S 'and' S' is true iff S is true and S' is true

Generalizing the procedure used earlier, we can say that a speaker U tacitly knows this principle iff he has a disposition such that:

(ΠS) (ΠP) ($\Pi S'$) ($\Pi P'$) if:

- (i) U is disposed to judge that S is true iff P
- (ii) U is disposed to judge that S' is true iff P'
- (iii) U hears an utterance having the form S 'and' S' , then U will judge that the utterance is true iff P and P' .

(The substitution classes for the variables S and S' , and P and P' , are names of sentences of the object language, and sentences of the metalanguage, respectively.) More difficult is the clause for an existential quantifier. It might run:

A sentence of the form $(\exists x)\widehat{\Phi}$ is true iff there is something y such that, letting β be its name, the sentence $\Phi^{\beta/x}$ is true.

This is a simplified 'Fregean' clause for an objectual quantifier.⁶ ' $\Phi^{\beta/x}$ ' abbreviates 'the result of substituting β for all occurrences of x in Φ '. Then we can say that U tacitly knows this principle iff he has a disposition such that:

($\Pi\Phi$) ($\Pi\psi$), if:

- (i) ($\Pi\beta$) (Πx) (If U tacitly knows the denotation of β is x , then U is disposed to judge that the sentence $\Phi^{\beta/x}$ is true iff x is ψ)
- (ii) U hears an utterance of the form $(\exists x)\widehat{\Phi}$ then U is disposed to judge that the utterance is true iff something is ψ .

(In this instance, the substitution class for Φ are structural-descriptive names of object-language propositional functions in the variable x .)

III

Professor Wright and I are agreed that tacit knowledge of the semantic rules of a language is a 'logical construction' out of the use of whole sentences. I have suggested that the idea that we may tacitly know one rather than another of two extensionally equivalent theories which differ in the amount of structure they discern in sentences leads one to the thought of a correspondence between the separable principles of a theory and a series of internal states of the subject, dispositionally characterized. Nevertheless, I would agree with Wright that to regard these states as states of knowledge or belief, that is to say, states of the same kind as are

⁶ For an account of the 'Fregean' approach to quantifiers, see my paper 'Pronouns, Quantifiers and Relative Clauses (I)', *Canadian Journal of Philosophy* 7 (1977), pp. 467-536. [Reprinted as chap. 4 in this volume. Ed.]

identified by the ordinary use of those words, is wrong and capable of leading to confusions of the kind he gestures at. To establish this point would require another paper. However, I shall say a brief word about it now, since I believe it is to this point, rather than the very idea of a psychological underpinning to the theory of meaning, that Wright's criticisms are legitimately directed.

There is no doubt in what the similarity between the states of tacit knowledge and the ordinary states of knowledge and belief is taken to consist. At the level of output, one who possesses the tacit knowledge that *p* is disposed to do and think some of the things which one who had the ordinary belief that *p* would be inclined to do and think (given the same desires). At the level of input, one who possesses the state of tacit knowledge that *p* will very probably have acquired that state as the result of exposure to usage which supports or confirms (though far from conclusively) the proposition that *p*, and hence in circumstances which might well induce in a rational person the ordinary belief that *p*. But these analogies are very far from establishing tacit knowledge as a species of belief. After all, similar analogies at the level of input and output exist between the state of a rat who avoids a certain food which has upset it in the past ('bait-shyness') on the one hand, and the belief that a man might have that a certain food is poisonous, on the other.

It is true that many philosophers would be prepared to regard the dispositional state of the rat as a belief. But such a view requires blindness to the fundamental differences which exist between the state of the rat and the belief of the man—differences which suggest that fundamentally different mechanisms are at work. We might begin with this disanalogy: the rat manifests the 'belief' in only one way—by not eating—whereas there is no limit to the ways in which the ordinary belief that something is poisonous might be manifested. The subject might manifest it by, for example, preventing someone else from eating the food, or by giving it to a hated enemy, or by committing suicide with it. These variations stem from the different projects with which the belief may interact, but similar variations arise from combining the belief with other beliefs. It might, for example,

lead to a subject's consuming a small amount of the food every day, when combined with the belief that the consumption of small doses of a poison renders one immune to its effects. (The existence of other beliefs induces a similar variability in the ways in which the belief that something is poisonous might be established.) It is of the essence of a belief state that it be at the service of many distinct projects, and that its influence on any project be mediated by other beliefs. The rat simply has a disposition to avoid a certain food; the state underlying this disposition is not a part of a system which would generate widely varying behaviour in a wide variety of situations according to the different projects and further 'beliefs' it may possess.

So, one who possesses a belief will typically be sensitive to a wide variety of ways in which it can be established (what it can be inferred from), and a wide variety of different ways in which it can be used (what can be inferred from it)—if we think of plans for intentional action as being generated from beliefs by the same kind of rational inferential process as yields further beliefs from beliefs. To have a belief requires one to appreciate its location in a network of beliefs; this is why Wittgenstein says, 'When we first begin to *believe* anything, what we believe is not a single proposition, it is a whole system of propositions. (Light dawns gradually over the whole.)'.⁷ To think of beliefs in this way forces us to think of them as structured states; the subject's appreciation of the inferential potential of one belief (e.g. the belief that *a* is *F*) at least partly depending upon the same general capacity as his appreciation of the inferential potential of others (e.g. the belief that *b* is *F*). After all, the principle of an *inference*, of *reasoning*, can never be specific to the set of propositions involved.⁸ Possession of this general capacity is often spoken of as mastery of a concept, and the point I am making is frequently made by saying that belief involves the possession of concepts (e.g. the concept of *poison*). Behind the idea of a system of beliefs lies that of a system of concepts, the structure of which determines the inferential

⁷ L. Wittgenstein, *On Certainty* (Oxford: Blackwell, 1969), sect. 141.

⁸ See T. Nagel, *The Possibility of Altruism* (Oxford: Clarendon Press, 1970), chap. 7.

properties which thoughts involving an exercise of the various component concepts of the system are treated as possessing. At the ground floor of the structure will be observational concepts whose possession requires the subject to be able to discriminate (in suitable favourable circumstances) instances of the concept. Inferential links connect these concepts with more theoretical concepts 'higher' in the structure, and they in their turn will be connected with concepts yet more remote from observation.

Concepts are exercised in the first instance in thoughts; beliefs may be regarded as dispositions to entertain thoughts in the 'believing mode'—i.e. to make judgements; if we think of belief in this way, we shall not be prepared to attribute to a subject the belief that *a is F* (for some particular object *a*, and property *F*) unless we can suppose the subject to be capable of entertaining the supposition (having the thought) that *b is F*, for every object *b* of which he has a conception. For example, we will not be inclined to explain a subject's actions by attributing to him the belief that *he is F* (e.g. he is in pain) unless we suppose him capable of entertaining the supposition with respect to individuals distinct from himself that they are *F* (e.g. that the person is in pain).⁹ Now, it is true that the 'believing mode' of thought cannot be characterized without reference to its influence upon the subject's actions; the traditional accounts of belief which I am largely following went wrong in trying to identify the difference between judgement and mere thought in terms of some introspectible feature of accompaniment of the thought. A judgement is (*ceteris paribus*) a thought one acts upon (if a suitable plan for action is derived from it). But we are now far away from the rat's disposition to avoid certain food. For one thing, there is an enormous gap between belief (a disposition to judge that *p* when the question whether *p* is raised) and action. Even though a subject believes a substance is poisonous he may not slip it to an enemy he wants to kill and knows no other way to kill either because he 'didn't think of it' or because, having thought of it, he 'forgot' what to do when the time came.

Tacit knowledge of the syntactic and semantic rules of the

⁹ See P. F. Strawson, *Individuals* (London: Methuen, 1959), chap. 3, sect. 4.

language are not states of the same kinds as the states we identify in our ordinary use of the terms 'belief' and 'knowledge'. Possession of tacit knowledge is exclusively manifested in speaking and understanding a language; the information is not even potentially at the service of any other project of the agent, nor can it interact with any other beliefs of the agent (whether genuine beliefs or other tacit 'beliefs') to yield further beliefs. Such concepts as we use in specifying it are not concepts we need to suppose the subject to possess, for the state is inferentially insulated from the rest of the subject's thoughts and beliefs. There is thus no question of regarding the information being brought by the subject to bear upon speech and interpretation in rational processes of thought, or of making sense of the subject's continued possession of the information despite incorrect performance, due to his 'not thinking' of the rule at the appropriate time etc. Remarks which Wright makes ('unconscious deployment of information' etc.) suggest that he considers the proponent of his third option as holding that tacit knowledge is a real species of belief, but with all the relevant inferential processes made by the subject somehow taking place outside his ken. This is certainly a mysterious and confused position.

I disagree with Professor Wright only in denying that the proponent of a structure-reflecting theory of meaning need have anything to do with it.

IV

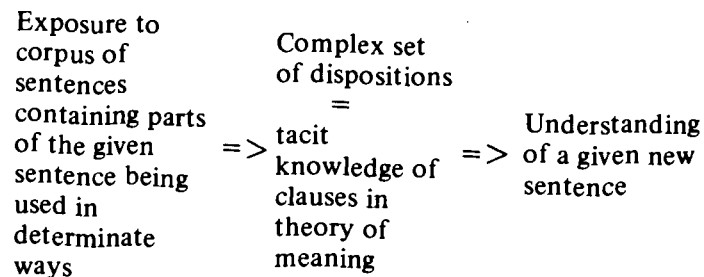
I come finally to the question of whether there is any sense in which the theorist of meaning provides an explanation of a speaker's capacity to understand new sentences. Now it is implicit in what has gone before that the notion of tacit knowledge of a structure-reflecting theory of meaning, explained as I have explained it, cannot be used to explain the capacity to understand new sentences. I have given a purely dispositional characterization of tacit knowledge, and though this does not just amount to a re-description of the speaker's capacities to understand sentences (including the new ones) what it provides, in addition to a description of those capacities, is not itself something that could be

involved in an explanation of them. The surplus concerned the form which an explanation of the capacities would take, and to say that a group of phenomena have a common explanation is obviously not yet to say what the explanation is. So I agree with Wright when he writes:

there is no reason to think that it is within the power of the sort of theory which we are considering to serve up anything which could rightly be considered an *explanation* of the infinitary character of competence with its object-language.¹⁰

But while I agree with this, I disagree with his claim that a proponent of structure-reflecting theories of meaning must somehow be committed to the view that they are providing an explanation of speakers' capacity to understand new sentences.

Nevertheless, I believe that there is a way of explaining a speaker's capacity to understand new sentences to which provision of a structure-reflecting theory of meaning is indispensable. For we can provide a genuine explanation of a speaker's capacity to understand a certain novel utterance by citing his exposure, in the past, to the elements of that sentence occurring in sentences whose meaning was, or was made, manifest. I envisage an explanatory chain like this:



This chain can be genuinely explanatory even though the last link of the chain by itself is not. The attribution to a subject of tacit knowledge of T_1 is neither more nor less explanatory of the capacity to understand the given new sentences than the attribution of T_2 , but tacit knowledge of T_2 belongs in an

¹⁰ Wright, *op. cit.*, p. 113.

explanation of the capacity to understand a new sentence because we understand how those dispositions might have been acquired as a result of exposure to the corpus of utterances which the subject has heard. Now, we can cite a subject's exposure to a corpus of utterances in explanation of his capacity to understand a new one only if we believe that the use of expressions in the new sentence is *in conformity with* their use in the previously heard corpus. Only in this case will we be able to show what set of dispositions the subject might have acquired which meets the two conditions: (1) exercising them yields the observed (and correct) interpretation of the new sentence; (2) would have been exercised in, and hence could have been acquired by exposure to, the previous use. Consequently, when a capacity to understand novel sentences is observed, the theorist of meaning has an indispensable role to play in its explanation, since he must exhibit the regularity between the old and the new.

I have more or less deliberately spoken in terms which might well offend some of those present, for I believe that some of those present, though not, I think, Professor Wright, believe that Wittgenstein's arguments on rule-following show that the ambition to exhibit such regularity must be based upon some kind of mistake. Perhaps this is so, and if it is so, I hope that we shall hear tonight why it is so. Since I do not have time to discuss the arguments of such philosophers, let me end by addressing two connected questions to them.

Is it their opinion that all capacities to understand novel sentences (to *know*, I stress, what they mean) are equally inexplicable, or do they believe that scope is provided for one kind of explanation of how it is that a speaker knows what a new sentence means when and only when it can be shown to contain elements which also occur in sentences with whose use he is already familiar?

On the assumption that their answer to my first question is 'Yes', I come to my second:

Do they think that it is sufficient to provide an explanation of the kind which the occurrence of familiar expressions makes possible simply by showing that the new sentence

does contain expressions which also occur in sentences with whose use he is already familiar, or do they believe, in view of the evident possibility of ambiguity, that something else must be provided? If so, how does this further part of the explanation differ from a statement of the regularity between the old use and the new?

Does Tense Logic Rest upon a Mistake?

This paper is a critical examination of the semantic foundations of tense logic. It belongs in a *Festschrift* for Donald Davidson not because the subject is one he has worked on, but quite simply because, without the benefit of his work on the theory of meaning, I should not have been able to write it. I hope that this proves the tribute I intend it to be.

Arthur Prior, the founder of tense logic, wrote:

If [an] expression constructs a sentence out of one other sentence it is an adverb or adverbial phrase, like 'not' or 'It is not the case that', or 'allegedly' or 'It is alleged that', or 'possibly' or 'It is possible that' . . . I want to suggest that putting a verb into the past or future tense is exactly the same sort of thing as adding an adverb to the sentence. 'I was having my breakfast' is related to 'I am having my breakfast' in exactly the same way as 'I am *allegedly* having my breakfast' is related to it, and it is only a historical accident that we generally form the past tense by modifying the present tense, e.g. by changing 'am' to 'was', rather than by tacking on an adverb. In a rationalized language with uniform constructions for similar functions we would form the past tense by prefixing to a given sentence the phrase 'It was the case that', or 'It has been the case that' . . . and the future tense by prefixing 'It will be the case that'.¹

Prior himself constructed such languages, and gave birth to the flourishing discipline of tense logic. To a considerable extent, tense logic has continued in the direction which Prior's initial impetus gave it, as a branch of, or at least a very close relative of, modal logic, and therefore it continues to rest upon Prior's original and fundamental assumption: that 'It was the case that' is an operator 'out of the same box as' 'It is

This paper, previously unpublished, was completed in the Spring of 1979 for inclusion in *Essays on Davidson: Truth and Interpretation*, B. Vermazen and M. B. Hintikka (eds.) forthcoming from Oxford University Press. Copyright © Antonia Phillips 1985. Various versions of the ideas in this paper have been presented in seminars in Oxford since 1974. I would like to thank Martin Davies, Kit Fine, Graeme Forbes, Toomas Kaarmo, John McDowell, Christopher Peacocke, and Barry Taylor for giving me their reactions to my ideas.

¹ *Papers on Time and Tense* (Oxford: Clarendon Press, 1968), p. 7.

possible that', and 'It is not the case that'. This assumption is clearly a semantic assumption, and stands in need of a justification which can only be provided by the construction of a theory of meaning which shows that the meaning of a past-tense statement depends upon the meaning of a present-tense statement in the same kind of way that the meaning of 'It is possible that (P)' depends upon the meaning of P. It is my contention that no adequate theory of this kind, and hence no adequate justification of tense logic as it is currently practised, has ever been provided.

I know that the claim that tense logic awaits a semantic justification will be greeted with astonishment by its practitioners, who will cite the existence of a 'semantics' for tense logics which is exactly parallel to the 'possible-worlds semantics' which have been provided for modal logics.² Such semantic theories employ as their central concept that of 'truth-at-a-time- t ' ('true _{t} ') which is supposed to be parallel to the notion of 'truth-with-respect-to-a-possible world' ('true _{w} '). Typically, these semantic theories for tense-logical languages are model theories, but there is no difficulty in principle in fixing upon a single, designated interpretation. In this case, for primitive tense logical sentences (which I shall represent by the schematic letters 'X', 'Y', etc.) there will be clauses of the form:

(1) For any time t , true _{t} (X) iff it is raining at t ,

and tense operators will be treated by recursive clauses like:

(2) For any time t , and any sentence S , true _{t} ('P' $\widehat{}$ (S)) iff there is a time t' , earlier than t , such that S is true _{t'} .³

² See, e.g., G. J. Massey, *Understanding Symbolic Logic* (New York: Harper & Row, 1970), App. H; and R. P. McArthur, *Tense Logic* (Dordrecht: Reidel, 1976), chap. 1.

³ Theories of this kind involve explicit quantification over times, and should not have been looked on with favour by Prior, who insisted (correctly, I think) that tense operators do not involve an ontology of times. (See *Papers on Time and Tense*, p. 118.) The inconsistency between the claim that tense operators are ontologically neutral and the adoption of the standard semantics is mentioned in G. Harman's review of P. T. Geach's *Logic Matters*, in *Foundations of Language* 13 (1975), pp. 127-33. Prior should insist upon a homophonic semantics in a tensed metalanguage, as is anyway suggested by his informal exposition: "It has been the case that Professor Carnap is flying to the moon" is true if and only if the present-tense statement "Professor Carnap is flying to the moon" has been

People presumably have a theory like this in mind when they say that tense logic is concerned with 'propositions whose truth-value varies with time'.

We can be content with this theory only to the extent that we understand its central notion, 'true _{t} ', and it is by no means clear what it means. For a reason I shall give in a moment. I do not believe that a theory without implications for the correctness or incorrectness of utterances can lay claim to the title of a semantic theory, and it is not clear what is the connection between 'true _{t} ' and such an evaluation of utterances. There are three possible interpretations of 'true _{t} ', and I shall consider the three semantic theories, T₁, T₂, T₃, which result when clauses like (1) and (2) are interpreted in these three different ways.

In what follows, I shall be working with a broadly Fregean conception of semantics; a species of the genus whose merits Davidson has so capably brought out. If we ignore context-dependence for a moment, a semantic theory for a language, on this conception, will have at its heart a *theory of reference* — a theory which assigns to every meaningful expression of the language something that can be regarded as that expression's *semantic value*. Frege himself advanced a theory according to which proper names and sentences have particular objects and truth-values as their respective semantic values, but neither of these choices is forced upon us by the general conception. The only fixed point is this: we should so choose semantic values for expressions that a theory of reference can *serve as* a theory of sense. Any formulation of the theory of reference will, of course, involve the references of expressions being specified in a particular way, and one of the indefinitely many formulations of the theory of reference must be capable of being selected and regarded as *showing* the sense of the expressions of the language.⁴ To say that

true.' (*Time and Modality* (Oxford: Clarendon Press, 1957) p. 9.) Important though the difference between such a semantic theory and that embodied in (1) and (2) is for certain purposes, it is irrelevant to those of the present paper, since the questions I shall raise about 'true _{t} ' have exact parallels for the significantly tensable predicate 'true' used in a tensed metalanguage.

⁴ The use of Wittgenstein's term in this connection is due to Dummett; see 'Frege's Philosophy', in *Truth and Other Enigmas* (London: Duckworth, 1978), p. 105, and Frege (London: Duckworth, 1973), p. 227.

a particular formulation of the theory of reference shows the sense of the expressions of a language is to say that one whose knowledge of the language consists only in his explicit knowledge of the theory of reference, thus formulated, would thereby be placed in a position to speak and understand it.⁵ Thus, both the sentences:

(3) 'Snow is white' is true

(4) 'Snow is white' is true iff snow is white

assign to the sentence 'Snow is white' the value True, but in (4) that value is identified in a way which allows us to say that (4) shows or displays the sense of the sentence. To think of sense in this way, as what is shown or displayed by one among many formulations of the theory of reference, is to take seriously Frege's metaphor of sense as a mode of presentation of reference, which certainly does not encourage us to expect that the sense of an expression can be given save in the course of actually presenting the reference. Finally, the requirement that the theory of sense be compositional, that it show how the senses of complex expressions are a function of the senses of their parts, is captured by the requirement that the theory of reference be compositional, that it show how the semantic value of a complex expression is a function of the semantic value of its parts.

It is the requirement that a theory of reference be capable of serving as a theory of sense which explains why any such theory must yield a characterization of the semantic values of particular utterances. If knowledge of the theory is to equip someone with a capacity to speak and understand the language, it seems reasonable to require that it enable him to determine that certain utterances are correct or incorrect. Such an evaluation of particular utterances is always implicit in theories of reference (truth) for classical, non-context-dependent languages, even though the theory explicitly treats

⁵ For him to *know* that he is in this position, he will additionally have to know something about the evidential basis of the theory of reference—e.g., that it is exclusively based upon evidence about how speakers use the language. For further clarification of how a theory of reference might serve as a theory of sense, see D. Davidson, 'Reply to Foster', in G. Evans and J. H. McDowell (eds), *Truth and Meaning* (Oxford: Clarendon Press, 1976), pp. 33–41.

only of sentence-types; but it is not usually mentioned because it is so obvious: we understand an ascription of the value True to a sentence-type as entailing that any utterance of that sentence-type is a correct utterance. The study of context-dependent languages merely forces us to be explicit about something that is implicit in our standard semantical practice; a semantic theory must ultimately assign semantic values to utterances (or to sentences-in-context), and must show how the semantic value of complex utterances (of complex sentences-in-context) is a function of the semantic values of simpler utterances (of simpler expressions-in-context).

I

The first interpretation of 'true_t', or of the phrase 'propositions whose truth value varies with time', is by far the most radical. On this interpretation, a proper appreciation of the semantic functioning of tense requires us to abandon the idea that particular historical utterances of tensed sentences are assessable, once and for all, as correct or incorrect. Rather, we must acknowledge that the evaluation of particular utterances must change as the world changes. Suppose *x* uttered 'It is raining' at *t*, when it is raining. Then, at *t*, his utterance was correct, but now it has ceased to rain, it has become incorrect.

We arrive at this position if we conceive of the connection between the ascription of 'true_t' to sentence-types and the evaluation of utterances on the model of the connection between the ascription of 'true' to sentence-types of a classical (tenseless) language and the evaluation of utterances of those types. For classical languages, we can say:

$$(5) (\forall S)(\forall u) [Of(S, u) \supset (Correct(u) \equiv True(S))],$$

and on the revolutionary interpretation of 'true_t' we can say

$$(6) (\forall S)(\forall u)(\forall t) [Of(S, u) \supset (Correct-at-t(u) \equiv True_t(S))].$$

According to T₁, the semantic value of an utterance is immediately derivable from the semantic value assigned by the theory to the type of which it is an utterance, just as in

classical languages. All utterances of the same type have the same semantic value, no matter when those utterances are made, but the fact that the sentence type is assigned a set of pairs of times and truth values as its semantic value is a reflection of the fact that an utterance is not to receive a single assessment as correct or incorrect, but rather an assessment which varies with time.

T₁ must be sharply distinguished from the third position I shall discuss, according to which 'true_{*t*}' is to be interpreted in such a way as to warrant:

$$(7) (\forall S)(\forall u)(\forall t)[(\text{Of}(S, u) \wedge \text{At}(t, u)) \supset (\text{Correct}(u) \equiv \text{True}_t(S))].$$

T₁ rejects an absolute 'correct' predicate of utterances; T₃ exploits it. T₁ advances the revolutionary idea that the evaluation of an utterance as correct or incorrect depends upon the time the *evaluation* is made (and so the evaluation varies); T₃ registers the commonplace idea that the evaluation of an utterance depends upon the time the *utterance* is made (and does not vary). According to T₁, the semantic value of each utterance of a tensed sentence is the same; according to T₃, the semantic value of an utterance of a tensed sentence depends upon when the utterance is made.

T₁ is such a strange position that it is difficult to believe that anyone has ever held it, but I detect it in some of the things that Prior says. For example, Prior cites, as an antecedent of his views on tensed statements, the views of Scholastic logicians, as interpreted by Geach:

For a Scholastic, 'Socrates is sitting' is a complete proposition, *enuntiable*, which is sometimes true, sometimes false; not an incomplete expression requiring a further phrase like 'at time *t*' to make it into an assertion.⁶

In another passage, also noted by Prior, Geach wrote:

May not the Stoics well have thought that though the truth value of

⁶ P. T. Geach, Critical Notice of J. Weinberg's *Nicolaus of Autrecourt*, *Mind* 58 (1949), pp. 238–45, quoted in *Past, Present and Future* (Oxford: Clarendon Press, 1967), p. 15.

'Dion is alive' changes at Dion's death, the sentence *still expresses the same complete meaning (lekton)?*⁷

I am not sure what view Geach means to be attributing to the Scholastics and Stoics. To say that the sentence-type 'Socrates is sitting' is a *complete* proposition, expresses a *complete* meaning, seems to imply that, as with classical sentences, to know what assertion is being made by an utterance of a tensed sentence all you need to know is which tensed sentence was uttered; you do not need further information to tie the tensed sentence down to a particular time, such as might be provided by the additional phrase 'at time *t*'. If this is what Geach intends, it would follow that such an 'assertion' would not admit of a *stable* evaluation as correct or incorrect; if we are to speak of correctness or incorrectness at all, we must say that the assertion is correct at some times and not at others.

Such a conception of assertion is not coherent. In the first place, I do not understand the use of our ordinary word 'correct' to apply to one and the same historical act at some times and not at others, according to the state of the weather. Just as we use the terms 'good' and 'bad', 'obligatory' and 'permitted' to make an assessment, once and for all, of non-linguistic actions, so we use the term 'correct' to make a once-and-for-all assessment of speech acts. Secondly, even if we strain to understand the notion 'correct-at-*t*', it is clear that a theory of meaning which states the semantic values of particular utterances solely by the use of it cannot serve as a theory of sense. If a theory of reference permits a subject to deduce merely that a particular utterance is now correct, but later will be incorrect, it cannot assist the subject in deciding what to say, nor in interpreting the remarks of others.⁸ What should he aim at, or take the others to be aiming at? *Maximum* correctness? But of course, if he knew

⁷ P. T. Geach, review of B. Mates's *Stoic Logic*, in *Philosophical Review* 64 (1955), p. 144. Geach makes it clear in this review that he does not claim that this conception of a proposition is coherent.

⁸ Defenders of Prior have denied that T₁ will have the consequence that an utterance made at *t* can become incorrect at *t'*, on the ground that utterances are temporary and perishing existents. But this is no help if we can say (as we surely must be able to say) of another utterance made at *t'* both that it has the same semantic value as the original sentence, and that it is incorrect.

an answer to this question, it would necessarily generate a once-and-for-all assessment of utterances, according to whether or not they meet whatever condition the answer gave. In fact, we know what he should do; he should utter sentence types true at the time of utterance. One who utters the sentence type 'It is raining' rules out dry weather only at the time of utterance; he does not rule out later dryness, and hence there can be no argument from the later state of the weather to a re-appraisal of his utterance. Utterances have to be evaluated according to what they rule out, and so different utterances of the same tensed sentence made at different times may have to be evaluated (once and for all) differently. They cannot therefore all be assigned the same semantic value.⁹

There is no objection to introducing an equivalence relation, 'says the same thing as', according to which all utterances of the same tensed sentence type say the same thing, nor to introducing an abstract object—a 'Stoic-proposition'—on the back of this relation. What we cannot do is to regard a Stoic-proposition as a complete proposition, in the sense that describing an utterance in terms of the Stoic-proposition it puts forward is a complete semantic description of the utterance. On this point, therefore, we must agree with Frege:

A thought is not true at one time and false at another, but it is either true or false, *tertium non datur*. The false appearance that a thought can be true at one time and false at another arises from an incomplete expression. A complete proposition or expression of a thought must also contain the time datum.¹⁰

⁹ I do not intend to rule out all tensing of the 'correct' predicate of utterances by this argument. Perhaps it makes sense to claim, as many have claimed, that a future tense utterance 'Tomorrow (P)' made on day 1 becomes correct (or incorrect) on day 2. But this notion of variable truth value is quite distinct from that employed in the semantic clauses (1) and (2), despite Prior's claim to the contrary (*Past, Present and Future*, p. 16). After all, unless there is also going to be a sea battle on day 3, the sentence type 'Tomorrow (There is a sea battle)' becomes *false* on day 2, not true. To avoid the very complex issues of determinism and fatalism, which are irrelevant to my limited, semantic point, I have concentrated on the case of the past tense.

¹⁰ G. Frege, *Kleine Schriften*, ed. I. Angelelli (Hildesheim: Georg Olms, 1967), p. 338.

I am sorry to have laboured these obvious points, especially since T_1 may have been held by no one. It is useful to have one sense of the phrase 'a proposition whose truth value varies with time' distinguished and out of the way, especially since it allows us to bring out an important difference between tense and modal logic. For it seems to me that the view which many people hold of the notion 'true_w' is one parallel to T_1 , but no arguments of the kind I have advanced against T_1 are effective against it. It is perfectly coherent to regard our utterances as putting forward thoughts which are true in some worlds and not in others. (Let us understand 'true-in- w ' on the model of 'bald-in- w '; whether or not this counts as an analysis, the statement that x is bald in w is understood in such a way as to entail that if w were actual, x would be bald, and so, equally, to say that the thought is true in w , is to say that, if w were actual, the thought would be true.) We do not have to insist that, really, the thought expressed by an utterance of a contingent sentence is that the actual world (the world of utterance!) is thus and so, and therefore to insist that *that* very thought would not have been true if things had gone differently. Why is this? The reason is simple; the ascription to thoughts of truth-in- w (truth that varies with worlds) is consistent with the ascription to them of simple truth, and hence is consistent with the straightforward ascription of correctness or incorrectness to utterances. Indeed, given the fundamental principle that there is a world ('the actual world') such that all and only truths are true with respect to that world, we may *derive* the absolute truth conditions of sentences (and utterances) from these 'truth-in- w ' conditions. Whatever structural parallels there are between time and modality, there are undeniable differences in *substance*, as, for instance, the difference between the outright assertion of the correctness of an utterance (correctness as things actually are) and the dubiously intelligible assertion of the present correctness of an utterance (correct at the present time) brings out. It is a consequence of these substantial differences that a theory of reference which assigns to a sentence as its semantic value a mapping of worlds to truth values can serve as a theory of sense in a way in which a theory which assigns

to a sentence a mapping from times to truth values cannot. To fail to see this difference between the theories is to be blinded by the structural parallels between time and modality to their substantial differences. After all, the knowledge that a course of action involves present happiness but future suffering must have an impact upon my decision-making which the knowledge that it involves happiness in the actual world, but suffering in other possible worlds will not, and it is upon this difference that the difference between the theories rests.

II

On the second interpretation of 'true_t', it has no intrinsic connection with the truth/correctness of utterances. According to T₂, the notion 'true_t' which occurs in clauses like (1) and (2) should be regarded as a notion wholly internal to the semantic theory, like 'satisfies' in a Tarskian theory. Truth conditions for utterances can only be derived from it by means of an additional stipulation, which is needed to complete the theory. In view of this, it would be better to imagine (1) and (2) rewritten with some novel term, say 'Obtains(*S*, *t*)', in place of 'true_t', and the connection made with the correctness of utterances via the additional principle:

$$(8) (\forall S)(\forall u)(\forall t)[(\text{Of}(S, u) \wedge \text{At}(t, u)) \supset (\text{Correct}(u) \equiv \text{Obtains}(S, t))].$$

Clearly (8) is very similar to (7), the principle which defines T₃, but there is an important difference; T₃ regards the principle (7) as guaranteed by the meaning of 'true_t', and hence as in no sense an additional stipulation. T₃ states the semantic recursion using a notion with immediate consequences for the correctness of utterances throughout — T₂ does not.

A parallel might help to make T₂ clearer. A supervaluational semantics recursively characterizes a semantic relation 'true-with-respect-to-a-valuation', or, for vague languages, 'true-with-respect-to-a-sharpening'. No one supposes that the absolute evaluation of utterances as true or false is somehow implicit in these relative-truth clauses, and hence, no one

supposes that a theory which only characterizes this notion is complete. (There is no such notion as that of 'the intended valuation', or 'the sharpening in force'.) Since the languages studied are not normally context-dependent languages, the required additional connection with absolute truth can be made at the level of sentence types, along the lines of:

$$(9) (\forall S)(\text{True}(S) \equiv (\forall v)(\text{True}_v(S)))$$

(i.e. a sentence type is true simpliciter iff it is true upon all sharpenings/valuations), but (9) is equivalent to:

$$(10) (\forall S)(\forall u)[\text{Of}(S, u) \supset (\text{Correct}(u) \equiv (\forall v)(\text{True}_v(S)))].$$

On the present interpretation of 'true_t', the theory stated in clauses like (1) and (2) stands equally in need of completion, and (8) is regarded as completing it in exactly the way (9) or (10) complete a supervaluational theory.

In effect, T₂ has the consequence that the utterance of a present-tense sentence (a present-tense sentence in context) has *two* semantic values, and is (perhaps harmlessly) ambiguous. Clause (1) of T₂ gives it a semantic value akin to that of a predicate—a set of times—while clause (8) assigns to it a semantic value appropriate to a complete speech act—a truth value. Just to give them names, let us call the first of these a 'predicate-value' and the second a 'sentence-value'. Now, a past-tense sentence in context is also assigned two semantic values, but neither of these is a function of the *sentence-value* which the present-tense sentence possesses in that context. Consequently, Prior's logical language, with its formulae like:

$$(11) X \supset P(F(X))$$

cannot be regarded by one who holds T₂ as a logically perfect language. A 'rationalized', or logically perfect language would separate the two semantic roles which an expression like 'X' in Prior's notation can play in a given context. For example, we could leave 'X' and 'Y' as schematic letters for expressions which have, in any given context, predicate-values, and form expressions which have, in any given context, sentence-values, by the addition of some 'closing' operator.

say 'N' (for 'it is now the case that'). Thus, in place of (11) we would write:

$$(12) N(X) \supset N(P(F(X)))$$

which clearly exhibits that it is the semantic value which X has in any given context, and not that which N(X) has in that context, which is an argument to the function yielding the semantic value which the sentence N(P(F(X))) has in that context. For a language of this kind, we could replace (8) with:

$$(13) (\forall S)(\forall u)(\forall t) [(Of('N', \widehat{S}), u) \wedge At(t, u)] \supset \\ (Correct(u) \equiv Obtains(S, t)).$$

In fact, similar considerations seem to me to apply to languages for which supervaluational semantic theories are given: such theories also make sentences-in-context ambiguous. All sentences, and hence all utterances of sentences, are assigned two semantic values: one, a set of valuations or sharpenings, and the other, a truth value. A logically perfect language would distinguish the two roles played by sentences, differentiating the non-sentence S , which has a semantic value of the first kind, from the sentence $\#(S)$, which has a value of the second kind. This would be achieved by replacing (9) or (10) with a clause for the operator '#':

$$(14) (\forall S)(True(\widehat{\#}(S)) \equiv (\forall v)(true_v(S))).$$

In such a language, we would be able to clearly differentiate a statement like:

$$(15) \#(S \text{ or not-}S)$$

from an instance of the Law of the Excluded Middle like:

$$(16) \#(S) \text{ OR NOT}(\#(S)).^{11}$$

Although T_2 , with its distinction between X and N(X), appears to bring nothing but the benefits of clarity, Prior was always very much opposed to it. He argued that the

¹¹ I have capitalized the truth functions, 'OR' and 'NOT' to distinguish them from the operators in (15) which do not map truth values on to truth values, but sets of valuations/sharpenings on to sets of valuations/sharpenings. There is a problem of unifying these two uses, similar to that posed for T_3 at the end of this section.

introduction of an operator like 'N' would make tense logic impossible.

To say that such an operator [a present-tense operator] is necessary is to say that the expressions to which we attach it would not be propositions, at all events not tensed propositions, without it. This in turn is to say that tense operators do not form propositions out of propositions . . . rather they form propositions out of merely juxtaposed nouns and verbs, or they form tensed propositions out of untensed ones. And from this it in turn follows that tense operators cannot be iterated . . . and to rule this would be practically to destroy tense logic before we have started to build it.¹²

But this argument is fallacious. Let us call the non-sentential expressions, like X and Y in the new language, 'situation expressions' or SEs. 'N' is therefore of category S/SE.¹³ But there is no reason why we should suppose that the past and future-tense operators are of the same category. The fact that these operators iterate shows that they are of the category SE/SE—mapping 'predicate-values' on to 'predicate-values', as specified by a clause like (2). This allows for strings of tenses of any length, though to form a complete sentence, the string must be 'closed' with an occurrence of 'N'.

Even if a sense of semantic fastidiousness encourages us to adopt a tense-logical notation which incorporates the operator 'N', all the principles and technical results of tense logic would no doubt remain expressible and derivable. Indeed one can suppose the operator already to be in use if one supposes that it is written as a blank and pronounced as a pause. However, there is one difficulty which the categorial indeterminacy of Prior's sentences enabled him to hide. Tense-logical systems must allow for the occurrence of connectives like 'and' and 'not' within the scope of tense operators, and thus they must be of the category SE/SE, SE or SE/SE. Only in this way can we form expressions like:

$$(17) N(P(X \wedge Y)).$$

Standard clauses for the connectives in tense-logical semantic theories run like this:

¹² *Time and Modality*, p. 10. A similar argument is found in *Past, Present and Future*, pp. 14–15.

¹³ I use the notation which D. K. Lewis uses for stating a categorial grammar; see his 'General Semantics', in D. Davidson and G. Harman (eds), *Semantics of Natural Languages* (Dordrecht: Reidel, 1972), pp. 169–218.

$$(18) (\forall S)(\forall S')(\forall t) (\text{true}_t(S \wedge S') \text{ iff } \text{true}_t(S) \text{ and } \text{true}_t(S')),$$

giving conjunction a basically temporal meaning (of simultaneity); the role of conjunction in connecting sentences with a truth-value—sentences which may have no temporal content at all—is not accounted for. Formally, the problem is parallel to that posed by the occurrence of the connectives both inside and outside the scope of quantifiers, and is amenable to similar solutions.¹⁴ A theory like T_2 appears to be committed to a solution of a Tarskian kind, on which (18) would remain the only statement of the semantic function of ' \wedge ', but on which its occurrence between complete sentences is accounted for because such sentences, including, for example, mathematical ones, are given ' true_t ' conditions—or conditions under which times satisfy them—though in such a way that true sentences are true at all times. Truth can then be defined as truth at all times, and the truth-functional role drops out from (18) as a special case. Interestingly enough, the alternative, 'Fregean' solution to the problem is represented by the theory which results from the third interpretation of ' true_t '—to which I now turn.

III

The difficulty we found with the first interpretation of ' true_t ' was that it did not provide for the stable evaluation of utterances as correct or incorrect. The difficulty we found with the second interpretation of ' true_t ' was that it provided for such an evaluation (and consequently, for an appropriate semantic value for tensed sentences in context) only by means of an additional stipulation, and this forced the acknowledgement of an ambiguity in tensed sentences in context, according to whether that additional stipulation did, or did not, have to be reckoned with. The third interpretation promises to steer between the difficulties of its predecessors. According to T_3 , ' true_t ' is always to be understood in such a way as to validate:

¹⁴ I have discussed this problem, and solutions to it, in my paper, 'Pronouns, Quantifiers and Relative Clauses (I)', *Canadian Journal of Philosophy*, 7 (1977), pp. 471–7. [See chapter 4, Ed.]

$$(7) (\forall S)(\forall u)(\forall t) [(Of(S, u) \wedge At(t, u)) \supset (\text{Correct}(u) \equiv \text{True}_t(S))],$$

so the semantic recursion proceeds throughout in terms of a notion with immediate consequences for the truth of utterances. T_3 , therefore, appears to assign to tensed sentences in context a unique, appropriate semantic value, while at the same time, clauses like (2) remain in force, apparently exhibiting the dependence of the semantic value of a past-tense sentence in context upon the semantic value of present-tense sentences in context.

However, it is important to be clear about the novelty of this proposal, for it involves the recognition of a hitherto unknown form of embedding. In all previously-studied forms of embedding, including the truth-functional and modal forms which Prior claimed to be exactly parallel, the semantic value which a complex sentence $\Sigma(e)$ has in a given context is a function of the semantic value which the expression e has *in that context*. This principle applies whether or not the embedded sentence contains context-dependent expressions. The semantic value which 'Possibly (I am eating my breakfast)' and 'It is not the case that (I am eating my breakfast)' have in any given context, is a function of the semantic value which 'I am eating my breakfast' has in that context. But T_3 asserts that the semantic value which the sentence 'P(X)' has in a context is a function of the semantic value which X would have in *another* context. For, on the present interpretation, the recursive clause (2) says *roughly*¹⁵ that the utterance of 'P(X)' is true iff the utterance of X at some earlier time would have been true. If T_3 is right, the interpretation of a tensed utterance forces us to consider the interpretation which other, perhaps only potential, utterances would have, and this is a quite unprecedented feature.

A parallel, resembling one given by Prior, might help to make the situation clear. Suppose that there is a language exactly like English, save that it possesses two additional operators, 'To the right', and 'To the left', which can be prefixed to sentences in the first person. A sentence like

¹⁵ I shall turn almost immediately to a more careful statement of the principle.

'To the left (I am hot)' as uttered by a speaker x at t is true iff there is at t on x 's left someone moderately near who is hot. These operators can iterate; for example, 'To the left (To the right (I am hot))' is entailed by, but does not entail, 'I am hot'. What makes it compelling to acknowledge that these sentences involve the new form of embedding is the occurrence of the first person pronoun within the scope of the operators. If, but apparently only if, we suppose that these operators are governed by the rule that a sentence of the form 'To the left' (S) is true, as uttered by x at t iff there is someone moderately near to the left of x such that, if he were to utter the sentence 'I am hot' at t , what he would thereby say is true, we can generate the postulated truth conditions, while continuing to suppose that the only role of the first person pronoun is that of denoting the speaker. Principle (2), as interpreted by T_3 , is a semantic rule of exactly the same kind as that which governs the operators in this hypothesized language.¹⁶ However, if this example serves to bring out the coherence of the new form of embedding, it surely highlights its novelty. There is nothing in 'Possibly (I am hot)' and 'Allegedly (I am hot)' to prepare us for it.

Many people will deny that T_3 does provide a coherent account of the semantics of a tensed language, on grounds that are perhaps familiar. Any such semantics must allow that the sentence 'In the past (There are no speakers)', as uttered now, expresses a truth, and to this end, it must be the case that there is a time t' , earlier than now, such that 'There are no speakers' is true _{t'} . But this cannot mean that, had someone uttered the sentence at t' , he would have spoken correctly for he would not have done so. If this is right, one may well feel that the principles which T_3 uses to derive the semantic value of one utterance from the semantic value of another, potential, utterance, are just not true.

¹⁶ Prior's example involves a language without pronouns in which the utterance of the simple sentence-type 'Hot' is true iff the speaker is hot. This provides a closer parallel to English tenses, but only at the cost of making the recognition of the new form of embedding considerably less compelling. A theory parallel to T_2 , which separates the two roles played by the expression 'Hot' would do equally well. The example is found in *Papers on Time and Tense*, pp. 134-44, and in *Worlds, Times and Selves* (with Kit Fine), (London: Duckworth, 1977), pp. 28-50.

Though this objection underlines the need for a more careful explanation of 'true _{t} ' than the rough approximation I have offered so far, I am not inclined to regard it as a decisive objection to T_3 . In the first place, though I am doubtful whether English incorporates any examples of the new form of embedding, it seems perfectly clear to me that a language could coherently do so. We could surely learn to play the game with the two operators 'To the left' and 'To the right', and in such a way that the sentence 'To the left (I am silent)' can be correctly uttered. Secondly, the difficulty to which the objection points is a perfectly general one. Every semantic theory which has the required consequences for the evaluation of utterances must employ a notion similar to 'true _{t} ' as interpreted by T_3 , and the objection is to an over-simple counterfactual analysis of this notion, rather than to the specific use to which it is put by T_3 . For example, Donald Davidson employs in his theory of reference the three-place relation 'True(S , x , t)', which he glosses as 'The sentence S is true as (potentially) uttered by x at t '.¹⁷ Davidson's truth theory will yield a theorem like:

$$(19) (\forall t)(\forall x)(\text{True}(\text{'I am silent'}, x, t) \equiv x \text{ is silent at } t)$$

which, by instantiation, entails:

$$(20) \text{True}(\text{'I am silent'}, \text{G.E.}, t^*) \equiv \text{G.E. is silent at } t^*.$$

With a suitable choice of t^* , we will have:

$$(21) \text{True}(\text{'I am silent'}, \text{G.E.}, t^*)$$

i.e. there is a time, t^* , such that the sentence 'I am silent' is true as potentially uttered by G.E. at t^* , but this cannot mean that, had G.E. uttered the sentence 'I am silent' at t^* , he would have spoken truly.

There must be a coherent explanation of the relation Davidson employs other than the simple counterfactual one, and there is no reason why T_3 should not employ it also. For the only novelty in T_3 is that, in evaluating a particular utterance, changes are rung on the temporal parameter in the truth relation, while for Davidson they remain fixed. Nor does it seem impossible to identify the required interpretation of

¹⁷ D. Davidson, 'Truth and Meaning', *Synthese* 17 (1967), p. 319.

'true_t'. We want to speak of the *actual* value of a *potential* utterance; a sentence type is true_t iff, were anyone to utter it at *t*, what he *would* thereby say *is* (as things stand) true. Using an operator 'A' for 'Actually', and David Lewis's notation '□→' for the counterfactual conditional, we can write:¹⁸

$$(22) \text{true}_t(S) \text{ iff } (\exists u)(\text{Of}(S, u) \wedge \text{At}(t, u)) \square \rightarrow \\ (\exists u)(\text{Of}(S, u) \wedge (\text{At}(t, u) \wedge \text{A}(\text{True}(u)))).$$

Although I believe that T₃ does provide a coherent account of the semantics of tensed sentences, I do not think that there is much evidence about the use of tense in English that should lead us to prefer T₃ to a more conservative theory like T₂. There certainly is no direct evidence, parallel to the occurrence of the pronoun 'I' in our imaginary language. Far from it; as Prior remarked:

English speakers find it hard to see these things quite clearly; for in English sentences the point of view of the speaker dominates even subordinate clauses . . . on the few occasions on which we use phrases like 'It was the case that' in English, they are not followed by the present but the past; we say 'It was the case that he *was* sick', not 'It was the case that he *is* sick'.¹⁹

In particular, 'now' is 'temporally rigid'; we certainly do not say 'It was the case that he is now sick', as we might if our language embodied the new form of embedding. However, there is the possibility of an interesting form of direct evidence. Do the very same connectives which join the complete sentences, 'and', 'or', etc., also occur within the scope of tense operators? Are there sentences which correspond to the Priorean formula 'P(X ∧ Y)', whose truth requires the existence of a time at which X and Y are *simultaneously* true? If so, it is something for which T₃ provides a very neat explanation. T₃ can take the *truth*-functional role of the connectives as basic, for by the time we come to compute the impact of '∧' in the sentence 'P(X ∧ Y)' we shall be considering its role in a potential earlier utterance of the sentence 'X ∧ Y'. Since the truth of both of these (present-tensed)

¹⁸ The present formulation of (22), a considerable improvement over an earlier formulation of my own, is due to Christopher Peacocke.

¹⁹ *Past, Present and Future*, p. 14.

sentences requires the occurrence of a suitable state of affairs at that earlier time of utterance, the truth of the whole (potential) utterance will require the *simultaneous* occurrence of those states of affairs, and so the requisite truth conditions will be given to the past-tense utterance.²⁰ This very simple explanation of the role of the connectives inside tense operators compares so very favourably with the cumbersome 'Tarskian' solution to which T₂ seems committed, that we must allow such uses of the connectives to constitute a fairly weighty consideration in favour of T₃, were we able to find any. However, it is far from clear that there are such uses in natural English.²¹

Whether or not English is a language of the kind tense logicians have studied is not the central question of this paper, and I have agreed that T₃ does provide an account of a possible form of embedding. My main concern is to stress its novelty; if T₃ provides the intended interpretation of tense-logical formulae, then there simply is not the parallel between tense and modal logics which Prior and others have claimed. For, to repeat, the semantic value a complex tensed sentence possesses in a context is, according to T₃, a function of the semantic value which the embedded sentence would possess in another context; this is not true of the semantic values of complex statements of a modal logic, or indeed of any other known logic.

This conclusion can be resisted in one of two ways. One can argue that T₃ does not involve the novel form of embedding I have ascribed to it, or one can argue that modal logic does. I shall briefly consider arguments of both kinds.

I have interpreted T₃ as assigning to sentences semantic values which depend upon the context in which they are

²⁰ I hope this makes it clear why I regard T₃ as embracing the 'Fregean' solution to the problem posed at the end of pt. II.

²¹ The force of any example would be dissipated, if it could be shown that the tense-operator concerned is a 'metric-operator', in Prior's sense, like 'It was the case *n* units ago that . . .', since such operators distribute over the connectives. As Barbara Partee points out (in 'Some Structural Analogies between Tenses and Pronouns in English', *Journal of Philosophy* 70 (1973), p. 602), a great many past-tense statements 'concern a particular time'. She wrongly thinks that the point rules out the recognition of tense operators in English, forgetting Prior's metric tense operators, but the point she makes does lessen the likelihood of finding clear evidence of the kind mentioned in the text.

uttered; in the words of David Lewis, semantic values that are *variable but simple*. But why should we not regard T_3 as assigning to tensed sentences *constant but complicated* semantic values—i.e. functions from contexts to the previously mentioned simple values?²² If this suggestion was adopted, then the (complicated) semantic value of 'P(X)' in a context c (which is the same as in any other context) is a function of the (complicated) semantic value which X possesses in that context (as in any other), and no novel form of embedding would be contemplated.

However, this proposal does not satisfy the constraint upon assignments of semantic value which derives from our requirement that a theory of reference be capable of serving as a theory of sense. If we try to suppose a theory assigning constant but complicated semantic values serving as a theory of sense, we have to suppose that there is a single piece of knowledge (about the way in which the truth value of a tensed sentence type depends upon context) which *suffices* for the understanding of all utterances of that sentence, even in the absence of any knowledge of when the utterances are made. But there is no such piece of knowledge. One certainly cannot be said to understand an utterance of a tensed sentence if one knows only which sentence it is an utterance of, and the general way in which the truth value of that tensed sentence depends upon the time of utterance—one needs, in addition, to know which time the sentence concerns, and hence, when it was uttered. Similarly, it is not sufficient to understand an utterance of the sentence 'You are hot' to know the general truth that 'You' refers to the person being addressed; one needs to know who is being addressed.

To pursue the second line of objection, and to argue that modal sentences involve the kind of embedding which T_3 discerns in tensed sentences, one must construe 'true_w' in the way T_3 construes 'true_t'. I have already tried to suggest that there is no reason to construe 'true_w' in this way, but what happens if we do? To say that there is a world w such that 'G.E. is an only child' is true_w is then to say that, if

²² See D. K. Lewis, 'Index, Context, and Content', presented at the Anniversary Symposium on Philosophy and Grammar, Uppsala University, June 1977.

someone were to utter the sentence 'G.E. is an only child' in w (as a sentence of English, it goes without saying) what he would have thereby said is, as a matter of fact, true. Since I am not an only child, he could not be taken to have expressed a world-neutral proposition, true in some worlds and false in others, since *that* proposition is, as a matter of fact, false. So the hypothesized utterance, and presumably any utterance, must be taken to express a 'world-indexed' proposition, asserting in this case that G.E. is an only child in w . On this view, we must regard ordinary statements as containing a reference to the world of utterance, in the way tensed statements contain a reference to the time of utterance. To think in this way about possible worlds seems to commit one to an unacceptable form of modal realism—the doctrine that other possible worlds exist in exactly the same sense in which the actual world exists, and differ from it merely in not being the ones we happen to inhabit.²³ A sane view of possible worlds begins when we conceive of the actual world as the world in which all and only true propositions are true. Such a conception of the actual world presupposes the conception of world-neutral propositions, and requires a reading of 'true_w' parallel to T_1 , not T_3 .

A great deal more needs to be said, but it would take us away from tense logic and deep into the study of modality. I hope to have given at least some plausibility to my claim that if T_3 is adopted, tense logic involves a quite novel form of embedding. If T_3 does provide the intended interpretation of tense-logical formulae, it is high time that the disguise by which Prior smuggled his baby into the family of Logics—its supposed family resemblances to more familiar logics—be cast aside, and the subject be seen for what it is.

²³ There is a difficulty here which I will ignore, which arises from the fact that we, and our utterances, exist in many different possible worlds.

Molyneux's Question

I

William Molyneux posed the following question in a letter to Locke:

Suppose a man born blind, and now adult, and taught by his touch to distinguish between a cube and a sphere of the same metal, and nighly of the same bigness, so as to tell, when he felt one and the other, which is the cube, which the sphere. Suppose then the cube and sphere placed on a table, and the blind man to be made to see; *quaere*, Whether by his sight, before he touched them, he could now distinguish and tell which is the globe, and which the cube?¹

This question aroused tremendous interest among philosophers and psychologists on both sides of the Channel, so much so that Cassirer was able to claim that it formed the central question of eighteenth-century epistemology and psychology.² In fact, Molyneux's Question raises many different issues, not all of which are of as great interest now as they were in earlier times. Nevertheless, there seems to me to be one issue at the heart of the controversy which is very important and upon which it is still possible to make some progress. In this introductory section, I shall try to identify this issue as sharply as I can. The issue I want to discuss arises only on the assumption that the blind possess genuine spatial concepts, and so I shall begin by saying a word about that background assumption. Having stated the question as precisely as is possible, I shall distinguish it from other issues

This paper is previously unpublished. Copyright © Antonia Phillips, 1985. [Written during the Winter of 1978, the paper was read at several universities in the first half of 1979. It must be emphasized that in its present form it is a first draft: Evans intended to make substantial revisions, and to incorporate new material. Had he had the opportunity of writing a final version, and preparing it for publication, he would have made acknowledgements where due. Ed.]

¹ Quoted in Locke's *Essay Concerning Human Understanding* II, ix, 8.

² E. Cassirer, *The Philosophy of the Enlightenment* (Boston, Mass.: Beacon Press, 1951), p. 108.

which have become entangled with it. In the second section, I shall try to make clear what is involved in the presumed ability of the blind to perceive spatially, and in the final section, I consider how the spatiality of vision is to be thought of in relation to that ability.

Even before proceeding with this programme, I want to make two points about the terms in which the question is posed. First, I shall make a simplification of the situation originally suggested by Diderot.³ Molyneux asked whether or not the blind man would be able to apply three-dimensional spatial concepts, such as *sphere* and *cube*, upon the basis of his newly acquired vision; he also seems to have been interested in whether the blind man would see things as at various distances from him.⁴ There is nothing wrong with these questions, but they are in a sense less interesting than whether he would be able to extend two-dimensional concepts, like *square* and *circle*, since there is less antecedent expectation that the newly sighted man would be able to appreciate the depth cues available in visual perception. Although most disputants in the controversy retained the terms in which Molyneux originally posed it, it transpires that their fundamental disagreement is about whether the blind man would be able to extend his two-dimensional concepts.⁵

Secondly, I must stress that Molyneux's Question is about

³ Diderot, *Lettre sur les Aveugles*, translated in M. J. Morgan's *Molyneux's Question* (Cambridge: Cambridge University Press, 1977), p. 108.

⁴ See the original letter Molyneux sent to Locke in 1688, four years before the one Locke cites in the *Essay* (in the Bodleian Library (MS Locke c. 16, fol. 92)). I owe this reference to W. von Leyden, *Seventeenth Century Metaphysics* (London: Duckworth, 1968), p. 277. See also the letter no. 1064 in *The Correspondence of Locke*, vol. 3, ed. E. de Beer (Oxford: Clarendon Press, 1978).

⁵ John Mackie claims, in his *Problems from Locke* (Oxford: Clarendon Press, 1976), p. 30, that Locke's negative answer was only to Molyneux's original question, and that he would have answered 'Yes' to Diderot's later version. I do not think that this interpretation of Locke can be sustained. See especially the passage immediately following the discussion of Molyneux's Question: '... sight ... conveying to our minds the ideas of light and colours which are peculiar only to that sense; and also the far different ideas of space, figure and motion, the several varieties whereof change the appearance of *its proper object*, viz. light and colours; we bring ourselves by *use* to judge of the one by the other.' (*Essay* II, ix, 9 —my italics.) This seems to express the straight Berkeleyan position on visual perception of space, and would obviously require a negative answer to Diderot's version of Molyneux's Question also.

whether a born-blind man *who can see* a circle and a square would extend his concepts to them. It is not a question about how soon after the operation, and via what process, a newly sighted man would be able to see. Molyneux's Question requires only that the newly sighted man would be able to have visual experiences of circles and squares without his, or his brain's, having had a chance to establish correlations between the old and the new information.

Several thinkers have returned a negative answer to Molyneux's Question not because they held a view about the difference between the spatial concepts of the blind and those connected with visual experience, but because they held that the blind do not have any genuinely spatial concepts at all. The first explicit statement of this view that I have been able to find occurs in Platner's *Philosophische Aphorismen* (1793), though antecedents can certainly be found in Berkeley's *New Theory of Vision*. Platner wrote:

In reality, it is time that serves, for the man born blind, as space. Remoteness and proximity only mean to him the time, more or less long, and the number, more or less, of intermediaries which he needs in passing from one tactual impression to another.⁶

Echoing this, Lotze wrote:

the space of a blind man may not be so much what we mean by space, as an artificial system of conceptions of movement, time and effort.⁷

The line of thought behind this position can be seen in the following passage from von Senden, the most recent, and most dogmatic, of its proponents:

Nothing is given to the blind man simultaneously, either by touch or the other senses; everything is resolved into successions . . . Only the variety furnished by a temporally ordered series of experiences can furnish him with knowledge . . .

Since nothing is given simultaneously to his senses as spatial, it must be mentally strung together in time . . . A spatial line must be replaced by a temporal sequence.⁸

⁶ Quoted in J. S. Mill, *An Examination of William Hamilton's Philosophy* (London: Longmans, 1872), pp. 283-4.

⁷ H. Lotze, *Metaphysic II* (Oxford: Clarendon Press, 1887), pp. 272-3.

⁸ M. von Senden, *Space and Sight*, trans. P. Heath (London: Methuen, 1960), pp. 285-6.

On this view, the perceptions of the blind cannot represent several objects existing simultaneously; they amount to no more than the *succession* of tactual experiences, perhaps linked by, or embroidered with, kinaesthetic and muscular experiences. When a blind man traces the outline of a square—e.g. by tracing the edge of one face of a cube, or a wire figure—he receives just such a sequence of impressions, and his concept of *square*, though it is abstract, and can be applied, for example, to the arrangement of houses in the village, remains the concept of a certain kind of *succession*, like the concept of a *fugue* or a *tune*.

It is not surprising that those who hold this view about the spatial concepts of the blind will say 'Not' to Molyneux's Question; as Bain put it:

But how a vision to the eye can reveal beforehand what would be the experience of the hand or the other moving members I am unable to understand.⁹

The implication here is perhaps not watertight, but I shall not elaborate since the claim about the spatial concepts of the blind surely cannot be sustained. It is true, and I think important, that spatial concepts are concepts that involve the thought of distinct objects or elements existing simultaneously. Nor do I wish to quarrel with the view that someone who has spatial *concepts* must also have the ability to form, or enjoy, *perceptual* representations of distinct but simultaneously existing objects. Of course, not all spatial concepts need relate immediately to perceptual experience. Once some spatial concepts are possessed, a great many other spatial concepts, like that of a *fifteen-sided plane figure*, can be understood by means of their relation to those basic concepts. Even what we like to think of as very elementary spatial concepts—such as *triangle* and *square*—can only fully be grasped via their connection with simpler concepts, such as *between*, *longer*, *straight*, *angle*, etc. But ultimately, possession of some concepts will require a capacity to use the concept in response to the appropriate perceptual experience, and if they, and all concepts explicable in terms of them, are to have the essential character of applying to

⁹ A. Bain, *The Senses and the Intellect*, 2nd edn., p. 376

arrangements of distinct but simultaneously existing objects, then the perceptions to which they relate must involve the presentation or representation of distinct objects. It follows that it is not sufficient for the possession of genuine spatial concepts that one can correctly use spatial terms of a public language, for it is possible that this could be done without the appropriate perceptual representations; in much the way that someone might be able to apply colour terms correctly by analysis of wavelength of light, or as someone might be able to apply 'to the right' and 'to the left' to sounds simply upon the basis of the difference in the time the sound waves meet the two ears, and with no spatial meaning at all.¹⁰

My disagreement with this tradition comes at the point where it is claimed that the blind cannot form or enjoy perceptual representations of space—of distinct objects existing simultaneously and related spatially. I do not wish to object, as one might object, on the ground that the blind can perceive distinct things simultaneously, by touching them with both hands, for example, or by feeling the parts of something small enough to be placed within the hand. For, although this is true, to locate this as the point of objection is to concede far too much to the theory of perception on which the position is based. I would still want to object to the conclusion even if the blind had only a single hand in a world in which all the objects were too large to be grasped in it.

It is unacceptable to argue from the successiveness of *sensation* to the successiveness of *perception*. One can surely make sense of the idea of a perceiving organism which uses a sequence of impressions or stimulations to build up a perceptual representation of a spatial world, in which the information contained in the sequence of stimulations is integrated into, or interpreted in terms of, a unitary representation of the perceiver's surroundings. As Gibson has been at pains to point out, the supposedly unproblematic spatial perception provided by vision depends to a very considerable

¹⁰ This last example comes from T. G. R. Bower's book *Development in Infancy* (San Francisco, Calif.: Freeman, 1974), p. 29. I quite agree with Bower that it is not enough to establish mastery of spatial concepts to show that babies are capable of being conditioned to respond differentially to stimuli whose differences we describe in spatial terms.

extent upon precisely such a process.¹¹ Now, I do not pretend that this concept is clear; indeed, it is one of the main tasks of this paper to get it clear. For the time being I shall appeal to an intuitive understanding of the idea. The point is made at this intuitive level by Pierre Villey, a blind Montaigne scholar who was obviously riled by the suggestion that the blind did not have genuine spatial concepts (and by the implication of intellectual inferiority that often went with it), and he wrote a book, from which the following quotation is taken, in order to refute it. Villey concedes that his tactual perception of the chair is successive, but writes:

. . . if, an hour after feeling it, I search in my consciousness for the memory of the vanished chair . . . I do not reconstruct it by means of fragmentary and successive images. It appears immediately and as a whole in its essential parts . . . There is no procession, even rapid, of representations . . . I couldn't tell in what order the parts were perceived by me.¹²

Of course, the attribution to a subject of a perceptual representation of space is to be justified upon the basis of behavioural and reasoning skills which it explains, but this anecdotal evidence is not worthless. If von Senden and his predecessors are right, Villey's memory of a chair would have to be ordered in time; yet his incapacity to remember the order in which he perceived the chair, or to acknowledge a succession in what is remembered, directly contradicts this.

There is, indeed, the possibility of a deeper argument on this point. Kant argued with tremendous force that it was not possible to have a conception of an objective world—a world whose states and constituents are independent of one's perception of them—without conceiving of that world as spatial, with oneself as located within it and tracing a continuous path through it. The argument connects objectivity with space—but space as conceived in 'genuine' spatial concepts, i.e. concepts which connect up with simultaneous perceptual representations of the world. (I shall occasionally call these 'simultaneous representations' and the concepts

¹¹ J. J. Gibson, *The Senses Considered as Perceptual Systems* (London: Allen & Unwin, 1968), chap. 13.

¹² Pierre Villey, *The World of the Blind* (London: Simpkin, Marshall, Hamilton, Kent & Co., 1922), pp. 183–4.

based upon them 'simultaneous concepts'.) If the concepts of the blind were really no more than concepts of certain kinds of succession in experience, then they could not be regarded as having a concept of an independently existing world at all. The blind would have to think as the phenomenologists would have us all think, in terms of actual and possible sense experiences.

In what follows, I shall assume that the blind do have simultaneous perceptual representations of space, and hence that they have genuine, simultaneous spatial concepts. Although there are many experiments which show that the blind are inferior to the sighted in spatial ability and reasoning,¹³ there are equally many which show the blind to be capable of behaviour which seems to show that they are able to form simultaneous representations of the world—whether in perception, or imagery.¹⁴ The issue I want to discuss, and which I believe was the issue on which Locke, Berkeley, Leibniz, Condillac, and others were taking up positions, is that of the relation between the perceptual representations of space attributable to the blind, and the perceptual representation of space available in visual perception. For this reason, I shall not consider the school of thought, represented most notably by J. S. Mill and Henri Poincaré, who held, partly with a view to avoiding the consequence that the spatial concepts of the blind and the sighted are different, that *our* spatial concepts have precisely the character assigned

¹³ See, e.g., S. Millar, 'Spatial Memory by Blind and Sighted Children', *British Journal of Psychology* (1975); 'Spatial Representations by Blind and Sighted Children', *Journal of Experimental and Child Psychology* (1976), and L. C. Hartlage, 'Development of Spatial Concepts in Visually Deprived Children', *Perceptual and Motor Skills* 42 (1976), pp. 255–8.

¹⁴ B. Jones, 'The Spatial Perception of the Blind', *British Journal of Psychology* (1975); M. Robin and P. Pecheux, 'Problèmes Posés par la reproduction des modèles spatiaux chez les enfants aveugles', *Perception* (1976); J. Juurmaa, 'Transposition in Mental Spatial Manipulation' (American Foundation for the Blind Research Bulletin no. 26, June 1973), pp. 87–134. For some further general reflections on the spatial concepts of the blind, see G. Revesz, *The Human Hand* (London: Routledge & Kegan Paul, 1958) and *The Psychology and Art of the Blind* (London: Longmans, 1950); K. Lashley, 'Psychological Problems in the Development of Instrumental Aids for the Blind', in P. A. Zahl (ed.) *Blindness* (New York: Haffner, 1962); G. Warnock, 'Significance for Philosophy', Appendix to von Senden, *op. cit.*; and D. Rosencranz and R. Suslick, 'Cognitive Models for Spatial Representations in Congenitally Blind, Adventitiously Blind and Sighted Subjects', in *New Outlook for the Blind* 70 (1976), pp. 188–94.

by von Senden and his predecessors to the concepts of the blind. Poincaré wrote:

To localize an object simply means to represent to oneself the movements that would be necessary to reach it. It is not a question of representing the movements themselves in space, but solely of representing to oneself the muscular sensations which accompany these movements and which do not presuppose the existence of space.¹⁵

These philosophers take account of a point which is of the greatest importance in the theory of space perception, but distort it. This fundamental point is that no explanation can be given of what it is to have a perceptual representation of space—to be given perceptually the information that objects of such-and-such a character are arranged in such-and-such a way in one's vicinity—except in terms of the behavioural propensities and dispositions to which such information gives rise. The distortion comes in at the point where an attempt is made to *reduce* spatial propositions to propositions about time, kinaesthetic and tactual sensation—at the point, that is, where Poincaré suggests not only that the information embodied in these perceptual representations cannot be explained without reference to bodily movements, but also that the concepts of the relevant bodily movements *can* be explained without reference to space. I shall return to this point below.

I have not placed Berkeley of the *New Theory of Vision* in the Mill–Poincaré tradition, despite the fact that there is much that he says with which they would agree. When Berkeley is thinking about 'outness'—distance away from the observer—he belongs squarely with Mill and Poincaré. In the *Principles*, he summarized the position he thought he had established in the *New Theory of Vision*:

. . . in strict truth the ideas of sight, . . . , do not suggest or mark out to us things actually existing at a distance, but only admonish us that ideas of touch will be imprinted in our minds at such and such distances of time, and in consequence of such and such actions.¹⁶

¹⁵ H. Poincaré, *The Value of Science* (New York: Dover, 1958), p. 47. See also J. S. Mill, *An Examination of William Hamilton's Philosophy*, 13.

¹⁶ *Principles of Human Knowledge*, ed. M. R. Ayers (London: Dent, 1975), sect. 44, p. 89.

But I do not think that Berkeley held that the *two-dimensional* concepts which we are able to form on the basis of touch—like the concept he called *tangible square*—were concepts of succession, and so I do not think that his reason for giving a negative answer to Molyneux's Question is the same as von Senden's. Berkeley is quite explicit:

Sounds, for example, perceived at the same instant, are apt to coalesce, if I may so say, into one sound: but we can perceive at the same time great variety of visible objects, very separate and distinct from each other. Now *tangible extension being made up of several distinct co-existent parts*, we may hence gather another reason that may dispose us to imagine a likeness or an analogy between the immediate objects of sight and touch.¹⁷

In view of this, it is certain that in answering Molyneux's Question negatively, Berkeley was taking up a position upon the most fundamental issue posed by that question—an issue which only arises on the assumption that the blind do have simultaneous spatial concepts. It is to that issue that I now turn.

To bring it out, let us consider first the position of a philosopher, whom I shall call 'V', who insists that, on the conditions given, the newly sighted man must be able to apply his concepts to the visually presented array—must, to use a convenient term, 'generalize'. (Curiously enough, there is no historical figure who has taken exactly this position; as I will explain, even Leibniz's 'Yes' to Molyneux's Question indicates a position weaker than V's.) For V, the case presented by Molyneux's Question is no different from that involved in the following speculation: whether a man born deaf, and taught to apply the terms 'continuous' and 'pulsating' to stimulations made on his skin, would, on gaining his hearing and being presented with two tones, one continuous and the other pulsating, be able to apply the terms correctly. Few of us have a doubt about the outcome of this experiment, but, more important, if the born-deaf man failed to apply the terms in this new case, we should feel obliged to interpret this as casting doubt upon his understanding of the terms which we thought we had introduced to him, just as if he had, incomprehensibly, been unable to apply the terms to

¹⁷ *New Theory of Vision*, ed. Ayers, sect. 145, p. 51 (my italics).

stimulations made on a hitherto unused part of his skin. We should say that he had not fully mastered the concept of *pulsation* simply because he had been presented with a case to which the concept manifestly applied, and had failed to apply it. No obligation attaches to one who holds this position to provide an alternative interpretation of the subject's previous utterances; it is not, indeed, part of his position that they are intelligible. What he does claim is that there is a unitary conceptual capacity which most people have with the word 'pulsating' which this born-deaf man must be acknowledged to lack.¹⁸

In the same way, V holds, there is a unitary conceptual ability associated in the case of most adults with the word 'square'—mastery of a single concept. Now, the fact of the matter is that this concept *applies* in the case of a visually presented square, or in the case of four points of light arranged in a square. If a man does not perceive the shape, or the points of light, then his inability to apply the term 'square' in no way casts doubt upon his understanding of it. But if we build into the description of the case that the newly-sighted man does see the square, or the four dots—does, if you like, have a visual experience of the same character as leads the normally sighted to apply the term—then he is presented with something which falls under the concept, and an incapacity to apply the term in this new case must show that he does not possess the (unitary) concept of a square.

V finds it impossible to conceive of a coherent concept which applies to items simultaneously existing, which is therefore exercised most directly in connection with perceptual representations of several distinct objects existing simultaneously, and yet which stops short of applying in the case where the items are visually perceived. V simply cannot find room for *two* genuine, i.e. simultaneous, concepts of a square, or for *two* genuine concepts of *between*, *straight line*, etc., each set of concepts generating its own geometry.

If V is sophisticated, he will make no appeal to the notion

¹⁸ [Evans subsequently noted that there is a disanalogy between this case and Molyneux's, for in the former the concepts *continuous* and *pulsating* apply literally to the experience itself, rather than figuring in the specification of its representational content. Ed.]

of *similarity*. His position is not that the tactual perceptual representation of a square *resembles* the visual perceptual representation of a square. His point is that if both are simultaneous representations, the only concept which he can understand applies (or seems to apply) in both cases.¹⁹

The opposing position is essentially that advanced by Berkeley in the *New Theory of Vision*, though I shall call its proponent 'B' so as to allow him to deviate a little from the historical Berkeley. B denies that there is a single concept *square*, which may or may not be possessed by the blind man, and whose possession is tested by whether he generalizes when he regains his sight. The sighted adult's use of the word 'square' rests upon two separable and conceptually unconnected abilities.²⁰ Both of the concepts apply to arrangements of simultaneously existing objects, but nevertheless they are distinct.

No one can say that B's position is, on the face of it, a wildly attractive one; to see why it has attracted so many adherents, we must look at the reason B can advance for opposing V's simple idea of the single conceptual capacity. Many reasons have been advanced, but the one I want to focus upon as being particularly important arises from the way B regards visual experience. B supposes that a subject could enjoy visual experience without regarding it as *of* a world distinct from himself at all. In this condition, the subject would not conceive of the items of which he was aware as *outside* himself, or as located at any distance from himself, but even so, B thinks, his visual experience would acquaint the subject with a two-dimensional mosaic of colours. Now, B supposes that it would be possible for the subject to abstract from this experience colour concepts, such as 'red', 'blue', etc., and B supposes that the subject might respond in a similar way to shape-resemblances in the colour mosaics, and thereby form (two-dimensional) shape concepts. Such are the concepts which B terms *visible square*,

¹⁹ The qualification 'seems to apply' is designed to take account of the possibility that either representation might be illusory. [Evans was dissatisfied with this paragraph and probably would have rewritten it. Ed.]

²⁰ The qualification 'conceptually unconnected' is required so as to preclude a trivialization of B's position; *tangible square* and *visible square* must not be analysable as *square and tangible*, and as *square and visible* respectively.

visible circle—abstracted from an experience which has no reference to an external reality at all. These concepts, B thinks, are the concepts which V is prepared to ascribe to the blind man. But this is what B cannot understand. How is it possible for a man who cannot form the least idea of visual experience to acquire a concept capable of being abstracted so directly from it—related to it in the same way in which colour concepts are related to it?

These reflections certainly do not constitute a decisive argument against V's position, but they do focus attention upon an unease which will not be removed until visual experience and its spatiality are fundamentally rethought.

I have located the main reason for B's position here, rather than in *anti-abstractionism* where it is sometimes located, for two reasons. First, Berkeley's only sound point concerns a spurious explanation (in terms of images) of our generalizing capacities; it cannot be used to place a limit upon the range of those capacities themselves. If Berkeley were to claim that the supposed concept of a square which V attributes to the blind man cannot exist, because, being free of sensory elements from either modality, it could not consist in any image, V should reply by citing Berkeley's own observation that the concept of a visible triangle cannot consist in an image either. Secondly, the anti-abstractionism considerations, even if they were valid, are considerations of the wrong kind to sustain B's position. They rule out V's concepts simply because they are a-modal (i.e. supra-sensible or neutral with respect to sense modality), and this would mean that B's objection to giving an affirmative answer to Molyneux's Question is no stronger than an objection to giving a similar answer about the man born deaf who has to generalize the concept *pulsating*. I want to give B a position which would enable him to discriminate between the two cases.²¹

If I have identified in the dispute between B and V the central issue raised by Molyneux's Question, then we can see

²¹ Berkeley himself was clear on this. See the *New Theory of Vision*, sect. 127: 'It having been shown that there are no abstract ideas of figure . . . the question now remaining is, whether the particular extensions, figures, and motions perceived by sight be of the same kind with the particular extensions, figures, and motions perceived by touch?'

that this issue was missed by two contemporary discussions of the question.

In a recent book on Molyneux's Question, the psychologist M. J. Morgan writes as though the main issue was one of *innateness*. He writes, for example:

Locke replied 'Not' to Molyneux's Question to avoid postulating a common representational scheme for the different senses, because such a schema implies an innate supra-sensible structure to the mind.²²

The reasoning comes out in the following passage:

The real question was whether he could name what he saw. And the answer to this was 'Not' because to name the visual impression he would have to compare it to some other idea, an idea common between touch and vision. Plainly, such an idea could not be a simple 'sensory impression'; it would have to be something transcending individual impressions, and to which these individual impressions could be referred. The *Essay* was written against such innate ideas . . .²³

It seems to me that the issue of innateness cuts right across the dispute between B and V. In the first place, I see no particular reason why the concept which V attributes to the blind man has to be innate. Whether it can be accounted for by empiricist learning mechanisms depends upon what we conceive those mechanisms to be. I suppose that, according to the most radical empiricist position, an organism has an innate similarity-space defined over *sensations*, and concepts simply result from a partitioning of that space.²⁴ It is perfectly true, on that position, that no account can be given of the spatial concepts which V attributes to the blind man. This is not because the required concept is supra-sensible (i.e. a-modal). There is nothing on the radical empiricist view that precludes sensations produced by the stimulation of different sense modalities being sufficiently close together in the innate similarity space for responses conditioned to the

²² Op. cit., p. 14.

²³ *Ibid.*, p. 7. In fairness to Morgan, it should be pointed out that he ends the book with the view that the nativist/empiricist distinction is not a very useful one in terms of which to think of approaches to Molyneux's Question.

²⁴ For the notion of 'innate similarity-space' see, e.g., W. V. O. Quine, *Roots of Reference* (La Salle, Ill.: Open Court, 1973), sect. 5, 'Some implicit standard . . . for ordering our episodes as more or less similar must therefore antedate all learning, and be innate.' See also Quine's 'Linguistics and Philosophy', in S. Hook, *Language and Philosophy* (New York: New York University Press, 1969), pp. 95-8.

one to generalize to the other. There is nothing particularly upsetting to an empiricist theory of concept formation in the suggestion that human subjects who are trained with the use of 'harmonious' in the case of sounds might generalize its use (without further training) to the case of certain combinations of colours. Rather, the difficulty arises in the case of shape concepts because on the simple model there is no way for the gap between succession and simultaneity to be bridged. However, such a radical empiricism is not an attractive position, nor is it clear that B could appeal to it, since he too acknowledges that the concept *tangible square* is a simultaneous concept.

A much more reasonable position would be to suppose that spatial concepts can be learned through the subject's capacity to *perceive* spatially. A perceptual spatial representation, although it embodies spatial information, is not a *conceptual* representation, and there is room for the explanation of our acquisition of *concepts* like *square*, with their characteristic generality, in terms of our exposure, in perception, to a range of squares.²⁵ It is not clear why the concept which V attributes to the blind man, and which is open to generalization, could not have been acquired in this way. Of course, it is B's thesis that no concept acquired in this way can be directly applied on the basis of vision, but that, V holds, is because B fails to take into account what is involved in tactual-kinaesthetic spatial perception, and has a false view about the nature of visual experience. Their dispute is about these things, not about innate ideas.²⁶

In the second place, far from innateness being an essential ingredient of V's position, it is an element of a perfectly possible variant of B's position. B holds that there is no *conceptual* connection between *tangible square* and *visible square*, but this leaves it open how the move from the visible

²⁵ For further explanation of the distinction between conceptual and perceptual representations, see the Appendix to this paper. [This Appendix must be one that Evans planned to write, for as far as I know it does not exist. See instead *The Varieties of Reference* (Oxford: Clarendon Press, 1982), chap. 5, pp. 122-9. Ed.]

²⁶ Nor is the innateness of the blind man's capacity to perceive spatially the ground of the dispute between them. On my construal of the debate, both B and V attribute to the blind man this capacity, hence arguments, or evidence, for the innateness of this capacity affects both of them equally.

to the tangible is made. Berkeley held that we *learn* the connection by experience, but an alternative hypothesis is that the connection is pre-programmed into the brain. (I assume that it is a coherent hypothesis with respect to anything which is learned that it should have been innate.) It follows from this observation that Molyneux's Question is not in fact a crucial experiment in the dispute between B and V, for although a negative answer would refute V, a positive answer would be consistent both with V's position, and with this nativist version of B's position.

In a second recent work on Molyneux's Question, Judith Jarvis Thompson takes the central issue to be whether a world is conceivable in which tactual circles give rise to the visual impression of a square, and tactual squares give rise to the visual impression of a circle. She writes:

I am inclined to think that what was in Molyneux's mind, because of which he drew that conclusion [that the blind man would not generalize] from those premises [the blind man has not yet attained the experience that what affects his touch so or so, must affect his sight so and so], was this: that what affects one's touch so or so *could* have affected one's sight such and such instead of so and so . . .²⁷

Borrowing an idea from Grice,²⁸ she attempts to show that no such world is conceivable. For example, when three tactual circles are brought together, empty space can be felt between them, but there is nothing corresponding to this in the case of three visual squares. I shall not go into detail here, because this line of thought seems to me to be beside the point. For what position would this reflection support on the debate between B and V? If Mrs Thompson's reasoning is cogent, and if the blind man can be expected to rehearse it, it follows that he can *work out* which visual shape corresponds to which tactual shape, *on the assumption that he knows that some such correspondence exists*. But, according to V, no such assumption is necessary. V holds that, on seeing the square or the four points arranged in a square, the blind man is confronted with an instance of his antecedently existing concept, and hence should be disposed to apply it

²⁷ 'Molyneux's Question', *Journal of Philosophy* 71 (1974), p. 637.

²⁸ H. P. Grice, 'Some Remarks about the Senses', in R. J. Butler (ed.), *Analytical Philosophy* (Oxford: Blackwell, 1962).

without any additional information or instruction at all. There is no question of his having to *work out* how to apply it. We don't suppose that the man who regained his hearing would need to be told that one of the sounds he was to hear was going to be continuous and the other pulsating in order to be able to apply his concept in the new case, and this is the position V takes on Molyneux's Question.

Although she does not mention any of this work, Mrs Thompson's paper is the last in a long line, beginning with Leibniz, of answers to Molyneux's Question. Leibniz held that the blind man would be able to work out which was which, on the ground:

that in the globe there are no points distinguishable on the side of the globe itself, all being level there without angles, whereas in the cube there are eight points distinguished from all the others.²⁹

Hence, Leibniz thought that if one was to set up a correspondence between the visual and the tactual, one could not map the visual circle on to the tactual square in such a way as to generate analogues in the visual world to each structural property and relation perceivable in the tactual world. Mrs Thompson and Leibniz focus on different examples of properties that would not be representable in the visual domain in the unnatural correspondence, but their arguments have the same structure. However, Leibniz realized the limitations of this argument, in a way which Mrs Thompson apparently does not. Leibniz wrote:

Perhaps Molyneux and the author of the *Essay* are not so far from my opinion as at first appears . . . If you will weigh my answer, ['Yes' to Molyneux's Question] you will find that I have put a condition upon it which can be considered as included in the question: it is, that the only thing in question is that of distinguishing, and the blind man knows that the two figured bodies which he must distinguish are there, and thus that each of the appearances which he sees is that of the cube or that of the globe.³⁰

In other words, Leibniz was refusing to support V's position, since by implication, he suggests that in the absence of instruction, the newly sighted man might enjoy visual experiences of squares and circles without being put in mind

²⁹ *New Essays on the Human Understanding*, IX, 8.

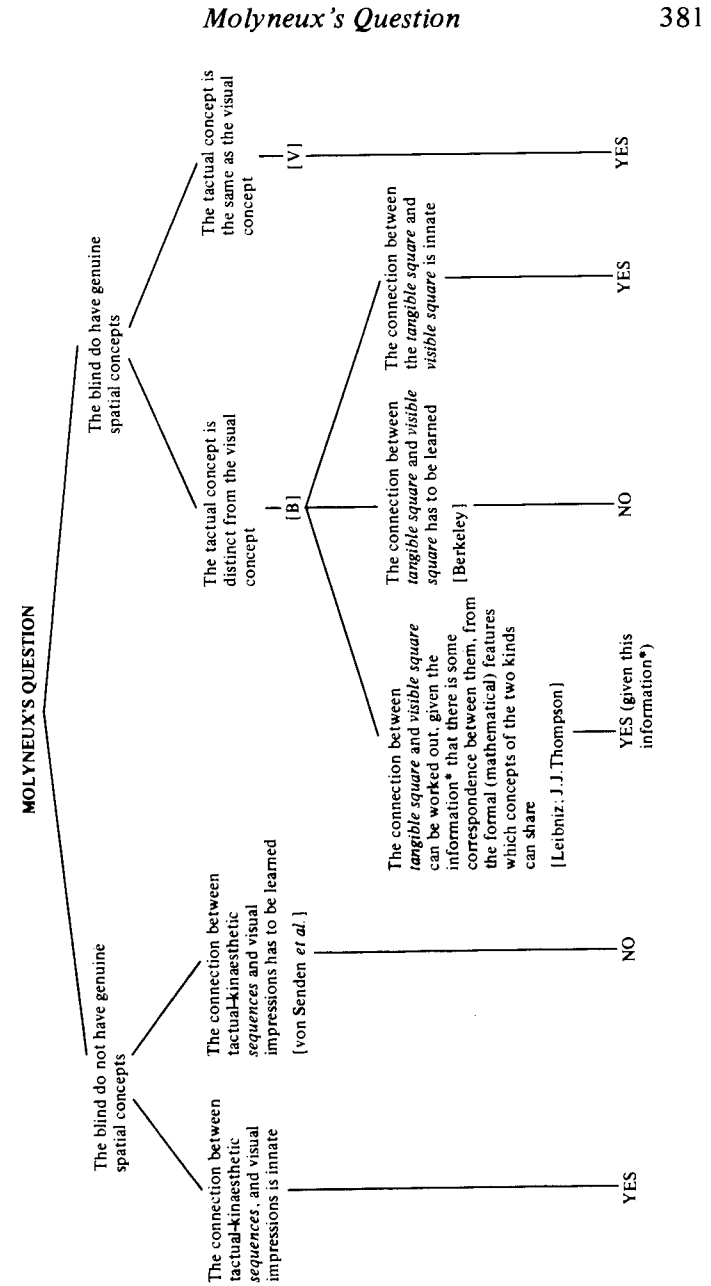
³⁰ *Ibid.*

of the shapes discernible by touch at all. Hence, Leibniz's qualified affirmative answer shows him to be an adherent of a version of B's position. After all, Berkeley himself was prepared to allow that 'the visible square is fitter than the visible circle to represent the tangible square'³¹ on the ground that the visible square had, as the visible circle did not, several distinguishable parts, and it would not matter to the fundamental disagreement that he has with V that the visible square is *uniquely* fitted to represent the tangible square. It remains the case that the one *represents* the other, rather than being both instances of a common concept. It remains the case, that is, that there is an intelligible and separable conceptual capacity whose range is restricted to the set of tactually perceived squares.

As we have seen, the result of Molyneux's experiment bears somewhat indirectly upon what I have identified as the main issue raised by his question. (See the diagram opposite.) It is true that a negative result to the experiment does undermine V's position, and it might be thought that enough negative answers had been collected to make further speculation unnecessary. However, almost all of the experiments which are cited as providing a negative result to Molyneux's Question do no such thing, for while it is true that subjects cannot name the circle and the square, this is because they do not have any visual figure perception at all but are restricted to a confusing succession of experiences of light and colour. Even Berkeley missed this point, and cites Chiselden's case in defence of his position; quite unsuitably for even as Berkeley reported it, the subject in that case 'knew not the shape of anything, nor any one thing from another, however different in shape or magnitude'.³² Von Senden cites only one or two cases in which it is claimed that the subject could distinguish the circle from the square, yet could not name them correctly, but even these results are not guaranteed to be relevant to the central issue, since a capacity to make gross 'same/different' judgements is far from establishing visual perception of the figure. A more recent case studied by Gregory seems to confirm a positive answer to Molyneux's

³¹ Berkeley, *New Theory of Vision*, sect. 142.

³² *The Theory of Vision Vindicated and Explained*, sect. 71.



Question, but unfortunately the experiments were only carried out a considerable time after the operation.³³ I have been able to track down only two reports of relevant experiments since then, and neither bears unequivocally upon the issue.³⁴ In view of the delicacy of the experimental conditions, it might be thought to be worth while to make a more theoretical examination of the issue by considering what background theories of perception might sustain the respective positions of B and V, and whether any arguments might be offered for or against them.

II

The next section of our path lies across a minefield. We cannot get closer to the dispute between B and V without getting at least a little clearer about what is involved in a blind subject's (forming a) simultaneous perceptual representation of his vicinity, and to do this, we must work within the framework of some theory of perception and perceptual experience. Fortunately our task is not the elaboration and defence of such a theory, and there will be little in the very general account that I shall put forward with which either party to the dispute is committed to disagree, so that once the minefield is crossed, we can continue our journey on safer terrain.

I want to begin by considering the spatial element in auditory perception. What is involved in a subject's hearing a sound as coming from such-and-such a position in space? I assume that the apparent direction of a sound is a *phenomenal* property of the perception of the sound; just as we hear sounds as high or low, discordant or harmonious, we hear them as coming from this or that direction. I have already mentioned one very important, negative, point: it is not sufficient for an organism to perceive the direction of

³³ R. L. Gregory and J. Wallace, 'Recovery from Early Blindness—a case study', in R. L. Gregory, *Concepts and Mechanisms of Perception* (London: Duckworth, 1974), pp. 65–129.

³⁴ A. Valvo, *Sight Restoration after Long Term Blindness* (New York: American Foundation for the Blind, 1971); H. Umezū, S. Torii, and Y. Uemura, 'Post-operative Formation of Visual Perception in the Early Blind', *Psychologia* 18 (1975), pp. 171–86.

a sound that it be capable of discriminating, i.e. responding differentially to, sounds which have different directions. '[A]n organism could perfectly well discriminate between values on all the proximal variables that specify position in the third dimension, and yet have no awareness of position in the third dimension *per se*.'³⁵ When we envisage such an organism, we envisage one that can be conditioned to respond differentially to those different values of the proximal stimulus which code the direction of sound, e.g. by pressing a button, yet in whom the difference in stimulus is not connected to any difference in spatial *behaviour*. 'Awareness of position in the third dimension' at least involves such a connection. George Pitcher makes this point effectively in his recent book. He considers the suggestion that 'position in the auditory field' is a phenomenal property without any intrinsic connection with behaviour, and writes:

... suppose the idea is legitimate; suppose, that is, that the direction from which a (phenomenal) sound comes is a purely auditory matter, that it is just another aspect of the perceiver's auditory sense-datum. If so, then it ought to be logically possible for someone to hear the direction from which a certain single sound (a bird-call, for example) is coming, and yet for him not to know in what direction he must point (or walk) if he is to point (or walk) in the direction from which the sound is coming, not to know in what direction he must look if he is to look in that direction, . . . and so on for all related abilities.³⁶

We do not hear a sound as coming from a certain direction, and then have to *think* or *calculate* which way to turn our heads to look for the source of the sound etc. If this were so, then it should be possible for two people to hear the sound as coming from the same place ('having the same position in the auditory field') and yet be disposed to do quite different things in reacting to the sound. Since this does not appear to make sense, we must say that having the perceptual information at least partly consists in being disposed to do various things (always given, of course, other

³⁵ T. G. R. Bower, 'Infant Perception of the Third Dimension and Object Concept Development', in L. Cohen and P. Salapatek (eds.), *Infant Perception* (New York: Academic Press, 1975), p. 34. Bower credits the point to Irving Rock.

³⁶ G. Pitcher, *A Theory of Perception* (Princeton, N.J.: Princeton University Press, 1971), p. 189.

beliefs and desires which make that reaction to the sound source a sensible thing to do).

The same point also comes out very clearly if we reflect upon how we might specify the spatial information which we imagine the perception to embody. The subject hears the sound as coming from such-and-such a position, but how is that position to be specified? We envisage specifications like this: he hears the sound *up*, or *down*, *to the right* or *to the left*, *in front* or *behind*, or *over there*. It is clear that these terms are *egocentric* terms; they involve the specification of the position of the sound in relation to the observer's own body. But these egocentric terms derive their meaning from their (complicated) connections with the actions of the subject. Some people—including, apparently, Freud—are only able to understand the word 'right' via the rule linking it to the hand they write with, and I suppose a similar defect might force someone to rely upon the connection between 'down' and the earth's surface, though such a person should not travel into space. But when these terms are understood in this way, they are not suitable for specifying the content of the information embodied in directional perception. No one hears a sound *as coming from the same side as the hand he writes with* in the sense, that, having heard it thus, he has to say to himself 'Now I write with this hand' (wiggling his right hand) 'so that sound must be coming from over there' (pointing with his right hand). Rather, he can immediately say to himself 'It's coming from over there' (pointing with what is in fact his right hand), and may then reflect as an afterthought 'and that's the hand I write with'. Thus Charles Taylor writes:

Our perceptual field has an orientational structure, a foreground and a background, an up and down . . . Now this orientational structure marks our field as essentially that of an embodied agent. It is not just that the field's perspective centres on where I am bodily—this by itself doesn't show that I am essentially agent. But take the up-down directionality of the field. What is it based on? Up and down are not simply related to my body; up is not just where my head is and down where my feet are. For I can be lying down, or bending over, or upside down; and in all these cases 'up' in my field is not the direction of my head. Nor are up and down defined by certain paradigm objects in the field, such as earth or sky: the earth can slope for instance . . .

Rather, up and down are related to how one would move and act in the field.³⁷

Auditory input, or rather, the complex property of auditory input which codes the direction of sound, acquires a spatial *content* for an organism by being linked with behavioural output in an advantageous way. But, at least in the case of adult human beings, the connection is very complex, for the behaviour which is appropriate to a sound at such and such a position is, when described in muscular terms, indefinitely various. This is not merely because the behaviour may involve the movement of different parts of the body; one can run, walk, crawl, or, as in the case of rats in a famous experiment, swim, to a target position. Even if we focus upon a particular kind of behaviour, such as reaching out with the hand for a rattle heard in the dark, there is a similar kind of complexity, since an indefinite range of reaching responses (identified in muscular terms) will be appropriate, depending upon the starting position of the limb, and the route it follows—which need not, and often cannot, be the most direct. It may well be that the input-output connections can only be finitely stated if the output is described in explicitly spatial terms—for example, 'extending the arm', 'walking forward two feet', etc.—and if this is so, there would certainly not be the reduction, claimed by Poincaré, of this egocentric spatial vocabulary to a muscular vocabulary. Such a reduction is not needed for the point being urged here, which is that the spatial information embodied in auditory perception is only specifiable in a vocabulary whose terms derive their meaning from being linked with bodily actions. Even given an irreducibility, it would remain the case that possession of such information is directly manifestable in uncalculated behaviour; there would just be indefinitely many ways in which the manifestation can be made.³⁸

I have used the term 'egocentric' in what is close to its literal meaning, and without intending to link my views with

³⁷ C. Taylor, 'The Validity of Transcendental Arguments', *Proceedings of the Aristotelian Society* (1979), p. 154.

³⁸ Egocentric spatial terms, and spatial descriptions of bodily movement would then form a structure familiar to philosophers under the title 'holistic'. For a study of concepts interrelated in this way, see C. A. B. Peacocke, *Holistic Explanation* (Oxford: Clarendon Press, 1979).

any others expressed with the use of the term. I shall occasionally allow myself the shorthand way of speaking in terms of information 'specifying a position in behavioural space', but in doing so I shall not be talking about information about a special kind of space, but about a special kind of information about space—information of the type one would possess if one had mastered an egocentric spatial vocabulary, and had received and understood information expressed in it. It is perfectly consistent with the *sense* that I have assigned to this vocabulary that its terms should *refer* to points in a public three-dimensional space.³⁹

Although we have told part of the story of what is involved in a subject's having an auditory perception embodying spatial information, we have not told the whole of it. The dispute between B and V concerns the conceptual response of a thinking subject to his perceptual experience, and so the perceptual representations which the debate forces us to consider are states of conscious subjects. When we were thinking of the (simultaneous) perceptual representation which a blind man might form of his immediate environment, we were thinking, however inchoately, of something the possession of which would constitute a perceptual *experience*, i.e. a state of consciousness. But, however addicted we may be to thinking of the links between auditory input and behavioural output in information-processing terms—in terms of computing the solution to simultaneous equations⁴⁰—it seems abundantly clear that evolution could throw up an organism in which such advantageous links were established long before it had provided us with a conscious

³⁹ Since coining the term 'behavioural space', I have found it used in roughly the same sense by C. B. Trevarthen; see 'Two Mechanisms of Perception in Primates', in *Psychologische Forschung* 31 (1968), p. 302: 'Animals act as though they were continuously cognizant of a space for behaviour around the body . . . In this world acts are made from the body as centre and origin. Therefore the spatial frame for activity has a symmetry imposed upon it; it is bisymmetric with the midplane of the body and polarized in the antero-posterior direction of the body axis. I shall call this body-centred space *behavioural space*.'

⁴⁰ For the mechanism of auditory localization, see, e.g., P. H. Lindsay and D. A. Norman, *Human Information Processing* (New York: Academic Press, 1977), pp. 178–88. For the writings of an addict, see J. Fodor, *The Language of Thought* (Brighton, Sussex: Harvester Press, 1976), pp. 42–53. One of the disadvantages of the addiction is that it tends to blur the distinction I am trying to explain.

subject of experience. If this point is not immediately obvious, it can be brought out by reflection on the following possibility. A conscious adult may display fairly normal responses to stimuli (including directional responses to spatially varying stimuli) and yet might have no conscious experience, might sincerely deny that he perceives anything at all. A dramatic illustration is provided by the case of a brain-damaged patient studied by Weiskrantz, who was able to point to a source of light despite claiming that he could not see anything at all.⁴¹

Reflecting upon this kind of case, philosophers and psychologists have thought that what is required for the application of our intuitive concept of conscious experience is that the subject be able to ascribe the experience to himself: to say or think 'I am having such and such an experience'. It is understandable, if one looks at matters in this way, that one should find the concept of little interest; surely, one might think, the experience can antedate thoughts about it. But, though it is true that our intuitive concept requires a subject of experience to have *thoughts*, it is not thoughts about the experience that matter, but thoughts about the world. In other words, we arrive at conscious perceptual experience only when sensory input is not only connected to behavioural dispositions in the way outlined—perhaps in some phylogenetically more ancient part of the brain—but also serves as the input to a thinking, concept-applying, and reasoning system, so that the subject's thoughts, plans and deliberations are also systematically dependent upon the informational properties of the input. (Psychologists sometimes refer to the link I have in mind as 'perceptual-verbal encoding'.) When there is such a link, we can say that he, the person, rather than just some part of his brain, receives and possesses the information. Of course, these thoughts are not epiphenomena; what a conscious subject does critically depends upon his thoughts, and so there must be links between the system I mention and behaviour. After all, it is only these links that enable us to ascribe any (conceptual)

⁴¹ L. Weiskrantz, E. K. Warrington, M. D. Saunders and J. Marshall, 'Visual Capacity in the Hemianopic Field following a Restricted Occipital Ablation', *Brain* 97 (1974), pp. 709–28. 'But always he was at a loss for words to describe any conscious perception, and repeatedly stressed that he saw nothing at all in the sense of "seeing", and that he was merely guessing' (p. 721).

content to those thoughts. Further the intelligibility of the system I have described depends upon there being a *harmony* between the thoughts and the behaviour to which a given sensory state gives rise, but this will seem adventitious only to those who forget that concepts exercised in thought are learned by an organism in whom the links between sensory input and behaviour have already been established. Now, I do not mean to suggest that only those information-bearing aspects of the sensory input for which the subject has concepts can figure in a report of his experience. It is not necessary, for example, that the subject possess the egocentric *concepts* 'to the right' etc., for him to have the experience of a sound as being from the right. All I am requiring for conscious experience is that the subject have *some* concepts, some thoughts, and that the content of those thoughts should systematically depend upon the informational properties of the input.

So, to return to our original question, we can say that a subject perceives the directionality of sound if *he* is in an informational state whose content is reportable (not necessarily by him) in egocentric spatial terms. And we must understand *the subject's* possession of the information in terms not only of his unreflective behaviour's systematic dependence upon the sensory input, but also, necessarily, of his thought's systematic dependence on that input too. We have not yet built in, or required, that the subject should be able to hear sounds from different positions simultaneously, but even in the absence of that requirement, we have, in this informational state, a 'simultaneous' spatial representation. For the subject hears a sound from one among indefinitely many simultaneously existing positions which define behavioural space. However, it is easy to understand what is involved in the subject's having a simultaneous representation in the stronger sense, of simultaneously hearing two sounds coming from different positions in space. He would then be in a complex informational state the content of which entails the egocentric location of two distinct sounds.

Let us bring these reflections to bear upon the 'simultaneous' perceptual representations of space, which earlier we argued the blind man could possess. The spatial information

available to him upon the basis of tactual-kinaesthetic perception is much *richer* than that available by hearing, and quite different perceptible phenomena, of course, are specified as located at different positions. However, there is a fundamental point of similarity; when we think of the spatial content of tactual-kinaesthetic perception, we also think of it as specifiable in egocentric terms. Indeed, when he uses his hand, the blind man gains information whose content is partly determined by the disposition he has thereby exercised—for example, that *if* he moves his hand forward such-and-such a distance and to the right he will encounter the top part of a chair. And when we think of a blind man synthesizing the information he receives by a sequence of haptic perceptions of a chair into a unitary representation, we can think of him ending the process by being in a complex informational state which embodies information concerning the egocentric location of each of the parts of the chair; the top *over there*, to the right (here, he is inclined to point or reach out), the back running from *there* to *here*, and so on. Each bit of the information is directly manifestable in his behaviour, and is equally and immediately influential upon his thoughts. One, but not the only, manifestation of this latter state of affairs is the subject's judging that there is a chair-shaped object in front of him.

We started off by thinking about what is involved in perceptions which specify the egocentric position of a stimulus, and we find that we have captured perceptions which convey, at least in a rudimentary way, *shape* or *figure*—i.e. perceptions upon the basis of which shape concepts could be applied. To make the transition, it is necessary to move from auditory spatial perception, which specifies the direction but not the distance of a sound, to a mode of perception which also conveys information about the distance of perceptible phenomena from the subject, so that he can think of a series of points as lying on a plane, and as all being equidistant from him. But no new theoretical departure is made; the content of this information is still specifiable egocentrically—we are still dealing with behavioural space.

It is a consequence of the fact that the spatial content of auditory and tactual-kinaesthetic perceptions must be

specified in the same, egocentric, terms, that perceptions from both systems will be used to build up a unitary picture of the world, and hence that spatial concepts applicable upon the basis of one mode of perception must generalize to the other. There is only one behavioural space.⁴²

III

There is nothing in the description that I have offered of the spatiality of tactual-kinaesthetic perception with which either party need disagree. Berkeley himself emphasized the egocentricity of the spatial information provided by the tactual-kinaesthetic mode of perception. For example, in the *New Theory of Vision* he writes:

. . . by touch he [the blind man] could not perceive any motion but what was up or down, to the right or left, nearer or farther from him; besides these and their several varieties or complications, it is impossible he should have any idea of motion.⁴³

And it becomes clear that one of Berkeley's main arguments for his negative answer to Molyneux's Question is precisely that the spatiality of vision has nothing to do with the egocentrically specifiable spatial information. In the same section, he continues:

He [the blind man] would not therefore think anything to be motion, or give the name motion to any idea which he could not range under some or other of those particular kinds thereof. But . . . it is plain that by the mere act of vision he could not know motion upwards or downwards, to the right or left, or in any other possible direction.

An angel, or unembodied spirit, who had no sense of touch, and who could not act, would have no notions of *up*, *down*, *left*, *right*, *forwards*, and *backwards*, but, says Berkeley, it could see perfectly well: 'i.e. having a clear perception of the proper and immediate objects of sight'.⁴⁴

We come here to a theoretical disagreement about the nature of visual perception which might be seen to underlie

⁴² See S. J. Freedman and J. H. Rekosh, 'The Functional Integrity of Spatial Behaviour', in S. J. Freedman (ed.), *The Neuropsychology of Spatially Oriented Behaviour* (Homewood, Ill.: Dorsey, 1968).

⁴³ *New Theory of Vision*, sect. 137.

⁴⁴ *Ibid.*, sect. 95.

the dispute between B and V. For, suppose one took a view of the nature of visual perception radically different from the one expressed by Berkeley, by holding that the spatiality of vision is not a primitive datum, but something to be explained in the way we have explained the spatiality of auditory and tactual-kinaesthetic perception. To hold this is to hold that any visual experience of distinct but spatially related phenomena must consist in the subject's possession of spatial information specifiable in egocentric terms. This would then provide a common basis for the application of at least certain fundamental spatial concepts. To explore this possibility further, I want to concentrate upon a streamlined version of Molyneux's Question.

It has been known for many years that direct electrical stimulation of the visual cortex in human subjects produces the experience of a flash of light, or a 'phosphene'.⁴⁵ It is also known that phosphenes can be produced in this way in subjects who have been blind for many years,⁴⁶ though, to the best of my knowledge, no attempt has ever been made to produce phosphenes by direct cortical stimulation of the congenitally blind.⁴⁷ The repeated stimulation of a given site of the visual cortex with a given intensity reliably produces a phosphene located at the same position in the visual field, and the simultaneous stimulation of two or more distinct sites produces the experience of simple patterns. However, as with the after-images, the phosphenes are experienced as moving when the eye moves. There has been a certain amount of research devoted to the possibility of exploiting this fact to provide a visual prosthesis for the blind.⁴⁸

⁴⁵ For a review, see G. S. Brindley 'Sensory Effects of Electrical Stimulation of the Visual and Paraviscual Cortex', in R. Jung (ed.), *Visual Centres of the Brain*, Handbook of Sensory Physiology 7, 3, b (Berlin: Springer-Verlag, 1973), chap. 26.

⁴⁶ W. H. Dobbelle, M. G. Mladejovsky, and J. P. Girvan, 'Artificial Vision for the Blind', *Science* 183 (1974), pp. 440-4.

⁴⁷ The closest we come to a study of phosphenes in the congenitally blind is in the report of W. Schodtman ('Ein Beitrag zur lehre von der Optischen Lokalisation bei Blindgeborenen', *Archiv für Ophthalmologie* 54 (1902), pp. 256-67) who claimed that 'pressure phosphenes' could be produced in such subjects by pressure on their eyes, and that these phosphenes were located in behavioural space; on their eyes, and that these phosphenes were located in behavioural space; located 'up' if the pressure was on the lower part of the eyeball, and 'down' if on the upper. However, there is a certain amount of doubt about whether this claim can be believed. I owe the reference to I. Rock.

⁴⁸ For a recent review, see W. H. Dobbelle, 'Current State of Research on

G. S. Brindley, the pioneer of this work, has made long-term cortical implants in two patients, and used them to produce recognizable patterns corresponding to letters of the alphabet, including a fairly good question mark.⁴⁹

The simplified version of Molyneux's Question that I want to consider is this. Suppose it is possible to produce in a congenitally blind man, by the use of a Brindley implant, a pattern of phosphenes exemplifying a shape which previously he had been able to name when he tactually perceived it (for example, a square, or the letter 'A'), would he then be able to name the shape correctly? This version of Molyneux's Question avoids certain difficulties in the original question which arise from the complexity of the information which a newly sighted man would receive upon opening his eyes, from the confusion introduced by the movement of his eyes and head, and from the need to separate figure and ground. But the essence of the question is preserved. And I am suggesting that one way in which V could defend his expectation that the blind man will be able to generalize is by arguing as follows. To have the visual experience of four points of light arranged in a square amounts to no more than being in a complex informational state which embodies information about the egocentric location of those lights; for example, one is perceived up and off to the left, another below it, a third up and off to the right, and so on. Now, we are assuming that the subject has been able to form simultaneous perceptual representations of the locations of tactually perceived objects, and this means that he has been in a complex informational state of just this kind before, perhaps when he felt the four corners of a wire square to be occupying these, or similarly related, positions in behavioural space. Of course, the perceptible phenomena apparently located at the various positions in behavioural space in the two cases are different, but the spatial ingredient of the information would be

Providing Sight to the Blind by Electrical Stimulation of the Brain', *Journal of Impairment and Blindness* 71 (1977), pp. 290-7. See also T. D. Sperling (ed.), *Visual Prosthesis* (New York: Academic Press, 1971).

⁴⁹ G. S. Brindley and W. S. Lewin, 'The Sensations Produced by Electrical Stimulation of the Visual Cortex', *Journal of Physiology* 196 (1968), pp. 479-93.

specifiable in the same vocabulary, so that if receipt of such information was sufficient to prompt application of the concept *square* in the tactual case, it is not clear why it should not do so in the visual case.

There is a complication. Although phosphenes are assigned positions 'in the visual field', the position does not involve the specification of distance from the observer; in this very pared-down case, visual localization is like auditory localization:

Sperling: Could the woman equate those phosphenes with any prior visual experience?

Brindley: Yes, certainly. She said that they were like stars in the sky. This raises the question of whether they appear to be distant or close to her, but when I tried to probe her on that, and when other people questioned her on that, she was not very consistent. I do not think that she has a definite impression that they are a long way away, or that they are close.⁵⁰

If the blind man was to apply the two-dimensional concept *square*, he would have to think of the points of light as equally far away. Although this is a natural assumption for sighted people when looking at the night sky, since none of the normal distance cues, like occlusion, is present, we cannot say that this would be a natural assumption for the blind man, who has never responded to those cues. So, strictly speaking, a defence of V's position based upon the view of the spatiality of vision being outlined, carries with it the need to enter a slight qualification to his affirmative answer to Molyneux's Question. However, I do not think that it significantly diminishes the interest of his conclusion.

V's position is therefore essentially that advanced by the great Scottish philosopher, Thomas Reid:

To set this matter in another light, let us distinguish betwixt the *position* of objects with regard to the eye, and their *distance* from it. Objects that lie in the same right line drawn from the centre of the eye, have the same position, however different their distances from the eye may be: but objects which lie in different right lines drawn from the eye's centre, have a different position; and this difference of position is greater or less, in proportion to the angle made at the eye by the right lines mentioned. Having thus defined what we mean by the position of objects with regard to the eye, it is evident, that as the real

⁵⁰ Telephone conversation with G. S. Brindley, in *Visual Prosthesis*, op. cit., p. 48.

figure of a body consists in the situation of its several parts with regard to one another, so its visible figure consists in the position of its several parts with regard to the eye; and as he that hath a distinct conception of the situation of the parts of the body with regard to one another, must have a distinct conception of its real figure; so he that conceives distinctly the position of its several parts with regard to the eye, must have a distinct conception of its visible figure.⁵¹

Visual localization is complicated in the normal case by the fact that the eyes can move, so that, even given a single orientation of the head, there is no simple correspondence between points on the two retinas and points in behavioural space. It is not necessary, however, for V to argue that the mechanism whereby account is taken of eye position in computing position in behavioural space must be present at the inception of visual experience. His position is that apparent location in behavioural space is an essential feature of any visual experience which permits the application of two-dimensional spatial concepts, not that the apparent location is accurate, or likely to be accurate, were the visual cortex to be connected to the retina in the standard way.

Nor is it necessary for V to argue that the capacity to apply visual shape concepts rests upon no more than the capacity for visual localization. It obviously does not. I have already mentioned the separation of figure and ground, which is closely tied to the capacity to perceive an object as the same as it moves about in the visual field. Furthermore, sighted people clearly have a capacity to respond to the purely visual similarities of things, as when they recognize a friend by his face, or detect a family resemblance in the faces of father and son, and the application of many visual shape concepts depends upon this capacity. V's position is that if a visual system is capable of providing the experience of distinct but spatially related phenomena then it is *at least* a system which provides the subject with information about the position such phenomena occupy in behavioural space, and that this fact provides a basis for the application of certain very fundamental spatial concepts which is common to vision and touch. What matters

⁵¹ Thomas Reid, *An Inquiry into the Human Mind*, ed. T. Duggan (Chicago, Ill.: University of Chicago Press, 1970), p. 113. I presume Reid must mean to be referring to 'the Cyclopean eye'.

to V is that the application of concepts like *straight*, *square*, *between* etc., should be independent of the capacity to respond to the characteristic look that things have, and not that the application of all shape concepts should be independent of that capacity.

There is certainly nothing in the literature on direct cortical stimulation that contradicts V's contention. What experimenters mean by 'point in the visual field' is 'apparent point in behavioural space'; when they map cells on to points in the visual field, the subject is asked to point to the apparent source of the light. This, however, may not impress B, since all those studied are late-blind; but it is extremely difficult to see how the spatiality of the experience could be established experimentally in the absence of any links with behavioural space. There *is* a certain amount of evidence that visual localization, and visual shape perception are subserved by different parts of the brain in certain mammals, and that with suitable lesions, they can be separated. Schneider claims that 'undercutting the superior colliculus [in a hamster] abolishes the ability to orient toward an object, but not the ability to identify it according to tests of pattern discrimination learning'.⁵² However, I do not think that this undermines V's position, which concerns the nature of conscious visual experience. It is not difficult to envisage feature detectors of the Hubel and Wiesel type, capable of responding selectively to certain pattern and shape features of the retinal stimulus, operating in the absence of any of the links between stimulus and behaviour which we earlier saw were necessary if the stimulus is to embody localization information. These feature detectors could then be exploited in discriminatory behaviour. But such a system could not provide the basis for the visual experience of shape. When we have the experience of seeing a square, we do not unaccountably find ourselves inclined to judge that there is a square somewhere in the vicinity; we possess information about each of the parts of the square and their relations to one another. The experience of seeing a square, we might say, is a *complex*

⁵² G. S. Schneider 'Two Visual Systems', in *Science* 163 (1969), p. 901. A similar but by no means identical distinction between 'ambient' and 'focal' perception has been suggested by C. Trevarthen, in the paper cited earlier.

psychological state. And this, V argues, is because it involves the possession of information about the location of each of the parts of the square in behavioural space. To this claim, Schneider's findings are plainly irrelevant.⁵³

So far, I have considered the position on the spatiality of vision which might sustain V's answer to Molyneux's Question to be an empirical theory about the visual system in humans. I want to end by canvassing the possibility that the position can be defended on *conceptual* grounds—the possibility that there are grounds for holding that it does not *make sense* to talk of a subject's perceiving spatially distinct points of light in an extended visual field, unless this can be explained in terms of the subject's receiving information about the location of phenomena in behavioural space. Now, I think sighted people *do* have an incapacity to conceive (i.e. imagine) a visual experience of an array of lights which are not at the same time referred to points in behavioural space, but I do not think that we can rest any weight on this incapacity, which B will argue shows nothing more than our extreme familiarity with the association of points in visual and behavioural space. However, there are weightier considerations.

The first argument is very familiar, and I shall only mention it briefly. B's notion of a visual field generates what we may call *necessarily private facts*. B presumably thinks of the subject's visual field as having an orientation—four distinguishable sides—so that the experience of A is distinguishable from the experience of \forall . (Were this not so, it would not make sense to speak of the motion of something across, or the rotation of something in, the visual field.) The visual field, then, has four sides, *a*, *b*, *c*, *d*, which can be identified from occasion to occasion, and what makes the

⁵³ The considerations of this paragraph strongly suggest that a great deal of the work on computer vision—computer simulation of visual perception—is based upon a mistake. The problem is conceived to be that of simulating the human subject's capacity to describe his environment, and so, fundamentally a problem of devising a sufficiently complicated pattern-recognizing program. Whatever it is that one may reach by this route, it does not remotely resemble the psychology of a conscious human subject, since there is nothing which corresponds to the human's non-conceptual representation of his environment—i.e. nothing which corresponds to visual experience. If the argument of this paper is along anything like the right lines, that defect will not be rectified until attention shifts to the study of programs for computers which control behaviour in a spatial world.

experience of A different from that of \forall is that in the first case the apex of the A is closest to the *a*-side, and in the second it is closest to the *c*-side. Now, on B's theory, the sides of the visual field cannot be distinguished by reference to anything outside the field, and consequently, in identifying them from occasion to occasion, the subject is engaged in the application of a necessarily private concept, something which I believe Wittgenstein has shown to be highly problematic. For there does not appear to be a distinction between the correct and the incorrect application of the names 'a', 'b', 'c', and 'd'. As I said, this argument is very familiar, and since the present application of the argument introduces nothing new, I shall not say anything further about it.

One who uses this argument of Wittgenstein's plays along with B's notion of the visual field, only to discredit it, so to speak, from the inside. The second argument I want to mention questions the legitimacy of B's describing the visual experience in spatial terms at all. We may start by taking note of the fact that many of the spatial descriptions of visual experience which we are inclined to give are obviously metaphorical. For example, I myself have been speaking of cortical stimulation 'producing a pattern of phosphenes arranged in a square', and the literature is replete with such metaphorical talk. It is clear that such descriptions cannot be taken literally; there are not literally four points of light, or indeed four things of any kind, arranged in a square. To think that when a subject seems to see four points of light arranged in a square, there really are four (mental) items *actually* arranged in a square is to commit the sense-datum fallacy. It might be better to call it 'the homunculus fallacy', to which it inescapably gives rise. One commits the homunculus fallacy when one attempts to explain what is involved in a subject's being related to objects in the external world by appealing to the existence of an inner situation which recapitulates the essential features of the original situation to be explained—by introducing a relation between the subject and inner objects of essentially the same kind as the relation existing between the subject and outer objects. Thus, we start by wondering what is involved in a subject's gaining knowledge of the spatial relations of outer objects, and we appeal,

quite correctly, to an inner, psychological, state—a perceptual experience 'of items disposed in the visual field'. We cannot then go on to take the 'visual field' description literally, by supposing that there are certain items which *in fact* stand in spatial relations. For the question arises again: how are we to understand the subject's capacity to gain knowledge of these relations?⁵⁴

We must therefore always be prepared to replace our metaphorical descriptions of experience in terms of mental items—colour patches, phosphenes, and the like—with conceptually more innocent descriptions. V's way of thinking about the spatiality of visual experience enables him to do this; 'the subject experiences four phosphenes arranged in a square' describes a subject as being in a complex informational state embodying the (non-conceptual) information (or misinformation) that there are four lights located at various positions in his immediate vicinity. Notice: the subject *has* this information, he does not confront it on an inner screen.

How can B cash the metaphor of the four lights in the visual field? He can certainly do nothing along V's lines. It is essential to V's way of avoiding the inner screen that he thinks of a visual experience as an informational, or representational, state—a state which can be assessed as true or false, and hence which refers to something outside itself. The spatiality of the experience is explained in terms of its embodying information about the spatial relations of things. But on B's view, points in the blind man's visual field bear no relation to points in physical space, and this precludes him from conceiving of visual experience as representational. If the existence and spatial relations of objects in the subject's immediate vicinity are not represented, it is difficult to believe that the existence and spatial relations of any other group of objects can be.

B wishes to hold that cortical stimulation of a congenitally blind man may cause him to have an experience which can be described in spatial terms, even though he does not perceive points of light as having positions in behavioural space.

⁵⁴ The link between the sense-datum fallacy and the homunculus fallacy is brought out well in Ryle's *Concept of Mind* (Harmondsworth: Penguin, 1963), pp. 200–11.

When we were content with metaphorical descriptions, there did not seem to be any difficulty—the simulation produced four phosphenes arranged in a square. But once we attempt to dispense with the metaphor, it is not clear that the description of the experience in spatial terms can be defended. If 'arranged in a square' cannot come in literally, as a description of the position of phosphenes, and if it cannot come in indirectly, as a description of the apparent position of lights in space, it is not clear that it can come in at all. It certainly cannot come in by virtue of the fact that it is the description that the subject himself is inclined to offer—at least not if B is to continue to give a negative answer to Molyneux's Question.

Much more needs to be said about both of the arguments against B's position which I have mentioned. For example, a full treatment of the subject would involve an extended discussion of the results of the inverting prism experiments.⁵⁵ I shall be content if I have shown how Molyneux's Question is linked to these other fundamental issues in the philosophy of mind and perception, and if I have shown that V is not entirely without resources to defend himself.

⁵⁵ For such a discussion, see I. Rock, *The Nature of Perceptual Adaptation* (New York: Basic Books, 1966).

Appendix I

Commentary on Jerry A. Fodor's 'Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology'

1 Much of what Fodor writes gives the impression that he has put his finger on the essential difference between two fundamentally opposed positions in psychology ('this long-standing dispute'). But the two positions which Fodor identifies, 'Rational Psychology', with its commitment to 'The Computational Picture of Mind' (CPM), and 'Naturalistic Psychology' (NP), are not only consistent, but complementary—as Fodor admits when the issue is explicitly raised ('none of the arguments *for* a rational psychology is, in and of itself, an argument *against* a Natural Psychology'.¹ See also what Fodor says is 'in principle the right way to look at things'²). The CPM does not officially concern itself with the semantic or representational properties of internal states, but it must *presuppose* that the formal operations it studies are defined over states which have content, since it makes sense to identify a process as a computation or a proof only if it is *truth-preserving*. So the CPM is necessarily an incomplete, or partial, picture of mind: it is 'debarred from raising questions about the semantic properties of mental representations',³ but it is committed to their existence. It must be completed by a theory of *content*, and Fodor nowhere dissents from the suggestion advanced by NP that it is 'organism/environment relations which (presumably) fix these [semantic] properties'.⁴ The 'bland and ecumenical conclusion' that NP and CPM are consistent and complementary is the right one; the dim prospects for a scientific

From *The Behavioural and Brain Sciences* 3, 1 (1980), pp. 79–80. Reprinted by permission.

¹ J. A. Fodor, 'Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology', reprinted in his book *Representations* (Brighton, Sussex: Harvester Press, 1981), p. 244.

² *Ibid.*, p. 233.

³ *Ibid.*

⁴ *Ibid.*, p. 244.

NP which Fodor then goes on to outline leave this conclusion untouched, and are as much to be lamented by a subscriber to the CPM as by anyone else.

If this is correct, it is not merely the tenor of several parts of the paper which is misleading. Fodor argues that the CPM commits one to 'methodological solipsism', to the view that 'how the world is makes no difference to one's mental states'.⁵ If this were so, the CPM *would* be inconsistent with the central tenet of NP, but I cannot see that it is so. Opposition to methodological solipsism is based upon an incapacity to understand how content can be assigned to the internal states of an organism in the absence of those organism/environment relations we naturally presume to fix it. A computational psychologist does not need to make any assumptions about the nature of an organism's environment, precisely because he treats its internal states like the uninterpreted formulae of a formal calculus, but for the same reason, this hardly constitutes a *commitment* to methodological solipsism. True, Fodor does not regard the CPM as wholly silent about ascriptions of content to internal states, for only certain such ascriptions are consistent with the 'formality condition', but I am equally unclear how methodological solipsism is supposed to follow from this.⁶

Let us go along with Fodor and suppose that no term for a representational state which meets the formality condition can be such as to entail that the representation is true—that such terms must be like those in 'S believes that p', and 'it seems to S as though p', rather than those in 'S knows that p' and 'S perceives that p'. (But see below.) It still does not follow from the fact that a psychological life is describable by means of statements no one of which, taken individually, entails a proposition about the organism's environment, that such a life could be enjoyed in the absence of a suitable

⁵ *Ibid.*, p. 228.

⁶ See *ibid.*, pp. 226–7 for Fodor's discussion of this condition. He writes: 'I take it that computational processes are both *symbolic* and *formal*. They are symbolic because they are defined over representations, and they are formal because they apply to representations in virtue of (roughly) the *syntax* of the representations . . . What makes syntactic operations a species of formal operations is that being syntactic is a way of *not* being semantic. Formal operations are the ones that are specified without reference to such semantic properties of representations as, for example, truth, reference, and meaning.'

environment. For we come back to the possibility, not so far ruled out, that what makes it the case that an internal state has the content that *p*, is (for example, and crudely) that it should be of a kind typically produced by the state of affairs that *p*, and typically giving rise, in the context of other suitable beliefs and desires of the agent, to behaviour appropriate to the state of affairs that *p*.

For example, although the statement 'it seems to S as though there is food to the right, in front of him' entails nothing about S's current environment, it could not be true if S did not eat, and if S's movement in the world, forward or backward, to the left or to the right, were not systematically dependent upon information from its environment. It would appear from his comments upon Winograd's work that Fodor would deny this, for content is there ascribed in the total absence of suitable organism/environment relations; but this aspect of Fodor's discussion of Winograd's work⁷ does not seem to be consistent with his subsequent claim that 'the machine doesn't know what it's talking about'.⁸

2 Along with several other philosophers, I hold that there are thoughts we have about particular objects which we simply could not have if those objects did not exist. For example, an internal state of a subject can be ascribed the content that *this table is round* only if there is a particular object it is about, and on which it is causally dependent. There is no neutral or 'existence-independent' specification of this content to which we can retreat if the subject is hallucinating. If there is no object (whether or not the subject believes there is), there is no content—no thought—and this means that if the ideal description of mental life enjoined by the formality condition is one which carries absolutely no commitments about the nature of the organism's environment, then it is an unattainable ideal.

I see nothing in Fodor's arguments in defence of the formality condition to undermine the arguments that have led

⁷ T. Winograd, *Procedures as a Representation for Data in a Computer Program for Understanding Natural Language* (Cambridge, Mass.: MIT Project MAC, 1971).

⁸ Fodor, *op. cit.*, p. 233.

philosophers to embrace the view I have outlined. The grain of truth upon which Fodor bases his defence of the formality condition is this: there cannot be two, *simultaneously existing*, internal states which differ in content without differing formally (i.e. physically), and this is because there cannot be two internal states which differ 'in causal role' (i.e. functionally) without differing formally (i.e. physically). But when we attempt to use the formality condition to rule out specifications of mental states which are existentially committing, we are not comparing two simultaneously existing internal states, but a single internal state in two different possible situations, and I see no reason why there should not be difference in content without difference in form when the states compared exist in different possible situations. For an internal state of a subject which is actually caused by the presence of an object in his environment can be functionally quite different from a formally indistinguishable state of the same subject in a counterfactual situation from which we abstract the object, so that the subject is hallucinating.

3 NP is identified more by slogans ('psychology is a branch of biology') than by any precise theses, but even so, I find it hard to understand why Fodor saddles it with a denial of the formality condition or with a restriction to transparent statements of propositional attitude. Elsewhere Fodor identifies the main idea of NP as that of making a 'science of the organism/environment relations which . . . fix these [semantic] properties', and it is against this possibility that his closing argument is directed. Essentially the objection is that 'the theory which characterizes the objects of thought is the theory of *everything*',⁹ so that a statement of what a word refers to, or of what a subject is thinking, attends the progress of special sciences like physics and chemistry. This seems incorrect. We can say, quite definitely and with no help from physics or chemistry, that, on the planet Yon, the word 'water' refers to *the stuff which is found in seas, lakes, and rivers, and which falls as rain on Yon*.¹⁰ To this, Fodor will object that it does not employ a nomologically appropriate description of XYZ. Actually, I think that Fodor is quite

⁹ *Ibid.*, p. 248.

¹⁰ *Ibid.*, p. 245.

wrong about the vocabulary appropriate to laws. He says that laws about salt must be stated using 'NaCl' rather than 'salt', but 'salt' is a perfectly projectible predicate, and if Kripke and Putnam are right about its semantics (as Fodor assumes they are) it is necessarily coextensive with 'NaCl'. But in any case, the whole issue of whether the descriptions of objects of thought or reference are nomologically appropriate seems to me to be irrelevant. A scientific NP will issue in laws of the form: $(x)(y)(z)(x\text{'s utterance of } y \text{ refers to } z \text{ iff } R(x, y, z))$, which requires only that R be 'nomological'. Statements of the object of the thought of a particular subject will not themselves be laws, or lawlike, but the *instantiations* of a law, and this carries no requirement upon suitable vocabulary. Fodor does employ the notion 'description under which . . . is law-instantiating', but I cannot imagine what he means by it.¹¹ The proper name 'Robin Roberts' will serve perfectly well in an instantiation of the law schematized above. I confess to not seeing what the fuss in the last pages is all about.

¹¹ *Ibid.*, p. 249.

Appendix II

A Further Reflection on Semantic Structure and Logical Form

Perhaps you will let me convey one thought about the subject of my paper which I have had since writing it. I was gripped too much by one point: my conviction—which still appears to me to be correct—that Davidson did not adequately warrant distinguishing inferences containing the logical constants from inferences which are supposed to depend upon the analysis of primitives. So I fashioned a very austere and absolute notion of structurally valid inference, which I then distinguished from all inferences which depended upon the particular assignments given to particular semantical primitives. Although this distinction is acceptable as it stands, I would now look at it in a rather different way. The second class is much too heterogeneous. It includes 'bachelor'-'male' inferences, as well as e.g. the logical inferences. But these ought to be distinguished, since the latter are clearly *formal* in a sense in which the former are not. I think we need to introduce the idea of inferences which are categorially valid *relative to the particular assignments made to certain chosen expressions*. Then we should distinguish inferences which at least at a certain point depend merely upon the category (or *kind* of assignment) of certain elements occurring in them (which include the logical inferences) from those which do not so depend. And then rather than placing my 'structurally valid' inferences at the centre of the stage, we see them as the limiting case of categorially dependent inferences—those which do not depend at any point upon the particular assignments made to lexical items. Putting the matter this way loses me none of my central points; the notion of categorially dependent inference plainly depends upon the

kind of interpretational semantics which I wanted to see. (The inference from A to B is categorially valid iff there is some *proper* subset S of the expressions occurring in A, B, such that, restricting attention to those interpretations which agree with the designated interpretation solely in what they assign to elements of S, upon any interpretation on which A is true, so is B.) And Davidson's claim that somehow a theory of *truth* captured the interesting distinctions between inferences is still seen to be wrong.

Sometimes S will be restricted to just one item—e.g. 'Someone' in the inference from 'John runs' to 'Someone runs'. But often we are interested in holding constant the interpretations of a whole set. Obviously, logical constants traditionally conceived provide us with the best examples of categorially dependent inferences, but they are certainly not unique in doing so, though it is probably the case that most of the categorially dependent inferences of non-logical words can only be revealed when we hold constant not only the interpretation of those words, but also [the interpretation] of the logical words. (This provides a way of identifying the logical constants.)

So I want to replace the picture of the bifurcation between two different kinds of inference, with the picture of a spectrum of diminishing categorial dependence, which has at the one end, my pure 'structurally valid' inferences, and at the other inferences (if there are any) with no categorial dependence at all. An analogy: Think of a logical fruit machine with a row of windows behind which rotate wheels. On each wheel there are only members of the same semantic category, though the same categorial wheel (with the lexical items in the same order) may be found behind several windows. The wheels are spun, and a bell sounds when a valid argument is found in the windows—i.e. when the sentences to the left entail the last sentence on the right. When the bell sounds, how many 'Hold' buttons do you have to press for you to keep getting valid arguments (keep the bell ringing) when the wheels are spun?—remembering only that if in the initial argument two or more wheels of the same category were involved, and showed the same lexical item in two or more windows, then those

wheels will be locked together so that, though different words are shown in those windows, those windows show the same words.¹

¹ It would appear that Evans has not written what he meant here. An amendment intended to replace the last sixteen words might read: 'so that, when the wheels are spun, those windows always display the same word as one another, even though which word it is they all show may vary over time'.

Index

- Altham, J. E. J., 117 n., 155
Ananyev, B. G., 284 n.
Anderson, S., 232 n.
Anscombe, G. E. M., 98 n., 319 n.
Ayers, M. R., 247 n., 317 n., 372 n.
- Bach, E., 92 n.
Bain, A., 367
Baldwin, T., 87 n.
Beer, E. de, 365 n.
Bennett, J., 255 n., 272 n., 286
Berkeley, G., 270-1, 274, 365 n.,
366, 370-2, 374-5, 378, 380, 390-1
Berlin, L., 276, 278
Bever, T. G., 333 n.
Binnick, R. L. 224 n.
Black, M., 298 n.
Blackburn, S., 87 n.
Bouveresse, J., 182 n., 291 n.
Bower, T. G. R., 368 n., 383 n.
Bresnan, J., 216 n., 224 n.
Brindley, G. S., 391 n., 392-3
Broadbent, D. E., 333 n.
Bure, T., 125 n., 187 n.
- Cargile, J., 72
Cassirer, E., 364
Castañeda, H. N., 98, 312
Chambers, F. K., 13
Chastain, C., 131 n.
Chomsky, N., 216 n., 220 n., 224 n.,
231-3, 241-2, 284 n., 327
Cohen, L., 116 n., 383 n.
Crossley, J., 210 n.
Culicover, P. W., 107 n., 110 n.
- Dahl, O., 134 n.
Davidson, D., 1 n., 16 n., 17 n., 25 n.,
29, 46 n., 49-50, 52, 53 n., 54-7,
59, 60, 62, 66, 68 n., 70 n., 72-3,
92 n., 95 n., 100 n., 102 n., 113 n.,
147 n., 153 n., 174 n., 224 n., 295 n.,
300 n., 317 n., 325-6, 343, 345,
346 n., 355 n., 359
Davies, J. W., 4 n.
Davies, M. K., 192 n., 210 n., 213 n.
- Delhaise, E., 10 n.
Descartes, R., 318
Diderot, D., 365
Dobelle, W. H., 391 n.
Donnellan, K. S., 17 n., 193-9, 206,
209, 211
Duggan, T., 394 n.
Dummett, M. A. E., 53 n., 83 n.,
86 n., 95 n., 125 n., 148 n., 155,
178, 180, 202 n., 275-6, 291 n.,
293, 294 n., 295, 297 n., 300 n.,
307 n., 326, 345 n.
- Eissfeldt, O., 19 n.
Evans, G., 85 n., 186 n., 191 n., 214 n.,
295 n., 335 n., 346 n., 364 n., 373 n.,
374 n., 377 n.
- Field, H., 33 n.
Fine, K., 358 n.
Fodor, J. A., 79 n., 224 n., 333 n.,
386 n., 400-3
Foster, L., 33 n.
Freedman, S. J., 390 n.
Frege, G., 74, 81, 83, 86, 180, 226,
233, 291, 303, 306-14, 316-18,
326, 345, 346, 350
Freud, S., 384
- Geach, P. T., 30 n., 76-9, 84 n.,
87-8, 91-5, 97-8, 100-3, 112,
114-18, 120-8, 132 n., 133 n.,
135 n., 136, 153, 155-63, 165-74,
175 n., 218, 223 n., 224 n., 298 n.,
326, 344 n., 348, 349
- Gellner, E., 4 n.
Gibson, J. J., 368, 369 n.
Girvan, J. P., 391 n.
Gödel, K., 15, 67 n.
Green, G., 224 n.
Gregory, R. L., 380, 382 n.
Grice, H. P., 16 n., 18 n., 202, 378
Guenther, F., 295 n.
Guenther-Reutter, M., 295 n.
Gunderson, K., 131 n.
Guttenplan, S., 94 n., 98 n., 125 n.,
319 n.

- Halle, M., 216 n., 224 n.
 Hamburger, H., 107 n.
 Hankamer, J., 134 n.
 Harman, G. H., 1 n., 5 n., 17 n., 68 n., 92 n., 95 n., 100 n., 102 n., 113, 123, 141 n., 142 n., 143 n., 147 n., 174 n., 203 n., 300 n., 326 n., 344 n., 355 n.
 Harms, R. T., 92 n.
 Hartlage, L. C., 370 n.
 Hermes, J., 298 n.
 Hintikka, J., 16 n., 46 n., 57 n., 153 n., 317 n.
 Hobbes, T., 249
 Hockney, D. J., 4 n.
 Holtzman, S., 322 n.
 Hook, S., 376 n.
 Humberstone, L., 210 n., 213 n.
- Jackendoff, R. S., 103 n., 224 n., 232
 Jacobs, R. A. 141 n.
 Jones, B., 370 n.
 Jung, R., 391 n.
 Juurmaa, J., 370 n.
- Kambartel, F., 298 n.
 Kant, I., 318, 369
 Kaplan, D., 16 n., 292, 303 n., 307 n., 311 n., 317 n., 321 n.
 Karttunen, L., 118 n., 224 n.
 Katz, J., 79 n.
 Kaulbach, F., 298 n.
 Keenan, E. L., 84 n., 97 n., 117 n., 147 n., 155 n., 171 n.
 Kiparsky, P., 232 n.
 Klima, E., 79 n., 107, 110-11, 219 n.
 Kripke, S. A., 1-3, 6, 7 n., 8, 10-13, 15, 18, 22, 24, 85 n., 100 n., 102 n., 104 n., 131 n., 178-9, 181-2, 183 n., 186 n., 192-3, 198-9, 208, 211, 300 n., 404.
 Kuno, S., 98 n.
- Ladefoged, P., 333 n.
 Lakoff, G., 92 n.
 Lashley, K., 370 n.
 Lasnik, H., 214 n., 216-17, 229-30, 232-41, 243, 245-7, 248 n.
 Leblanc, H. 53 n., 100 n.
 Leibniz, G. W. von., 370, 372, 379, 380
 Leich, C., 322 n.
 Levin, M., 194 n.
- Lewin, W. S., 392 n.
 Lewis, C. I., 251 n.
 Lewis, D. K., 18 n., 68 n., 71 n., 155, 175 n., 355 n., 360, 362
 Leyden, W. von., 365 n.
 Lindsay, P. H., 386 n.
 Lindstrom, P., 67
 Locke, D., 258
 Locke, J., 273, 274 n., 364, 365 n., 370, 376
 Long, P., 299 n., 300 n.
 Lotze, H., 281, 366
- McArthur, R. P., 344 n.
 McCawley, J. D., 141 n., 147 n., 174 n.
 McDowell, J. H., 20 n., 25 n., 49 n., 85 n., 184 n., 186 n., 191 n., 291 n., 295 n., 296 n., 346 n.
 Mackie, J. L., 272 n., 365 n.
 Marcus, R. B., 85
 Margain, H., 25 n.
 Marshall, J., 387 n.
 Massey, G. J., 344 n.
 Mates, B., 83 n., 190 n., 349 n.
 Mill, J. S., 270-1, 282, 366 n., 370-1
 Millar, S., 370 n.
 Miller, G. A., 216 n., 224 n.
 Mladejovsky, M. G., 391 n.
 Molyneux, W., 364, 365
 Montague, R., 64, 65 n., 69 n., 156 n.
 Moore, G. E., 319 n.
 Morgan, M. J., 224 n., 365 n., 376
 Mostowski, A., 67
- Nagel, T., 337 n.
 Norman, D. A., 386 n.
- Parret, H., 182 n., 291 n.
 Partee, B. H., 95 n., 97, 98 n., 107 n., 156 n., 236 n., 361 n.
 Peacocke, C. A. B., 188 n., 360 n., 385 n.
 Pecheux, P., 370 n.
 Perry, J., 291 n., 292, 300, 303, 312-21
 Pitcher, G., 66 n., 383
 Plantinga, A., 194 n.
 Platner, E., 281, 366
 Poincaré, H., 282-3, 370-1, 385
 Prior, A., 155, 343, 344 n., 348, 349 n., 350 n., 353-5, 357, 358 n., 360-1, 363
 Putnam, H., 404

- Quine, W. V. O., 26-7, 30, 33, 36-7, 40, 43, 46-8, 68, 76, 87 n., 88, 127 n., 136 n., 153 n., 159, 162 n., 167 n., 325, 326 n., 376 n.
 Quinton, A. M., 272 n., 298 n.
- Reid, T., 271, 274 n., 393, 394 n.
 Reinhart, T., 110 n.
 Rekosh, J. H., 390 n.
 Rescher, N., 29 n., 57 n., 155
 Revesz, G., 282, 370 n.
 Robin, M., 370 n.
 Robinson, H. W., 11 n.
 Rock, I., 383 n., 391 n., 399 n.
 Rosenbaum, P. S., 141 n.
 Rosencranz, D., 370 n.
 Russell, B., 4 n., 158, 178, 183, 186, 187, 191 n., 201, 202 n., 296, 299-300, 326
 Ryle, G., 73 n., 92 n., 398 n.
- Sag, I., 134 n.
 Salapatek, P., 383 n.
 Saunders, M. D., 387 n.
 Schiffer, S. R., 18 n., 25 n., 194 n., 198-9
 Schneider, G. S., 395 n., 396
 Schock, R., 186 n.
 Schodtman, W., 391 n.
 Searle, J. R., 4 n.
 Senden, M. von., 291, 366, 369, 370 n., 371-2, 380
 Shemyakin, F. N., 284 n.
 Shoemaker, S., 319 n.
 Sosa, E., 4 n.
 Sperling, T. D., 392 n., 393
 Stalnaker, R., 190 n.
 Straaten, Zak van, 249 n.
 Strawson, P. F., 2, 66 n., 179 n., 201, 202 n., 249-54, 256-8, 260-1, 264, 266-8, 279-81, 285-7, 289, 290, 338 n.
- Stroud, B., 25 n.
 Suslick, R., 370 n.
 Swanson, J. W., 33 n.
- Tarski, A., 35 n., 52, 81-3, 86, 235 n.
 Taylor, B., 25 n., 146 n.
 Taylor, C., 384, 385 n.
 Taylor, I., 11
 Tennant, N. W., 117 n., 155
 Thomason, R., 190 n.
 Thompson, J. J., 378, 379
 Torii, S., 382 n.
 Trevarthen, C. B., 386 n., 395 n.
- Uemura, Y., 382 n.
 Umezū, H., 382 n.
- Valvo, A., 382
 Vendler, Zeno, 64 n.
 Villey, P., 283, 369
- Wallace, J., 87 n., 99 n., 100 n., 155, 295 n., 382 n.
 Warnock, G., 370 n.
 Warrington, E. K., 387 n.
 Weiskrantz, L., 387
 Wexler, K., 107 n.
 Wheeler, S. C., 56 n.
 White, R., 299 n., 300 n.
 Wiggins, D., 180, 181
 Wilson, W. K., 4 n.
 Winograd, T., 402 n.
 Wittgenstein, L., 5, 272-3, 281, 318, 319 n., 322 n., 324, 337, 341, 345 n., 397
 Wood, O. P., 66 n.
 Wright, C., 322-6, 333-6, 339, 340-1
- Zahl, P. A., 370 n.