

THOUGHT, REFERENCE, and EXPERIENCE

THEMES FROM THE PHILOSOPHY OF GARETH EVANS

> EDITED BY JOSÉ LUIS BERMÚDEZ

THOUGHT, REFERENCE, AND EXPERIENCE

This page intentionally left blank

Thought, Reference, and Experience

THEMES FROM THE PHILOSOPHY OF GARETH EVANS

Edited by

JOSÉ LUIS BERMÚDEZ

CLARENDON PRESS • OXFORD

OXFORD UNIVERSITY PRESS

UNIVERSITY PRESS

Great Clarendon Street, Oxford 0x2 6DP

Oxford University Press is a department of the University of Oxford. It furthers the University's objective of excellence in research, scholarship, and education by publishing worldwide in

Oxford New York

Auckland Cape Town Dar es Salaam Hong Kong Karachi Kuala Lumpur Madrid Melbourne Mexico City Nairobi New Delhi Shanghai Taipei Toronto

With offices in

Argentina Austria Brazil Chile Czech Republic France Greece Guatemala Hungary Italy Japan Poland Portugal Singapore South Korea Switzerland Thailand Turkey Ukraine Vietnam

Oxford is a registered trade mark of Oxford University Press in the UK and in certain other countries

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

> > © the several contributors, 2005

The moral rights of the authors have been asserted Database right Oxford University Press (maker)

First published 2005

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, without the prior permission in writing of Oxford University Press, or as expressly permitted by law, or under terms agreed with the appropriate reprographics rights organization. Enquiries concerning reproduction outside the scope of the above should be sent to the Rights Department, Oxford University Press, at the address above

You must not circulate this book in any other binding or cover and you must impose the same condition on any acquirer

> British Library Cataloguing in Publication Data Data available

Library of Congress Cataloging in Publication Data Data available

Typeset by Newgen Imaging Systems (P) Ltd., Chennai, India Printed in Great Britain on acid-free paper by Biddles Ltd., King's Lvnn, Norfolk

ISBN 0-19-924896-6 978-0-19-924896-4

1 3 5 7 9 10 8 6 4 2

ACKNOWLEDGEMENTS

This project was begun by Dr Rick Grush of the University of California at San Diego. I am grateful to Peter Momtchiloff of Oxford University Press for asking me to take over when Dr Grush withdrew, to the contributors who were already on board for their patience with a new editor, and to the contributors who signed up subsequently for fitting into the volume so seamlessly. Preparation of the final manuscript was greatly facilitated by the research assistance of Santiago Amaya who also prepared the index. This page intentionally left blank

CONTENTS

	Contributors	ix
	Introduction José Luis Bermúdez	I
1.	Evans's Frege John McDowell	42
2.	Names in Free Logical Truth Theory R. M. SAINSBURY	66
3.	Plural Terms: Another Variety of Reference? Ian Rumfitt	84
4.	Abandoning Coreference Ken Safir	124
5.	Evans and the Sense of "I" José Luis Bermúdez	164
6.	Information Processing, Phenomenal Consciousness, and Molyneux's Question JOHN CAMPBELL	195
7.	"Another I": Representing Conscious States, Perception, and Others Christopher Peacocke	220
8.	Space and Objective Experience QUASSIM CASSAM	258

viii	Contents	
	dentity, Vagueness, and Modality . J. Lowe	290
R	References	311
Ir	ndex	321

CONTRIBUTORS

José Luis Bermúdez is Professor of Philosophy and Director of the Philosophy–Neuroscience–Psychology program at Washington University in St Louis. He is the author of *The Paradox of Self-Consciousness* (1998), *Thinking without Words* (2003), and *Philosophy of Psychology: A Contemporary Introduction* (2005).

John Campbell is Willis S. and Marion Slusser Professor of Philosophy at the University of California, Berkeley. He is the author of *Past, Space and Self* (1994), and *Reference and Consciousness* (2002).

Quassim Cassam is Professor of Philosophy at University College London. He is the author of *Self and World* (1997).

E. J. Lowe is Professor of Philosophy at the University of Durham, UK. He is author of *Kinds of Being* (1989), *Subjects of Experience* (1996), *The Possibility of Metaphysics* (1998), *A Survey of Metaphysics* (2002), and *The Four-Category Ontology* (2005).

John McDowell is a University Professor of Philosophy at the University of Pittsburgh. He prepared Gareth Evans's *The Varieties of Reference* for publication (1982). He is the author of *Mind and World* (1994, 1996), and of two collections of papers: *Mind, Value, and Reality* (1998), and *Meaning, Knowledge, and Reality* (1998).

Christopher Peacocke is Professor of Philosophy at Columbia University. His books include *Sense and Content* (1983), A Study of Concepts (1992), Being Known (1999), and The Realm of Reason (2004).

Ian Rumfitt is a University Lecturer in Philosophy at Oxford and a Fellow of University College. He has published many papers, chiefly in philosophical logic, and his book *Frege's Logical Theory* is forthcoming from Oxford University Press.

Ken Safir is Professor of Linguistics at Rutgers University. He is the author of *Syntactic Chains* (1985), *The Syntax of Anaphora* (2004), and *The Syntax of (In)Dependence* (2004).

Contributors

R. M. Sainsbury is Professor of Philosophy at the University of Texas at Austin, and Susan Stebbing Professor of Philosophy at King's College London (part-time). He is the author of *Russell* (1979), *Paradoxes* (1995), *Logical Forms* (2000), *Departing from Frege* (2003), and *Reference without Referents* (2005).

JOSÉ LUIS BERMÚDEZ

The framework for the papers in this volume is set by the wideranging philosophical contributions of Gareth Evans, who died in 1980 at the age of 34. In the papers gathered together in the posthumously published *Collected Papers* (*CP*) and in *The Varieties* of *Reference* (*VR*) (prepared for publication from drafts and lecture notes by John McDowell after Evans's death) Evans made a number of important contributions to the philosophy of language, the philosophy of thought, and philosophical logic. Some of these contributions have been extensively discussed and assimilated into ongoing debates and discussion, but considerable areas of his thought remain relatively unknown. I hope that this volume will contribute to Evans being widely recognized as one of the most original philosophers of the second half of the twentieth century.

This introduction is in two parts. Evans was a far more systematic thinker than is usual in contemporary analytical philosophy, and the first part gives an overview of his thinking. The second part of the introduction summarizes the papers and provides any necessary background.

EVANS: INFLUENCES AND OVERVIEW

Evans's philosophical work is best viewed against the backdrop of four powerful and in many ways competing currents in the philosophy of the 1960s and 1970s. Two of these currents were firmly

I am grateful to Mark Sainsbury and Christopher Peacocke for written comments on an earlier draft of this introduction.

rooted in Oxford, Evans's intellectual home throughout his career, while two originated in the very different philosophical climate of North America.

Inevitably for an Oxford philosopher of his generation, Evans was exposed to the very different philosophical concerns and styles of Michael Dummett and Peter Strawson. Dummett's Frege-inspired philosophy of language was a powerful influence throughout Evans's career, as was Strawson's neo-Kantian project of using the techniques of conceptual analysis and transcendental arguments to plot the limits and structure of our conceptual scheme.¹ At the same time Evans was very open to, and informed about, developments in philosophy on the other side of the Atlantic. As a philosopher deeply versed in philosophical logic, Evans was influenced by the powerful cluster of ideas about modality and designation that came in the wake of the semantics for quantified modal logic proposed by Saul Kripke, Ruth Marcus, and others. A final important influence on his philosophical framework was Donald Davidson's proposal to develop a theory of meaning for a regimented version of natural language in terms of a broadly Tarskian theory of truththe so-called Davidsonian program in semantics.

Evans shared Dummett's powerfully developed conviction that the best approach to the philosophy of language is broadly Fregean, incorporating a truth theory at the level of reference and a theory of understanding at the level of sense, supplemented by a theory of force that explains how language is used to perform different types of linguistic act.² Like Dummett, Evans held that the most problematic notion in the theory of meaning is the notion of sense, and also like Dummett, Evans was troubled by Frege's lack of concern with the crucial notion of what it is to grasp a sense. Evans's most systematic and substantial work, *The Varieties of Reference*, is, among other things, an essay on the relation between a theory of sense and an account of linguistic understanding. Yet the emphasis there is rather different from that of Dummett. Dummett's principal

¹ The most comprehensive statement of Dummett's philosophical outlook remains *Frege: Philosophy of Language* (1973), but the essays in *Truth and Other Enigmas* (1981) provide a more accessible introduction, as does Dummett's well-known essay "What is a theory of meaning? (II)" (1976), originally published in Evans and McDowell 1976. The best single source for Strawson is *Individuals* (1959).

² The overall picture is sketched out in Dummett 1976: 74. Dummett's most sustained exploration of the theory of force is Dummett 1973: ch. 13.

concern is with what it is to understand a sentence (to grasp the sense of a sentence), and, as is well known, this leads him away from treating truth as the central notion in a theory of sense. We cannot, Dummett argues, give a satisfactory account of what it is to grasp the truth conditions of a sentence that is not effectively decidable, and hence the notion of truth cannot be at the core of a theory of understanding in the way that it is standardly taken to be. Evans does not directly engage with this aspect of Dummett's thought (although Evans's writings suggest a robust realism that would not be sympathetic to Dummett's anti-realist conclusions). His concern is more directly with the theory of understanding at the sub-sentential level-and in particular with what it is to understand those linguistic expressions that have the job of picking out individuals (the class of referring expressions). Evans holds that an account of what it is to understand referring expressions should be formulated in the context of a more general account of what it is to think about individuals. Our language contains referring expressions because we think about objects in certain ways, and those expressions work the way they do because of how we think about the objects that they pick out. In this sense, then, he thinks that the philosophy of language should in the last analysis be answerable to the philosophy of thought.

Evans's views about the direction of explanation in our thinking about language mark a significant divergence from Dummett (noted and discussed by Dummett in his 1991b). Dummett has consistently argued that the philosophical analysis of thought can proceed only via the philosophical analysis of language. In fact, he takes this principle (which he attributes to Frege) to be the defining feature of analytical philosophy (Dummett 1994). Evans, in contrast, holds that we can elucidate the nature of thought independently of the nature of language. The idea that the direction of explanation goes from thought to language is not new to Evans. It is clearly built into the Gricean program in semantics, for example, since that program aims to elucidate linguistic meaning through speakers' communicative intentions. In contrast to Grice and his supporters, however, Evans confronts the obvious challenge that this throws up of giving a substantive and independent account of the nature of thought. Part II of VR is a sustained investigation of the different types of what Evans calls singular thoughts. Singular thoughts are, roughly, thoughts about specific and identifiable individuals (as opposed, for example, to quantificational thoughts that are not generally targeted

José Luis Bermúdez

at specific individuals). Paradigm singular thoughts, in addition to thoughts that might be expressed using proper names, are perceptual beliefs of the sort that might be communicated through demonstrative expressions such as "this" or "that" and beliefs about oneself or the present moment that might be expressed through token-reflexive expressions such as "I" or "now" (see also "Understanding demonstratives" (Evans 1981b)). Both Dummett and Evans think that thoughts are the senses of sentences, and hence that to explore what is distinctive of, say, "I"-thoughts is to explore the sense of the first person pronoun. But Evans, unlike Dummett, holds that we can investigate "I"-thoughts without proceeding via the sentences that express them.

Evans's understanding of the category of singular thoughts marks a further significant divergence from Dummett. Singular thoughts, for Evans, are Russellian. They involve ways of thinking about objects that require the existence of the object being thought about. A singular thought requires the existence of the object being thought about because thinking it depends upon the ability to pick out, or otherwise keep track of, the object in question.³ In developing a Russellian account of singular thoughts, Evans took himself to be following Frege. Despite Frege's apparent concession that sentences with empty referring terms can have a sense, and hence that there can be thoughts with non-denoting senses. Evans maintains that a proper understanding of Frege's notion of sense shows this to be impossible. The best way to understand Frege's notion of sense, Evans maintains, is through the metaphor of a mode of presentation, and, he argues, there cannot be a mode of presentation of something that does not exist. This way of understanding sentences with empty referring terms and thoughts with non-denoting senses is diametrically opposed to Dummett's own theory of sense (and indeed to Dummett's interpretation of Frege). John McDowell's contribution to this volume (Chapter 1) assesses Evans's interpretation of Frege.

Although Evans's ideas about the Russellian nature of singular thoughts are diametrically opposed to Dummett's, they have a certain affinity with well-known views of Peter Strawson's (thus bringing us to the second of the four influences in Evans's philosophical

 $^{^3}$ The case for the Russellian status of singular thoughts is made in the first three chapters of *VR* and worked out for specific classes of singular thoughts in part II of that work.

framework). In "On referring" (Strawson 1950), his well-known essay on Russell's theory of descriptions, Strawson argued that particular uses of sentences with non-referring singular terms can fail to make statements that are true or false. There is, according to Strawson, a logical relation between a sentence expressing a singular thought and a sentence asserting that the object being thought about exists, such that the first sentence can only be true or false when the second sentence is true. When this relation of presupposition fails to hold, no statement is made. Strawson's theory of presupposition is clearly situated at the level of language. Evans can be seen as extending Strawson's account to the level of thought-and, indeed, of explaining why the relation of presupposition holds. If it is indeed the case that a thinker cannot entertain a singular thought without being appropriately connected with the relevant object, then it is fully to be expected that no such thought can be expressed linguistically if the object in question does not exist.

Developing this line of thought of course requires clarifying what is to count as being appropriately connected with the object. Evans's thinking here exemplifies one of his many points of difference from Kripke and other philosophers drawing philosophical consequences from the semantics of quantified modal logic. As part of his repudiation of those theories of sense that understand the sense of a proper name in terms of definite descriptions holding true of the bearer of that name, Kripke developed a very austere account of the type of connection with an object required in order to refer to it successfully (Kripke 1972, 1980). According to Kripke, an utterance of a name refers to an object just in case there is a series of reference-preserving causal links going back from the utterance in question to the initial occasion or process when the name was bestowed. A reference-preserving link is one where a speaker intends to use the name to pick out the same object as the person from whom she learnt the name. This type of account is deeply antithetical to Evans's understanding of thought and language. In his first publication, "The causal theory of names" (1973), Evans offers a nuanced critique of Kripke's account. He maintains that the information that speakers associate with a name is indeed crucial to fixing reference, but concedes that information is individuated by causal origin (what matters is not the object that best fits the information that speakers possess, but rather the object from which that information originates). In VR, however, Evans puts more distance between himself and those theories of the content of thought loosely based on Kripke's account of names (what in VR §§3.3 and 3.4 he collectively terms the Photograph Model). The fundamental emphasis in VR is on working through the implications of the principle (Russell's Principle) that thinking about an object requires discriminating knowledge of that object (see particularly VR ch. 4). Most of the cases of discriminating knowledge that Evans discusses clearly involve some sort of causal link with the object being thought about, but what is important is the thinker's capacity to exploit that link in order to distinguish the object from all other things.

This concern with discriminating knowledge signals a further connection between Evans and Strawson. In Individuals (1959) Strawson undertakes the neo-Kantian project of mapping the structure of our conceptual scheme. For Strawson our conceptual scheme is firmly anchored in our ability to think about individual objects (what he terms *particulars*). Our ability to think about objects is inextricably tied to our ability to identify and reidentify them, and it is through our ability to identify and reidentify objects that we are able to identify and reidentify places and hence get a grip on the spatial structure of the world.⁴ The idea that there is an intimate connection between thinking about objects and thinking about space is very much to the fore in Evans. It is a prominent theme in his discussion of demonstrative identification in VR chapters 6 and 7 and is discussed with particular reference to Strawson's Individuals in "Things without the mind" (1980a): see further Cassam's contribution to this volume (Chapter 8). Thinking about an object that one is currently perceiving requires being able both to locate that object relative to oneself and to integrate one's personal, egocentric space with one's understanding of the objective layout of the environment (with what psychologists term a *cognitive map*).

In at least one respect Evans is more Kantian than Strawson. One of the principal themes of Kant's *Critique of Pure Reason* is the relation between self-consciousness and the objectivity of the world. This theme is not directly explored in Strawson's *Individuals*, but is

⁴ Strawson and Evans adopt the same three-way classification of types of discriminating knowledge (as Evans notes at VR 89 n. 2, referring to Strawson 1971). One has discriminating knowledge of an object if one is currently perceiving it; if one could recognize it were one to be presented with it; or if one knows distinguishing facts about it.

very much to the fore in Evans's discussion of self-identification in VR chapter 7. Thinking about ourselves requires discriminating knowledge no less than thought about any other type of object, and Evans analyzes the discriminating knowledge involved here in terms of the thinker's ability to locate herself within a cognitive map of the environment. These themes are discussed in Bermúdez's contribution to this volume (Chapter 5).

It is because Evans takes thinking about objects to be such a highly sophisticated cognitive ability that he allows for so many ways in which our thoughts can fail to connect to the relevant objects, and correspondingly many ways in which we can have the illusion of a thought without thinking a genuine thought. Yet, not all thoughts about objects can fail in this manner. It is only singular thoughts that run the risk of giving "the illusion of thought".⁵ This raises the question of how to delineate the class of singular thoughts. Which ways of thinking about objects are Russellian, and which not? Here Evans's analysis of thought proceeds hand in hand with his analysis of language. The category of singular thoughts is delineated through the referring expressions that appear in the sentences that express them. A singular thought is a thought whose canonical linguistic expression would involve a genuine referring expression. But what is a referring expression?

Here we see once again the influence of Kripke and possible worlds semantics. In VR 2.4 Evans addresses the question of why definite descriptions should not be counted as referring expressions. After rejecting the arguments Russell provided in "On denoting", Evans offers an argument based on the behavior of definite descriptions in modal contexts. If we treat definite descriptions as referring expressions, we will need to relativize the relation of reference to a possible world. Only thus will we be able to capture the much discussed difference between wide and narrow scope readings of definite descriptions in modal contexts (the difference between, on the one hand, saying that the object that satisfies a definite description in this possible world has a particular property F in another possible world and, on the other, saying that the object that satisfies

 5 Evans's concern in VR is almost exclusively with the "illusions of thought" that arise when one fails to be appropriately connected to the object one is thinking about. But he would no doubt have been sympathetic to the idea that thoughts can fail with respect to their predicative component—when, for example, concepts fail to be clearly specified.

José Luis Bermúdez

that same definite description in another possible world is F in that world). But no such relativization of the reference relation is necessary in the case of "standard" referring expressions. As Evans puts it, "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" (*VR* 56). In essence, Evans identifies referring expressions as a semantic kind in virtue of how they can be most illuminatingly and economically treated within a semantic theory—and his preferred semantic theory is one that treats referring expressions as rigid designators in the way that Kripke suggested.

In one sense Evans's characterization of referring expressions is rather narrow. He holds, for example, that all referring expressions are singular, thus excluding by *fiat* the possibility of plural terms counting as referring expressions-an exclusion with which Rumfitt takes issue in his contribution to this volume (Chapter 3). In another sense, however, he drastically expands the scope of referring expressions. Some of Evans's most interesting (and least well known) contributions to semantic theory come in his explorations of the semantics of pronouns in his two-part "Pronouns, quantifiers, and relative clauses" (1977) and in the more synoptic "Pronouns" (1980). Evans is particularly concerned with pronouns that have quantifier expressions as antecedents—pronouns such as the "her" in "every woman loves her mother". Bound pronouns are frequently thought to be the natural language analogs of the logician's bound variable (the "x" in " $\forall x Fx$ "). On this view, most emphatically developed by Geach (1962), there is no sense in which pronouns with quantifier antecedents can be treated as referring expressions. In opposition to this, Evans develops what he terms a coreferential treatment of bound pronouns, showing how to provide a semantics for bound pronouns that trades on the fact that any sentence with a quantifier as antecedent can be paired with a sentence (or set of sentences) that has a singular term as antecedent. Evans suggests that by evaluating sentences with quantifier antecedents in terms of appropriately paired sentences with singular term antecedents we can understand bound pronouns in referential terms, and hence as semantically similar to pronouns with singular antecedents. Just as the "her" in "Alexandra beats her donkey" corefers with "Alexandra", so too does the "her" in

"Some woman beats her donkey" corefer with the proper name in the sentence or sentence whose truth underwrites the truth of the existentially quantified sentence (in this case, of course, "Alexandra"). Some of the implications of Evans's account for linguistics are explored in Safir's essay in this volume (Chapter 4).

Evans devoted considerable attention to a type of referring expression that he termed *descriptive names* (VR §§1.7, 1.8, 2.3). Descriptive names are names whose reference is fixed by description in such a way that they will always take wide scope in modal contexts. Evans's favorite example of such a reference-fixing stipulation is

(1) Let us call whoever invented the zip "Julius"

but in fact there are a number of descriptive names in common currency. "African Eve", generally taken to refer to a specific individual who lived 200,000 years ago and is our earliest common ancestor, is a good example. Descriptive names are interesting for two reasons. First, and in opposition to much of the post-*Naming and Necessity* discussion of proper names and definite descriptions, they are both descriptive and rigid (that is, they refer to the same object in all possible worlds in which they refer at all). Second, they are not Russellian. A sentence with a non-referring descriptive name can still express a genuine thought ("African Eve did not exist" might be a true sentence). The implications for semantic theory of recognizing the category of descriptive names is discussed in Sainsbury's contribution to this volume (Chapter 2).

Insofar as descriptive names are rigid designators that clearly express a descriptive Fregean sense, recognizing the existence of descriptive names is already compromising the insistence of Kripke and other direct reference theorists that referring expressions cannot have a Fregean sense. As far as Evans is concerned, however, this is just the tip of the iceberg. Like Dummett, Evans thinks that Frege's notion of sense is an indispensable tool for understanding language, as well as being the thread that links together the analysis of language and the analysis of thought. VR is a sustained attempt to show that, despite the widely held view to the contrary, there is no incompatibility between a referring expression functioning as a rigid designator and having a Fregean sense (see VR 2.5 for an emphatic statement of this point). An important element in the case he makes is pointing out, surely correctly, that descriptive names,

José Luis Bermúdez

and definite descriptions in general, are a poor model for Frege's notion of sense.⁶ Evans's reading of Frege's notion of sense begins with a relatively neutral characterization (VR \S 1.4, 1.5). We cannot, Evans holds, think about an object unless we think about it in a particular way. The sense of a referring expression is a particular way of thinking about the object it picks out, such that anyone who is to understand a sentence featuring that expression must think about its referent in that way. It is clear from Kripke's extended discussion in Naming and Necessity that no definite description or set of definite descriptions can satisfy this requirement in the case of a proper name. But it should be equally clear that there is no reason, even in the case of a proper name, to think that the sense of a referring expression should take a descriptive form-still less so for those referring expressions that are not proper names (other than descriptive names, of course). Thinking about an object in a particular way requires having discriminating knowledge of that object. But, as pointed out earlier, thinking of an object as the unique satisfier of a definite description is only one way of possessing distinguishing information about it-and distinguishing information is only one form of discriminating knowledge.

In VR Evans explains how the discriminating knowledge requirement can be met for each of the principal categories of referring expressions. In the case of what he calls "one-off referential devices" such as demonstratives and personal pronouns, perception and recognition are the fundamental forms of discrimination. Things are more complex for proper names (VR ch. 11). Here Evans emphasizes the existence of name-using practices and distinguishes between the producers and the consumers of such practices. The producers of a practice have dealings with the bearer of a name, and their use of the name is bound up with their abilities to recognize the bearer of the name in a way that clearly satisfies the discriminating knowledge requirement. Consumers of the name, in contrast, use names with the intention of participating in specific name-using practices, with the particular practice fixed by information deriving from the bearer of the name (information that might be partially or even wholly inaccurate). Here the discriminating knowledge requirement is not directly met (as Evans notes: VR 387 n. 13).

⁶ The point has been made many times (Dummett 1973: 110–11; Bell 1979). It is astonishing that it is not more widely recognized.

One can see Evans's discussion of proper names in "The causal theory of names" (1973) and VR chapter 11 as aiming simultaneously to do justice to Kripke's modality-inspired insights, while blunting the more drastic conclusions that might be drawn from them. In "Reference and contingency" (1979) Evans applies a similar strategy to Kripke's arguments for the existence of contingent a priori truths, arguing (against Dummett and others) that there is nothing paradoxical about the existence of truths that are a priori and contingent, and (against Kripke) that there is nothing particularly interesting in the fact that such truths exist. One very interesting feature of this paper is his attempt to undercut Kripke's trademark modal arguments, all of which draw conclusions about the content of a sentence from its behavior in modal contexts.

In §III of "The causal theory of names" Evans distinguishes between the proposition a sentence expresses and its content. Propositions are understood in the standard way as functions from possible worlds to truth-values, while the notion of content is picked out by Frege's intuitive criterion for sameness of content (viz. that two sentences have the same content just if it is impossible for someone to understand both while believing one and disbelieving the other). It is widely accepted that two sentences can have different contents while expressing the same proposition (in virtue of being true in exactly the same possible worlds). Evans, however, argues that sentences can be epistemically equivalent (have the same content) while expressing different propositions, which means of course that it is illegitimate to make inferences about content on the basis of behavior in modal contexts. Consider "Julius" once again. "Julius" and "the inventor of the zip" embed differently in modal contexts, since "the inventor of the zip" can take narrow scope within modal operators in a way that "Julius" cannot. Yet, argues Evans,

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. Belief states are individuated by the evidence that gives rise to them, the expectations, behavior, and further beliefs that may be based upon them, and in all of these respects the belief states associated with the two sentences are indistinguishable. (*CP* 202)

This bold argumentative move may fail to convince. One might wonder, for example, why the two sentences "Julius is F" and

José Luis Bermúdez

"the inventor of the zip is F" should behave differently in modal contexts if they are synonymous in the way that Evans suggests. There seem to be no features of a sentence, other than its content, that could explain how a sentence behaves in modal contexts. But those who find this line of argument persuasive will still have to answer Evans's challenge to explain what difference there is between the belief that Julius is F and the belief that the inventor of the zip is F. Evans is surely correct that philosophers have played too fast and loose with modal arguments about content.⁷

Let us turn now to the fourth and final influence on Evans, the Davidsonian program in semantics. In addition to the influential collection Truth and Meaning: Essays in Semantics, which Evans edited with John McDowell in 1976, Evans engages with Davidson's project of using a Tarskian theory of truth as a theory of meaning in his papers "Semantic structure and logical form" (1976) and "Semantic theory and tacit knowledge" (1981a) and in VR §1.8. In the Introduction to Truth and Meaning Evans and McDowell strongly align themselves with one aspect of Davidson's conception of the form a theory of meaning should take. This is its divergence from what they term translational semantics. Translational semantic theories offer mappings between sentences of the object language and semantic representations of those sentences—what are, in effect, translation rules. Suppose that the meaning of object language sentence S is given by the metalanguage sentence S'. A translational semantics would give a rule stating the relation between S and S'—a rule of the form "'S' means 'S'", as one might have a French-German translation rule stating that "'Il pleut' means 'es regnet'". The key objection that Evans and McDowell make to translational semantics is that it fails to provide a theory of understanding.

What we ought to be doing is stating what the sentences of the language mean, stating something such that, if someone knew it, he would be able to speak and understand the language...But there is no escaping the fact that one could have a competence based upon the mapping relation, and yet not know what a single sentence of the language meant. A speaker-hearer would know that only if he knew what sentences of the theory's language meant; but this is knowledge of precisely the kind that was to be accounted for in the first place. (Evans and McDowell 1976: p. ix)

⁷ It is worth noting, though, that some neo-Kripkean arguments in this area are not dependent upon modal contexts in ways that make them straightforwardly susceptible to Evans's challenge. See Soames 2002.

A theory of meaning, then, must be a theory of understanding. This gives a necessary condition for a theory of meaning: no theory can qualify if a speaker could know it without understanding the language for which the theory is being given. This condition is clearly satisfied by Davidson's theory, which is designed to yield a T-sentence of the form "p" is true iff p" for every sentence of the language. Davidson avoids the translational fallacy because his recursive theory of meaning yields theorems that *use* the meaning-specifying sentence rather than *mention* it. No one could grasp a theorem such as

(2) "Madrid is the capital of Spain" is true iff Madrid is the capital of Spain

without understanding the sentence "Madrid is the capital of Spain".

Nonetheless, Evans does not follow Davidson in the claims he makes on behalf of a recursive truth theory. Evans's divergence from Davidson comes across very clearly in "Semantic structure and logical form" (1976), where he takes issue with Davidson's claim that a recursive truth theory will provide a satisfying explanation of what makes an inference valid. Davidson holds that a satisfying account of validity will have to show that valid inferences are valid in virtue of the structure of their premises and conclusion, so that formal validity is more fundamental than material validity. Evans glosses the requirement that this imposes upon a recursive truth theory as follows:

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* for that inference $S_1 ldots S_{n-1} + S_n$. 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. (*CP* 53)

Two features of a Tarskian truth theory explain how the requirement is met in individual cases. The first is that the truth theory picks out a class of expressions as logical constants. Davidson, like many others before him, thinks that we should look to the logical constants to explain what secures structurally valid inferences. Second, and slightly less obviously, structurally valid inferences depend upon premises and conclusion having common elements. These must be semantically common elements, and it is the job of a theory of truth to display the logical structure of sentences in a way that will make manifest what these semantically common elements are and how inferences can trade on them.

Attractive though this picture is (and of course this conception of what makes an inference structurally valid is very widely held), Evans finds it wanting. There are important classes of structurally valid inferences whose validity cannot, without considerable contortion, be traced to the logical structure of their constituent sentences. The non-logical parts of a sentence have a structure that can validate certain inferences, so that structural validity can, Evans thinks, be a function of semantic structure no less than of logical structure. The semantic structure of a sentence is a function of the semantic kinds to which its constituent expressions belong. These semantic kinds determine the inferential properties of expressions falling under them as a function of the type of entity that can serve as their semantic values. Consider, for example, the sentence "Clare is a skinny ballet dancer". We are interested in explaining the structural validity of the inference from "Clare is a skinny ballet dancer" to "Clare is a ballet dancer". According to Evans, this follows from a correct account of the word "skinny", which is functioning here as an attributive adjective. Attributive adjectives take as their semantic values functions from sets to subsets of those sets. In this case the function will be from the set of ballet dancers to the subset of just those ballet dancers who are skinny. Clearly, then, any skinny ballet dancer will be a ballet dancer, and the inference from "Clare is a skinny ballet dancer" to "Clare is a ballet dancer" will be valid in virtue of its semantic structure.

In stressing the need to discern semantic structure by what he calls *interpretational semantics* (VR 33; CP 61), Evans diverges from Davidson, since a Tarskian truth theory of the sort that Davidson thinks can serve as a theory of meaning does not assign semantic values to any expressions other than proper names. Evans's interpretational semantics is quite clearly Fregean in inspiration. Nonetheless, as he makes clear at VR 34, Evans is proposing interpretational semantics as a validation for a Tarski-style truth theory, rather than as a replacement for it. At the very least, he thinks, any adequate truth theory must be in harmony with a coherent interpretational semantic theory. And in fact it may be the case that one can read off from what a truth theory explicitly says about a type of expression the class of semantic values that an interpretational semantics would assign to it.

In addition to the shift to interpretational semantics, Evans proposes a further extension of Davidson's approach to meaning. Evans's guiding thought is that a theory of meaning must be a theory of understanding. One implication of this is that we must be able to explain speakers' understanding of a language (and in particular their ability to understand novel sentences of that language) in terms of their knowledge of a theory of meaning. This is not something that can be achieved merely by spelling out the form that the theory of meaning must take (presumably the conjunction of a coherent interpretational semantics and a suitable truth theory). We need to know how to understand what it is to have knowledge of such a theory, which is something that Davidson does not address and takes himself not to need to address.

Some philosophers have expressed skepticism about the prospects for an informative and explanatory account of what it would be to know a theory of meaning. Plainly, the only knowledge of a theory of meaning that speakers can be claimed to have is tacit or implicit knowledge, and it has been argued that no account of tacit knowledge can be anything other than a redescription of a speaker's competence (Quine 1972; Wright 1981). The putative problem is that the meaning theorist has no way of identifying which of a number of extensionally equivalent theories of meaning a speaker can properly be described as tacitly knowing. Evans confronts this line of argument in 'Semantic theory and tacit knowledge' (1981*a*). He begins from the thought that tacit knowledge of a theory of meaning involves having a disposition corresponding to each axiom of the theory. These dispositions, he argues, should be interpreted in a full-blooded sense, as requiring causally efficacious categorical bases.

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) . . . Understood in this way, the ascription of tacit knowledge 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 behavior 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. (CP_{329-30})

This way of thinking about the causal basis of comprehension offers a way of distinguishing between extensionally equivalent meaning theories, and hence of responding to the skeptical challenge. It also explains how a hearer can understand a previously unencountered sentence S. The hearer understands S in virtue of a complex set of dispositions corresponding to the constituents of S and derived from exposure to those constituents in different sentences and different contexts.

Evans's discussion of tacit knowledge deploys a theoretical notion that marks him off from all four of the philosophers we have been considering-the notion of nonconceptual content (although he only uses the expression "nonconceptual content" in VR). Evans is happy to concede that the causally effective inner states in terms of which he understands tacit knowledge are very different from "ordinary" states of knowledge or belief. Although there are overlaps on both the output side (in that someone with the tacit knowledge that p will be disposed to do many of the same things as someone who believes that p in the ordinary sense of belief) and the input side (in that someone will typically acquire the tacit knowledge that p in circumstances that might well induce the ordinary belief that p). Evans maintains that there is a fundamental difference in kind between tacit knowledge and the propositional attitudes. Although he does not himself put it in these terms, states of tacit knowledge are essentially one-track rather than multi-track dispositions. Whereas tacitly knowing a principle of a semantic theory manifests itself in a narrowly defined disposition to react to sentences and to use words in certain ways, the dispositions associated with beliefs have a significantly wider application and are at the service of a far wider range of projects.

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. (*CP* 337)

Evans's discussion of the distinctiveness of beliefs (and other propositional attitudes) returns to the important connection between semantic structure and inference. It is because propositional attitudes have structured contents that they can feature in such a complicated web of inferential connections. A thinker's appreciation of the inferential power of any given belief reflects her ability to appreciate the inferential power of other beliefs that have constituent concepts in common with it. As Evans puts it, "behind the idea of a system of beliefs lies that of a system of concepts, the structure of which determines the inferential properties which thoughts involving an exercise of the various component concepts of the system are treated as possessing" (*CP* 338).

Tacit knowledge states, in contrast, are single-track dispositions because they lack this structure. This makes them a class of what in VR Evans terms nonconceptual information states (VR §§5.2 and 6.3). In VR Evans's emphasis is on nonconceptual perceptual states and the role that they play in demonstrative thought. The discriminating knowledge required for demonstrative thought about an occurrently perceived object takes the form of an ability to keep track of it over time. Thinkers are able to do this when they stand in the appropriate information links to the object. These information links are nonconceptual, issuing in practical capacities to locate the object relative to oneself and to other objects-practical capacities that can be viewed as one-track dispositions in just the same way as the tacitly known principles of a semantic theory. Although similar ideas had been mooted by other theorists (two very different examples can be found in Stich 1981, which distinguishes propositional attitudes from what Stich terms subdoxastic states, and in the distinction between analog and digital content made in Dretske's 1981 discussion of the content of perception), Evans offers an original and sustained development of the idea that there might be content-bearing states that do not involve the exercise of concepts and that are fundamentally different from conceptual states.⁸ Campbell's contribution to this volume (Chapter 6) explores Evans's distinction between conceptual and nonconceptual content in the context of Evans's discussion of Molyneux's question (the question, discussed by Locke, of whether a congenitally blind person able to discriminate certain shapes by touch would be able, were they suddenly to acquire the capacity to see, to discriminate the same shapes using vision alone).

 8 Evans's discussion of nonconceptual content in VR inspired a number of theorists to explore the notion further. A number of the contributions to the ensuing debate are collected in Gunther 2003.

PAPERS

A central theme in VR is the project of reconciling two ideas frequently treated as incompatible—the idea that singular terms are Russellian, on the one hand, and the idea that they have Fregean senses, on the other. As Evans freely acknowledges (CP 291 n.), it was **John McDowell** who originally proposed to ascribe Fregean sense to Russellian singular terms (McDowell 1977). Fittingly, McDowell's contribution to this volume (Chapter 1) assesses Evans's claim in the first chapter of VR that this was in fact Frege's considered view of singular terms, despite his apparent willingness to ascribe sense to empty names.

McDowell agrees with Evans that the Russellianism discernible in the early (pre "On sense and reference" (1892)) Frege remains "submerged" in Frege's later thought, rather than being definitively discarded. It is true that Frege does seem explicitly to countenance the possibility of sentences with empty names expressing genuine thoughts, suggesting that we can understand what is going on here by analogy with fictional discourse. But McDowell agrees with Evans that this is fundamentally incompatible with how Frege understands thoughts. A thought, for Frege, is a postulate to the effect that the world is a certain way. The way the world is postulated to be is the truth condition of the thought—the way the world has to be for the thought to be true. To judge the thought is to judge that the truth condition holds. But, McDowell argues, we cannot apply the notion of a truth condition to the putative thoughts expressed by sentences with empty names. If the putative thought does indeed have such a truth condition, then there seems every reason to hold that the condition is not satisfied if the sentence expressing the thought contains an empty singular term-in which case the putative thought would come out as false, rather than as lacking in semantic value. This means that the putative thoughts expressed by sentences with empty names cannot have truth conditions-and the very sketchy comments that Frege makes about fiction give us no handle on what thoughts without truth conditions would be, or what it would be to judge them.

In addition to this line of argument (which develops points made by Evans at 1982: 24–5), McDowell identifies the stress in Evans's Frege on the object dependence of singular thoughts as part of a broader intellectual project—the project of "integrating our rational

powers with our natural situatedness in the world". These stem from McDowell's distinctive way of understanding Frege's introduction of the notion of sense. For McDowell, Frege's prime motivation for introducing the notion of sense is to accommodate the normative requirements of rationality.

The controlling aim of Frege's introduction of *Sinn* is to provide a conception of thoughts—possible contents of propositional attitudes and speech acts—that, in conjunction with a repertoire of concepts of kinds of propositional attitudes partly explained in terms of rational relations between them (for instance, that rationality precludes believing and disbelieving, or withholding judgment with respect to, the same thing), yields descriptions of ways in which minds are laid out such that the descriptions put a subject's rationality in question only when the subject's rationality is indeed open to question. (p. 48 below)

A thinker can rationally judge an object both to be F and not to be F if the thinker is thinking about the object in a different way (under a different mode of presentation) on each occasion. But this does not imply that we should characterize those modes of presentation in completely object-independent terms. Quite the contrary: to entertain a thought is already to be directed at the world, and in the case of a singular thought that directedness involves contextual relations that can only hold if the object exists. Traditional interpretations of Frege have not paid sufficient attention to these contextual factors (thereby losing touch with our "natural situatedness in the world"), while revisionist accounts of reference and content have lost touch with the idea that thinking involves the exercise of rationality. Evans's Frege tries to do justice to both constraints.

As we saw in the first part of the introduction, Evans does not think that all referring expressions are Russellian. Descriptive names (such as "Julius", introduced by the stipulation that "Julius" refers to the inventor of the zip) are referring expressions, since they satisfy the criterion of rigidity, and yet can plainly have a sense in the absence of a bearer. The implications that this has for the theory of truth are the subject of **R. M. Sainsbury**'s "Names in free logical truth theory" (Chapter 2). As Evans himself notes, the existence of descriptive names in a language means that a truth theory for that language cannot have a classical logic. The axiom for "Julius" takes the form:

```
(3) For all x ("Julius" refers to x iff x = Julius)
```

In classical logic (3) entails that Julius exists, contravening our assumption that descriptive names are not Russellian. Evans proposes a truth theory formulated in a negative free logic (a free logic is a logic in which the constants do not carry existential import, so that it is not the case that, for a given constant a, $\exists x (x = a)$). Within a negative free logic, even though (3) does not entail the existence of Julius, the existence of a non-descriptive and hence Russellian name such as "Hesperus" is nonetheless entailed by an axiom such as

(4) "Hesperus" refers to Hesperus,

since (4) is false if Hesperus does not exist. The significant difference between (3) and (4) is that (4) is a simple sentence, whereas (3) is a complex sentence (a simple sentence is one in which an n-place predicate is concatenated with n simple names). In negative free logic it remains the case that every simple sentence with an empty name is false.

Sainsbury's chapter explores a tension between Evans's need for a free logical framework, on the one hand, and his emphasis on truth theories as theories of interpretation, on the other. In using a truth theory as a theory of interpretation, we use the T-theorems of the theory rather than its axioms. That is, in interpreting a speaker's utterance of, say, "Hesperus is large", we deploy the T-theorem

(5) "Hesperus is large" is true iff Hesperus is large

without proceeding via an axiom such as (4). Even though the interpreter typically uses such an axiom in deriving the T-theorem, the interpretation will exploit the theorem regardless of how it is derived. The problem is that the distinction between Russellian and non-Russellian names cannot be captured at the level of the T-theorems in a free logical framework. In the case of (5) the right-hand side of the biconditional is a simple sentence, and hence false in the (counterfactual) case where Hesperus does not exist. Since in such a situation "Hesperus is large" will also be false, the biconditional will itself be true. As Sainsbury puts it, this obscures the distinction between Russellian and non-Russellian names in a way that seems to "emasculate" Evans's insistence on the object dependence of non-descriptive names, given his fundamental claim that an utterance with an empty singular term fails to express a genuine thought.

The resolution of the tension that Sainsbury offers on Evans's behalf exploits a principled modification of the standard truth theories that he (Sainsbury) has developed in previous writings in order to accommodate demonstratives and indexicals (see the papers in Sainsbury 2002). According to Sainsbury, the T-theorems for sentences involving demonstratives such as "that man" contain two parts. The first part essentially sets the scene for the utterance in question, by specifying the person whom the speaker is picking out, the place and time, and so on, while the second part gives the truth condition. So, for example, a T-theorem for Hermione's utterance of the sentence "That man is disgraced banker" might take the form:

(6) Speaking of Bob on I January 2004, Hermione said that he was a disgraced banker, and her utterance is true iff Bob is a disgraced banker whose fall from grace took place prior to I January 2004.

Since this idea of a two-part T-sentence is independently motivated, Sainsbury feels able to apply it to mark the distinction between Russellian names and descriptive names. The basic idea is that the object dependence or object independence of any given object language name will be given by the first, scene-setting part of the relevant T-theorem, in virtue of indications of the scope of the corresponding metalanguage names. Sainsbury follows Evans in using Russell's own scope-indicating device of square brackets. Prefixing an expression with a constant in square brackets indicates that the expression is false unless the constant bound by the square brackets has a bearer. So, for example, "[a] ~ Fa" will be false if "a" fails to refer, while "~ [a] Fa" will be true in the same circumstance. Clearly, Russellian names take the widest possible scope, and this will be clearly indicated in the appropriate T-theorems. A specimen such T-theorem is

(7) [Hesperus] "Hesperus is large" is true iff Hesperus is large.

The scope indication would be different for descriptive names, as follows:

(8) "Julius did not invent the vacuum cleaner" is true iff [Julius] Julius did not invent the vacuum cleaner.

(7) comes out as false if "Hesperus" fails to refer, while (8) is true if "Julius" fails to refer, thus giving the right result. As Sainsbury

José Luis Bermúdez

recognizes, his emendation imposes modifications elsewhere in Evans's semantics, but he argues that these remain well within the spirit of Evans's project. The final section of his paper explores Evans's motivations for thinking that natural language names must be either Russellian or descriptive.

Although Evans is studiously attentive to distinctions within the class of referring expressions (such as the distinction between Russellian names and descriptive names) and to the differences between referring expressions and other types of singular term, such as definite descriptions, he is curiously silent on the question of whether there can be such a thing as plural reference. He is emphatic that not all singular terms are referring expressions, but seems unquestioningly to accept that all referring expressions must be singular terms. Ian Rumfitt's contribution "Plural terms: another variety of reference?" explores the case for what he terms referentialism about plurals. The plurals in question include compound names, such as "William and Mary", collective names such as "the Aleutian islands", compound expressions such as "my teammates and I", and plural indexicals and demonstratives ("these flowers" and "we"). (Rumfitt leaves to one side plural descriptions such as "the men who independently discovered the calculus".) There are two ways of defending referentialism about plurals. The first way is to treat the referent of a plural term as semantically singular, so that plural terms come out as singular terms that refer to a distinctive type of object. The most obvious candidate objects are sets (e.g. the set of flowers to which I am currently pointing), mereological fusions (e.g. the fusion of the members of my cycling team), and aggregates (where aggregates, unlike sets, are physical objects with spatio-temporal locations). Rumfit rejects all three versions of singular referentialism in favor of a version of referentialism characterized by the following three theses:

- (i) Semantic predicates, such as "designate" and "satisfy", remain constant in sense whether they attach to a plural subject or a singular subject.
- (ii) In the principles that state the designations of plural terms the expressions that follow the verb "designates" (or "refers to") are themselves plural expressions.
- (iii) The designation relation does not distribute, so there is no entailment from " α refers to x & y & z" to " α refers to x".

In defending this type of referentialism, Rumfitt engages (in \S_4) with Dummett's neo-Fregean argument that plural phrases should be understood predicatively rather than referentially, and provides a limited defence (in \S_5 and 6) of Boolos's proposal to regiment arguments involving plurals by using second-order quantifiers and predicate variables.

In the final section of his paper Rumfitt deploys Boolos's secondorder account of the logic of plurals to explore how plurals behave in modal contexts. In the context of Evans's understanding of referring expressions, this is a very important question. As we saw earlier, Evans holds that genuine referring expressions must be rigid. The thesis has some plausibility for singular terms, but, as Rumfitt notes, the admissibility of plural referring expressions has a potentially disruptive effect. If there are non-rigid plural referring expressions, then they can only be accommodated by a semantic theory that relativizes reference to a possible world in precisely the way that Evans tries to rule out. Rumfitt himself argues that some varieties of plural are rigid, but he is agnostic about others. What would it be for a plural term to be rigid? Rumfitt holds that the rigidity of singular terms is given by appropriate instances of the following two schemata:

 $(Ria) \ x = a \rightarrow \Box (Ex \rightarrow x = a)$ $(Rib) \ \Diamond (x = a) \rightarrow x = a$

The first schema is effectively the principle of the necessity of identity (with "Ex" to be read as "x exists", while the second schema states that no object could possibly be identical to a unless it is actually identical to a. There are, Rumfitt holds, plural analogs of these principles. In the following "Tx" is to be read as "x is one of the T-things" and " E^2 (T)" as "the T-things exist".

```
\begin{array}{ll} (R_{2a}) & Tx \rightarrow \Box (E^2 \ (T) \rightarrow Tx) \\ (R_{2b}) & \Diamond Tx \rightarrow Tx \end{array}
```

In English, (R_{2a}) states that if something is one of the T-things, then, necessarily, if the T-things exist, x is one of them, while (R_{2b}) states that it is possible for something to be a T-thing only if that thing is actually a T-thing.

The simplest case comes with compound names formed by concatenating two or more rigid singular terms (as in "William and Mary"). Here, Rumfitt argues, the compound name inherits the

José Luis Bermúdez

rigidity of its singular constituents, and he offers a derivation to show how two singular terms each satisfying (*R1a*) will concatenate to form a compound term that satisfies (R_{2a}) . There is a certain intuitive plausibility in the claim that, if William is one of William and Mary, then he will be one of William and Mary in any possible world in which William and Mary exist. But other varieties of plural terms are less straightforward. Consider collective names such as "The Aleutian islands", for example. Is it really the case that, if a given island x is one of the Aleutian islands, then in any possible world in which the Aleutian islands exist, x must be one of them? Are there really no possible worlds in which there are more Aleutian islands than there are in this possible world? Plural demonstratives are also problematic. Let "these people" pick out the people in the line at the fast food restaurant. It seems plausible to say that had Martha not decided that the line was too long, Martha might have been one of them. But this is incompatible with $(R_{2}b)$. As Rumfitt notes, however, arguments in this area need to be formulated carefully. It is important not to prejudge the issue by tacitly assimilating plural terms to plural descriptions, for example. If "the Aleutian islands" is read as "the group of islands off the west coast of Alaska", then it seems hard to deny that there are possible worlds in which an island that is actually one of the Aleutian islands is not in the appropriate archipelago. And we also need to make sure that we are not tacitly importing additional uses of the demonstrative, imagining an utterance of "these people" made to a different group of people in line for the fast food restauranta group that on this occasion includes Martha. Rumfitt is surely correct when he says that these issues will need to be investigated further before we can make a full assessment of Evans's account of referring expressions.

Although Evans does not discuss pronouns in VR, his extended discussion of pronouns in "Pronouns, quantifiers, and relative clauses" (1977) and "Pronouns" (1980) is continuous with his discussion of proper names, demonstratives, and indexicals. The second paper was written in an interdisciplinary spirit, aiming to build bridges between the philosophically motivated semantics he proposed for pronouns and the extensive discussion of pronouns and anaphora in linguistics. As **Kenneth Safir** brings out in his paper "Abandoning Coreference" (Chapter 4), Evans's treatment of pronouns has had considerable influence within linguistics.

Evans made two fundamental contributions to the study of pronouns. The first (discussed briefly above) is his Frege-inspired development of a co-referential treatment of bound pronouns, which has the theoretical advantage of allowing both bound pronouns and pronouns with a singular term as antecedent to be understood in referential terms. To take an example that Safir also discusses, Evans's semantics allow us to give a uniform account of the pronoun "his" in the two sentences "John loves his mother" and "Everyone loves his mother". Evans's second contribution was to identify a class of pronouns that do not fit this model. These are what he termed *E-type pronouns*. These are pronouns with quantifier antecedents that nonetheless do not appear to operate as bound variables. Evans gives the following example.

(9) John owns some sheep, and Henry vaccinates them.

We cannot, Evans argues, understand (9) as saying of some sheep that they are such that John owns them and Henry vaccinates them. This would leave open the possibility, which (9) rules out, that John owns some sheep that Henry does not vaccinate. Evans proposes a different semantics for E-type pronouns, according to which they are singular terms whose denotation is fixed by a description recoverable from their quantifier antecedent. So, the denotation of "them" in (9) is fixed by the description "sheep that John owns".

In "Pronouns" (1980), Evans uses his semantics of pronouns to argue against an influential treatment of pronouns by Howard Lasnik (1976). Lasnik argued that relations of coreference between pronouns and their antecedents were not fixed by linguistic rule, but instead by pragmatic, extralinguistic factors. The paradigm case of pronominal coreference for Lasnik would be a sentence such as

(10) I'm glad he's gone

said of someone who has just left the room, where the person who has just left the room is suitably salient to both speaker and hearer. Of course, coreference cannot be completely unconstrained, and Lasnik proposes the following rule excluding certain forms of coreference;

(11) A name cannot be c-commanded by a coreferent noun phrase.

The technical notion of c-command is explained below (pp. 129–30), but the basic idea is that the position of pronouns in the hierarchical

José Luis Bermúdez

architecture of a sentence places restrictions on the noun phrases with which they can corefer. Evans makes a number of objections to Lasnik's proposal. He is concerned, for example, that it completely obscures the commonalities between pronouns with singular antecedents and pronouns with quantifier antecedents, arguing that these two types of pronoun form a semantic natural kind completely distinct from pragmatic uses of pronouns. In this context, Evans makes an important distinction (emphasized by Safir) between *dependent reference* and *intended coreference*. A pronoun is referentially dependent upon a noun phrase when it corefers with that noun phrase and picks up its referent from it. But there can be intended coreference between a noun phrase and a pronoun without dependent reference. A pronoun can corefer with a noun phrase even though it does not derive its reference from that noun phrase. Evans gives the following discourse as an example:

(12) What do you mean John loves no one? He loves John.

It seems clear (and it follows from Lasnik's noncoreference rule) that "He" cannot be referentially dependent upon the second occurrence of "John". Nonetheless, there can be intended coreference between the two (because the pronoun is referentially dependent upon the first occurrence of "John", which stands in the relation of intentional coreference to the second occurrence of "John").

In opposition to Lasnik, therefore, Evans proposes that all cases of coreference can be understood in terms of a linguistic rule enforcing a dependency relation and a principle blocking dependent reference in certain circumstances. Safir takes issue with Evans's central claim that coreference is enforced by linguistic rule, and (as the title of his paper suggests) argues that the notion of coreference is not the correct notion to use in thinking about pronouns. On the general picture that Safir favors, a number of linguistic principles constrain the possibilities of coconstrual (the notion he prefers to coreference, for reasons that will become apparent shortly), but there are no rules mandating the coconstrual of a pronoun with an antecedent (or following) noun phrase. One of these constraining principles is effectively Evans's reformulation of Lasnik's non coreference rule in terms of dependent reference. Safir terms this the Independence Principle (IP):

(13) IP: If X c-commands Y, then X cannot depend upon Y.

Introduction

Yet, Safir argues, the Independence Principle must be supplemented. He notes that those contexts, such as (12), where there is intended coreference without referential dependence are all contexts where the coreference is surprising—where there is, as he puts it, an expectation of noncovaluation. The Independence Principle does not explain where there should be such an expectation. We can obtain such an explanation, however, if we take the failure of dependence in such cases to be generated by what he terms the Form-to-Interpretation Principle (FTIP):

(14) FTIP: If X c-commands Y and Z is not the most dependent form available in position Y with respect to X, then Y cannot be directly identity dependent on X.

Safir proposes that expectations of noncovaluation arise in just those cases where the failure of referential dependence is generated by FTIP—this is what he terms Pragmatic Obviation (although it should be stressed that the FTIP itself is not a pragmatic principle in the sense made familiar by Grice, but rather an algorithm for computing dependency relations between formal representations).

One of the key claims of Safir's chapter is that Evans was fundamentally mistaken in thinking about the behavior of pronouns in terms of coreference, where two terms are coreferential just if they both refer to the same object/person. Safir argues that a number of linguistic phenomena show that there can be coconstrual without coreference. These are cases where a pronoun is to be understood in terms of the noun phrase on which it is referentially dependent, even though the pronoun cannot strictly speaking be described as referring to the referent of that noun phrase. One such linguistic phenomenon is effected by what Safir terms *proxy* readings of pronouns, where the pronoun refers to an object that in some sense stands proxy for the object picked out by its antecedent. Here are two examples:

- (15) As they strolled through the wax museum, Fidel could not help thinking that he would have looked better in a uniform, and Marlene could not help thinking that she would have looked better without one.
- (16) Patton realized that he would be vulnerable to a flanking movement.

In (15) "Fidel" and "Marlene" refer to the relevant individuals, while the pronouns refer to their respective waxwork models.

In (16) it is Patton's army, which is commanding from a distance, rather than Patton himself, that is vulnerable. The pronouns are referentially dependent upon their proper name antecedents, even though they do not refer to those antecedents.

Most of Evans's discussion of referring expressions in VR is taken up with demonstratives (such as "that" and "this", whose reference is fixed in part by an accompanying ostension) and indexicals (such as "now", "here", and "I", where reference is fixed as a function of the context of utterance), a topic that he also takes up in "Understanding demonstratives" (1981b). José Luis Bermúdez's contribution to the volume (Chapter 5) focuses on Evans's discussion of the first person pronoun. The overall aim of Evans's treatment of indexical expressions is to show, in opposition to direct reference theorists and almost all contemporary philosophers of language, that we can and should apply the Fregean model of sense to indexicals and demonstratives. Evans's position is particularly significant, since demonstratives and indexicals are frequently taken as prime examples of referring expressions that do not have a Fregean sense. For Evans, the need for the notion of sense arises because the successful use of indexicals and demonstratives to pick out objects is subject to the constraints imposed by Russell's principle. The requirements of Russell's principle are met in different ways by different types of indexical and demonstrative expression, but Evans's treatment of "I" incorporates many of the points that emerged in discussing other types of expression, such as "here" and "this". Evans takes a very metaphysical approach to explaining the sense of the first person, treating the task as requiring an exploration of the nature of self-consciousness.

Evans's account of the discriminating knowledge of oneself required to use and understand the first person pronoun has three components. We can understand the first as a way of fleshing out Frege's well-known remark that "everyone is presented to himself in a special and primitive way in which he is presented to no one else" (Frege 1918: 25). According to Evans, everyone who uses the first person pronoun with understanding does so in virtue of ways they have of thinking about themselves that are both primitive and not available to anyone else. What makes this possible is that self-conscious thinkers are suitably receptive to certain types of information—information that is distinctive because it feeds into judgment without requiring the thinker to identify a particular

Introduction

object as the source of that information. In the case of information about oneself that is derived from autobiographical memory, somatic proprioception, or introspection, for example, there can be no question but that one is oneself the source of the relevant information. There is no gap between judging on the basis of the appropriate information channels that there is *F*-ness, and judging that one is oneself *F*. These sources of information are *immune to error through misidentification relative to the first person pronoun*, to use the terminology that Evans adopts from Shoemaker (1968).

The second component in Evans's picture has to do with the "output" side of those first-person judgments for which information that is immune to error through misidentification serves as input. As Perry and others have emphasized, first-person thoughts have particular and direct implications for action. This is particularly evident when we are dealing with judgments based on information that is immune to error through misidentification, as with Perry's own well-known example of my seeing that a bear is coming towards me. There is no sense in which I can judge that the bear is coming towards someone and wonder whether I am that person. By the same token, nothing further is needed to rationalize an immediate reaction.

Since we are capable of entertaining thoughts about ourselves that involve possibilities to which we have no informational links, the sense of "I" cannot be exhausted by our sensitivity to information that is immune to error through misidentification in the required sense. Whereas judgments based on information sources that are immune to error through misidentification do not involve identifying a particular person as oneself, these more complex judgments do involve discriminating knowledge of oneself, and hence bring us back to Russell's principle. Evans's account of how Russell's principle is met for the first person hinges on the subject's ability to locate himself within the objective spatio-temporal world. As he puts it, "to know what it is for $[\delta = I]$ to be true, for arbitrary δ , is to know what is involved in locating oneself in a spatio-temporal map of the world (Evans 1982: 211). This knowledge, which Evans analyzes in terms of the ability to bring egocentric and objective ways of thinking about space into harmony, allows a thinker to determine the truth condition of any thought about himself (whether that thought involves predicates that are susceptible to error through misidentification, or immune to such error).

José Luis Bermúdez

Evans's stress on sensitivity to information that is immune to error through misidentification leads him to affirm the privacy of "I"-thoughts. Since the sense of the first-person pronoun depends upon the ability to exploit sources of information that concern only the thinker and are available only to the thinker, it follows that "I"-thoughts cannot be shared. Whereas Evans claims that the privacy of "I"-thoughts does not compromise their objectivity, Bermúdez argues that the shareability of "I"-thoughts is required both by the requirements of communication and by the need for objective thoughts to exist and have a truth-value independently of anyone thinking them. This leads him to propose what he terms the symmetry constraint upon a satisfactory account of "I"-thoughts. The symmetry constraint requires the token-synonymy (in a given context) of an utterance that I would express with the first-person pronoun and an appropriately related utterance that you might express with the second-person pronoun, so that your utterance of "You are F" comes out as synonymous with my utterance that "I am F", in a way that allows you to utter the negation of what I utter in the relevant context.

Although Evans's own account of the sense of "I" fails to meet the symmetry constraint, Bermúdez shows how it can be modified to meet it. He argues that Evans is wrong to include the information component in the sense of "I". Consider a first-person judgment of the form "I am F", where the information that one is F is derived from sources that are immune to error through misidentification. An account of the thought expressed will have to reflect the fact that it is derived from information that is appropriately immune to error through misidentification. But Bermúdez suggests that this fact is part of the sense of the predicate, rather than of the first person pronoun. Where the property of *F*-ness is such that there is no gap between judging on the appropriate grounds that F-ness is present and judging that one is oneself F, an account of what it is to possess the corresponding concept (the concept of F-ness, or the sense of "F") will include a first-person clause to that effect. Shifting the information component away from the sense of the first-person pronoun in this way means that privacy is not an immediate consequence of Evans's account of "I", but it does not vet show how it might meet the symmetry constraint. Bermúdez argues that the location component of the sense of "I" allows the symmetry constraint to be met. When you describe me as being F,

Introduction

your use of the second-person pronoun depends upon your knowledge of what it would be to locate me on a spatio-temporal map of the world. This knowledge, which is part of the sense of the second-person pronoun "you", is the very same knowledge that underpins my use of "I". Your use of a particular token of "you" in a given context and my use of a particular token of "I" in that same context both exploit the same practical ability.

Evans's innovative use of the concept of information to illuminate the nature of content is at the forefront of John Campbell's contribution, "Information processing, phenomenal consciousness, and molyneux's question" (Chapter 6). Campbell addresses some of the central issues raised by Evans's paper "Molyneux's question" (1985b), exploring how Evans's approach to Molyneux's question is informed by his account of the content of conscious perceptual experience, as developed in VR. Campbell takes issue with Evans's account of the content of perceptual experience, and in particular with what he sees as its conflation of subpersonal informationprocessing content and personal-level experiential content.

The underlying problem that motivates Evans in "Molyneux's question" has to do with the concepts of shape. It seems clear that our concepts of shape are derived from our experience of shape, and that we can have this experience through either the modality of touch or the modality of vision. Even though our tactile experience of shape is very different from our visual experience of shape, we appear to have unitary concepts of shape, rather than (as Bishop Berkeley suggested) specific concepts of shape tied to specific sensory modalities. But the fundamental differences in the phenomenal character of tactile and visual experience mean that there is a genuine and important question as to how the shape concepts acquired and used on the basis of vision can be the same as those acquired and used on the basis of touch. This is the question that Molyneux originally posed by asking whether someone born blind who has learned to discriminate certain shapes by touch would be able, were they suddenly to acquire the capacity to see, to discriminate the same shapes using vision alone.

Campbell presents Evans as arguing from the concept of egocentric space to the unitary nature of our shape concepts (and hence to an affirmative answer to Molyneux's question). Our shape concepts have the content they do in virtue of their connections to our egocentric ways of thinking about space. These egocentric ways of

José Luis Bermúdez

thinking about space are themselves a function of how we behave within our immediate environment and, as Evans puts it, "there is only one egocentric space, because there is only one behavioral space" (1982: 160). The content of shape concepts is a subspecies of egocentric spatial content, and since egocentric spatial content has to be understood through the ways we act upon the environment, the unitary nature of our behavioral repertoire secures the unitary nature of our shape concepts. We do not have distinct shape concepts for the tactile and visual modalities, because we do not have distinct behavioral repertoires associated with touch and vision. As Campbell brings out, however, there are two ways of interpreting the relation between egocentric spatial content and behavior. We can effectively assimilate the two, so that the egocentric spatial content of perception is actually constituted by its implications for action. On this view, perceiving an object at a particular egocentric location is to perceive it as something that might be reached by a certain set of movements, or that might be acted upon in certain ways were one to make the appropriate set of movements. Or we can interpret the distinction so that the egocentric spatial content of perception is the categorical ground of actions derived from perception, so that we perceive an object as something that can be reached by an appropriate set of movements precisely because we perceive it as having a particular egocentric location. Campbell notes that Evans's response to Molyneux's question works only on the first reading, but argues that the second reading is preferable.

Putting this objection to one side, however, Campbell finds a deeper problem with Evans's argument. The issue here is the extent to which sameness of egocentric spatial content can properly be expected to be transparent to the perceiving subject. Evans's argument from egocentric spatial content will yield an affirmative answer to Molyneux's question only if the perceiving subject is himself able to identify the sameness of egocentric spatial content that Evans claims to exist. Even if we grant that visual experience of shape and tactile experience of shape ultimately have the same content, there is a question whether this sameness of content is manifest to the perceiver. When Evans discusses conceptual contents (the type of contents that can be the objects of propositional attitudes and the meanings of sentences), he holds them to be constrained by what he calls the Intuitive Criterion of Difference, which entails that if two sentences express the same thought, this must be manifest to the thinker. But as Evans stresses, the content of perceptual experience is not conceptual content; rather, it is nonconceptual content. As we saw in the first part of the Introduction, nonconceptual content is not individuated by the holistic considerations of inferential role associated with conceptual thought. The question arises, therefore, whether nonconceptual content has the transparency property. At this point Campbell finds serious tensions in Evans's discussion. Evans has a unitary notion of nonconceptual content applicable both to subpersonal computational states (such as those involved, for example, in the subpersonal computation of the distance and direction of the source of a sound) and to conscious perceptual experiences (such as one's hearing a sound as coming from an object located at a certain distance and in a certain direction). The picture that Evans develops is one on which subpersonal information-processing contents become conscious when they become available to a "thinking, concept-applying, and reasoning system" (1982: 158). Nonconscious subpersonal information-processing contents clearly lack the transparency property. One might plausibly expect the specialized mechanisms responsible for processing tactile information to be insulated from the mechanisms that process visual information. Yet Evans thinks that once they become conscious, their sameness or difference is manifest to the thinking and perceiving subject. Campbell argues that this is a mistake—a mistake that reflects a more general problem with Evans's discussion of nonconceptual content. The mistake comes from assimilating the content of perceptual experience to the content of subpersonal information-processing modules. Campbell argues that we need to recognize that the content of perceptual experience is fundamentally different both from the content of subpersonal modules responsible for processing sensory inputs and from the content of beliefs and other propositional attitudes. These three types of content are related in complex ways that have not yet been adequately explored. The first step in clarifying these relations is making sure that they are kept carefully distinguished.

Whereas Campbell's contribution is focused on the nature of perceptual experience and the particular type of content that it has, **Christopher Peacocke**'s chapter, "'Another I': Representing conscious states, perception, and others" (Chapter 7), explores how we think about conscious perceptual experience, both our

José Luis Bermúdez

own and that of others. This investigation continues Peacocke's long-standing project (Peacocke 1992, 1999, 2004) of explicating the possession conditions of concepts. As suggested by the title of his chapter, which alludes to Zeno's characterization of a friend as "another I", one of the themes of the paper is the relation between the first-person and third-person aspects of our perceptual concept (where a perceptual concept is the concept associated with perception through a particular sensory modality, such as the concept *seeing*). We can see Peacocke's overall project as affirming the priority of the first-person dimension of perceptual concepts, while avoiding the epistemological and other problems that tend to go with a first-person/third-person asymmetry. According to Peacocke, perceptual concepts are what he calls deeply first-personal, because one's knowledge of what it is for any arbitrary person to be a perceiver will depend upon how one knows that one is oneself a perceiver.

Peacocke's starting point is what he terms the Core Rule, which he initially discusses in the context of the concept of seeing. The Core Rule for vision is that when a thinker sees that it is the case that p, he is entitled to judge "I am seeing that p". One important feature of the Core Rule is that it places the justificatory weight of a perceptual self-ascription upon the episode of perception, rather than upon any quasi-perceptual awareness of one's own perceptual experience. It is, in Peacocke's phrase, an "Outside-In" rule. Nor is this characteristic of the Core Rule a function of the factivity of seeing (of the fact that one cannot see that p unless p is actually the case). Peacocke offers an Extended Core Rule to cover cases where a perceiver is entitled to self-ascribe a visual experience even though that experience is illusory (and may be known to be illusory). The Extended Core Rule states that if a thinker is in a state that is subjectively as if she is seeing that p, then she is entitled to judge "I am having a visual experience as of p".

As Peacocke notes, the Core Rule and the Extended Core Rule respect Evans's observation in VR that a subject's "internal state cannot in any sense become an *object* to him. (He is *in* it.)" (1982: 227). We learn about our own experiences *through* our experience of the world, not by taking those experiences as objects. Nonetheless, the Core Rule differs significantly from Evans's own account of perceptual self-ascription. According to Evans, a thinker

works backwards from perceptual experience to perceptual selfascription in the following way:

He goes through exactly the same procedure as he would go through if he were trying to make a judgment about how it is at this place now, but excluding any knowledge of an extraneous kind. (That is, he seeks to determine what he would judge if he did not have such extraneous information.) The result will necessarily be correlated with the content of the informational state which he is in at that time. Now he may prefix this result with the operator 'It seems to me as though...'. This is a way of producing in himself, and giving expression to, a cognitive state whose content is *systematically* dependent upon the content of the informational state. But in no sense has that state become an object to him. (1982: 227–8)

Evans's proposal in effect calls for an act of the recreative imagination, expecting the thinker imaginatively to undergo an act of trying to determine how things are that abstracts away from the sort of background knowledge that thinkers would typically employ to determine whether the world really is as it perceptually appears to be. As Peacocke notes, the type of abstraction that Evans describes seems in many cases to stand in the way of accurate self-ascription of perceptual experiences. There are many cases where how we experience the world is directly a function of our background knowledge. Phonetically there is no difference between the English sentence "Peter leaped" and the German sentence "Pieter liebt". Yet one's experience differs depending upon how one hears the relevant phonemes, which in turn seems to depend upon information that would count as extraneous for Evans (such as one's knowledge that one is in Heidelberg, Mississippi, for example, rather than Heidelberg, Germany).

Peacocke shares with Evans (and a number of other contributors to this volume) a sense that philosophical accounts of concepts and conceptual abilities must respect psychological facts about how the concepts in question are acquired and deployed. Perceptual concepts have been extensively studied in the context of children's acquisition of what has come to be known as a "theory of mind", and Peacocke's discussion of the deeply first-personal nature of perceptual contexts is developed in the context of developmental data on young children's understanding of perception. The data he considers include,

José Luis Bermúdez

for example, the fact that 2-year-old toddlers do not seem to be able to appreciate that they can see something that cannot be seen by someone on the other side of an opaque screen, or that older children seem not to understand that someone in a different location who sees the same object is nonetheless likely to see a different part of that object. According to Peacocke, these and other phenomena are to be explained by young children's naïve use of the Core Rule in otherascriptions. Their ascriptions of perceptual experiences to others are based on straightforward extrapolation from their own case, whereas thinkers with a full grasp of perceptual concepts will appreciate that there are significant constraints and conditions upon applying the Core Rule in all cases where other perceivers are not in the same location and under the same conditions. Moreover, full mastery of the practice of ascribing perceptual states to others requires not simply working outwards from the application of the Core Rule in one's own case, but also appreciating that the other person is herself capable of deploying the Core Rule to make perceptual selfascriptions. In this respect, as Peacocke notes, the deeply first-personal nature of perceptual concepts fits well with simulationist approaches to understanding other minds-both because it suggests a development in terms of imagining what it would be like to perceive the world from another person's perspective and because it depends upon reasoning in accordance with the Core Rule rather than (as a more theoretical approach would require) explicitly formulating and applying the Core Rule.

Evans was interested not just in how we should explain what experience is and how we think about it, but also in the more Kantian question of whether there are any necessary conditions upon our having the type of experience that we have. He addresses this question in "Things without the mind" (1980*a*), which addresses Strawson's discussion of the relation between space and objective experience in chapter 3 of *Individuals* (1959) and was originally published in a Festschrift for Peter Strawson. Evans and Strawson are both discussing what makes it possible for us to experience and think about an objective world, where an objective world is a world that can exist whether or not we are perceiving it. Both, although in different ways and on different grounds, defend the Kantian thesis that space is a necessary condition for objective experience and hence that we cannot think about or experience an objective world that is not a spatial world.

As Quassim Cassam points out in "Space and objective experience" (Chapter 8), there is a number of different ways of developing what he terms the Spatiality Thesis (ST). One might, for example, argue that the existence of space is a necessary condition for objective experience. This line of argument, which Cassam labels ST_e, attempts to derive metaphysical conclusions from epistemic conditions upon experience. Alternatively, one might argue that the *idea* of space (ST_i) or the *perception* of space (ST_p) are epistemic conditions of objective experience. These different arguments are not completely independent of each other. Since perception is factive, an argument for ST_n might turn out to be an argument for ST_e, although it might equally turn out that what really matters for objective experience is not the perception of space (in the factive sense), but rather the experience as of space. In any case, Evans's arguments in "Things without the mind" are focused primarily on ST_i, whereas Kant's own arguments in the Critique of Pure Reason seem to be intended to support ST_p. (It is not altogether clear how to interpret Strawson's arguments in Individuals, but they seem to be directed at ST_i.) Cassam's paper argues that Evans's two principal arguments for ST_p are unsuccessful, but offers in their place a more directly Kantian argument for ST_p.

Strawson originally set up the issues by asking us to consider a No-Space world composed only of sounds and inquiring whether it would be possible for a thinker in such a world to think of it as an objective world. The conclusion he reached is that objective experience would be available to such a thinker only if the auditory universe effectively had built into it an analog of space (what Strawson calls the Master-sound). In "Things without the mind", Evans considers and rejects Strawson's arguments, for two reasons. The first is that the Master-sound is not as quasi-spatial as Strawson claims, since there is no need for it to be ordered in the way that space must be ordered. The second is that the Master-sound is not really required, since what is really important is that the thinker's experience should exhibit the type of regularity that would make it appropriate to think of its course as due simultaneously to how the world is laid out and to the subject's movement through it. This second criticism of Strawson introduces one of Evans's key ideas in this area. This is the idea that our perception of the world is bound up with what he calls "a simple theory of perception", by which he means a quasi-theoretical understanding of what make it possible for us to have the experiences that we do.

Evans's first argument for ST_i, which Cassam terms the Argument from Perception, is based on this idea that a simple theory of perception is an indispensable basis for the idea of existence unperceived. Such a theory can contain, Evans argues, only two possible explanations for why something perceptible should not be being perceived at a given moment. It might appeal to deficiencies in the perceiver or to the absence of factors causally necessary for perception. Evans argues that deficiencies in the perceiver alone cannot yield the idea of existence unperceived, because the subject has no independent understanding of what it would be to be suitably receptive other than to be perceiving the relevant object, which makes citing a lack of receptivity to explain the absence of perception completely uninformative. In contrast, incorporating spatial notions into the simple theory of perception gives a clear explanation of why a perceptible phenomenon might not be perceived (one might be at the wrong location, for example). As Cassam points out, however, Evans gives a convincing account of how a simple theory of perception incorporating spatial concepts would allow a thinker to make sense of existence unperceived, but does not make the case that this is the only way of making sense of existence unperceived. He finds a similar difficulty for Evans's second argument for ST_i (the Substance Argument), which is based on the idea that we need to think of things in terms of primary properties in order to think of them as persisting in a way that would allow them to exist unperceived. It may well be the case, Cassam argues, that persistence and existence in space go together in our conceptual scheme, but Evans is trying to establish the stronger conclusion that existence in space is a necessary feature of any conceptual scheme that allows for persisting objects that can exist unperceived.

After rejecting Evans's attempts to establish ST_i Cassam turns to a more directly Kantian argument for ST_p —for a necessity claim about the perception, rather than the concept, of space. The central claim here is that the perception of space is necessary for perceiving objective particulars, because one cannot perceive an objective particular without perceiving it as having spatial properties such as shape, extension, and location. Cassam recognizes that the thesis needs to be circumscribed to deal with objective particulars, such as sounds, that might be thought not to have spatial properties, and

Introduction

accordingly proposes ST_{pb}, according to which the perception of space is necessary for objective experience insofar as it is necessary for the perception of bodies. It is obviously not the case that a perceiver must perceive all of an object's spatial properties. In fact, Cassam argues, it is only the perception of location that can plausibly be argued to be strictly necessary for the perception of bodies. Cassam suggests that ST_{pb} is far more promising than the version of ST_i defended by Evans, not least because it offers the possibility of explaining what might ground the necessity claim. Whereas Evans follows Strawson in seeing necessities in this area as essentially conceptual necessities, secured by the structure of our conceptual scheme, Cassam sees the possibility of an empirical and/or metaphysical basis for the necessity claim. He points to two possible argumentative strategies. The first highlights the role that information about spatial location has been suggested to play in solving the so-called binding problem, while the second derives from the fact that material objects are individuated according to their location. In each case it appears that we have the material for an explanation of why the Spatiality Thesis holds.

Turning now to a very different aspect of Evans's thought, E. J. Lowe's "Identity, vagueness, and modality" (Chapter 9) evaluates Evans's very influential one-page article "Can there be vague objects?" (1978). It is a familiar idea that our language is vague; that there are terms for which it is indeterminate whether they apply to certain objects. There are borderline cases of hairlessness for which there is no fact of the matter whether they count as bald or not. Far more contentious is the idea that reality itself might be vague-that there might be, as Lowe puts it, ontic vagueness in addition to semantic vagueness. Evans's article contains an attempted reductio ad absurdum of the idea that there might be vagueness actually in the world, in addition to in our description of the world. The argument is as follows. (I follow Lowe in using the more perspicuous lambda notation over Evans's own notation. The expression " $\lambda x[Fx] a$ " is to be read as "a has the property of being F" and " $\nabla \phi$ " as "It is not determinately the case that ϕ ", with $\Delta \phi$ " as "It is determinately the case that φ ").

(i) $\nabla (a = b)$ (ii) $\lambda x [\nabla (x = a)] b$

(ii) $\forall x \in V$ (x = a)(iii) $\neg \nabla (a = a)$ (iv) $\neg \lambda x [\nabla (x = a)] a$ (v) $\neg (a = b)$ (vi) $\Delta \neg (a = b)$

Evans's point is that assuming the indeterminate identity of a and b means that there is a property that b has (ii) but a does not (iv). This is the property of being indeterminately identical to a. We are assuming that it is indeterminate whether a is identical to b, but there is clearly no indeterminacy about a's identity to itself. Hence, an application of the principle of the non-identity of discernibles yields the conclusion that a is not identical to b (v). Since this conclusion is reached by a series of truth-preserving steps from premises that are being assumed to be true, it follows that it is determinately true, and hence that there is, contradicting (i), a fact of the matter about whether a is identical to b (vi).

Lowe suggests that Evans's argument is subtly question-begging. He grants that there is no indeterminacy about a's identity to itself, but notes that Evans's application of the principle of the non-identity of discernibles rests upon taking it to be equally determinate that there is a property that b has but a lacks, viz. the property of being indeterminately identical to a. Lowe claims, however, that this is not warranted by the original assumption. If it is indeterminate whether a is identical to b, then it must also be indeterminate whether b has the property of being indeterminately identical to a. One can only assume, as Evans does, that it is a fact that there is a property that a lacks but b has if one tacitly assumes that it is determinately true that a is not equal to b, which is of course the conclusion of the argument.

Lowe's principal charge against Evans's argument, though, is not that it is question-begging, but rather that it is invalid. He draws an instructive comparison with the well-known family of arguments for the non-contingency of identity inspired by Ruth Barcan and Saul Kripke. As he notes, Evans's argument against ontic vagueness can be easily transformed into an argument for the non-contingency of identity by reading Evans's indeterminacy operator (" ∇ ") as "It is contingently the case that". Construed in this manner, moreover, the argument would not be question-begging in the manner just identified. If *a* is contingently identical to *b*, then it seems clear that the property of being contingently identical to *b*.

40

Introduction

Yet there is a fallacy of equivocation between the step from (iii) to (iv) and the step from (iv) to (v). Continuing to read " ∇ " as "It is contingently the case that", the step from (iii) to (iv) is the move from "It is necessarily the case that a = a" to "It is not the case that *a* has the property of being contingently identical to *a*". Presumably what makes it true that it is not the case that *a* has the property of being *contingently* identical to *a* is the fact that *a* has the property of being *necessarily* identical to *a*.

It is here that the putative fallacy of equivocation arises, since there are two very different properties that might be in play here. The first is the property, which a shares with b and every other object, of being necessarily identical to itself. The second is the property, peculiar to a, of being necessarily identical to a. It is uncontentious, according to Lowe, that "It is necessarily the case that $a = a^{"}$ entails "a has the property of being necessarily identical to itself", which validates the move from (iii) to (iv). But this will not secure Evans's conclusion, since b also has this property, thus blocking the application of the principle of the non-identity of discernibles in the move from (iv) to (v). The step from (iv) requires *a* to have the property of being necessarily identical to *a*. But if this is the property in question, Lowe argues, the step from (iii) to (iv) no longer seems plausible. The thesis that *a* is necessarily identical to itself is a substitution instance of a law of logic (the law of the reflexivity of identity). But the thesis that *a* is necessarily identical to *a* is a controversial metaphysical claim (and in fact a close relative of the controversial metaphysical claim that the argument is trying to establish). Lowe maintains that Evans's argument commits an analogous fallacy.

Evans's Frege

JOHN McDOWELL

1. It helps to start with Russell. In the Theory of Descriptions, Russell gives an account of the logical form of sentences with definite descriptions in what might intuitively be thought of as subject position—sentences of the form 'The F is G'. Russell's parsing amounts to 'There is exactly one thing that is F, and it is G'.¹

This distinguishes the logical form of such sentences from the logical form of sentences with genuinely referring expressions— 'logically proper names'—in what, in this case, really is subject position.

In Russell's conception, sentences with logically proper names in subject position express thoughts (or propositions) whose availability to be thought or expressed depends on the existence of the objects referred to by the logically proper names.

Suppose someone sets out to express a thought by uttering a sentence containing a definite description. Russell urges that lack of an object that uniquely satisfies the description cannot imply absence of a thought expressed. Definite descriptions owe their capacity to contribute to the thought-expressing powers of sentences to the fact that they are constructed in a familiar way out of independently significant words. Their significance, like that of a specification, should be independent of whether or not there is anything the specification fits.

Admittedly there is a superficial parallel between the form 'The F is G' and the form 'a is G' (where 'a' marks a place for a logically

Thanks to José Luis Bermúdez for comments on an earlier draft.

¹ See Russell 1905. The focus on monadic predications is of course inessential.

proper name). If there is an object that uniquely satisfies what replaces 'F' in an instance of the former, something is true of the description that is also true of the logically proper name in an instance of the latter. Both determine a certain object as what the thought that is being expressed concerns, in the sense that the truth or falsity of the thought depends on how it is with that object. Indeed, we could use that formulation to define a certain conception of singular reference.

But on Russell's view this superficial parallel masks a deep difference. Where absence of a suitable object would require us to find no thought expressed if we assimilated the two logical forms, the Theory of Descriptions finds a thought that is no worse than false. If we focus on this difference, we shall not be inclined to make much of the conception of singular reference that groups definite descriptions—though only those that do single out objects—and logically proper names together.²

2. In Russell's conception, thoughts expressible with the help of logically proper names—genuinely subject–predicate thoughts—are accessible only to thinkers who are 'acquainted' with the objects in question.³ And on his official account of acquaintance, one is acquainted only with things that figure in one's immediate consciousness, conceived in a rather Cartesian way: bits of the sensory given, bits of what is given with similar immediacy in recollection, and (as long as Russell believes in it) one's own self. Outside the range of one's acquaintance, one can direct thoughts at objects only in the way the Theory of Descriptions provides for.

So the apparatus of the Theory of Descriptions cannot have its application limited to sentences that actually contain definite

² This is the questioning of the 'traditional grouping' of singular referring expressions for which Evans gives Russell credit (1982: 1–3). For his own part, Evans goes on to argue that 'Russell's criterion *for being a referring expression* simply will not stand up' (p. 53, citing the treatment of 'descriptive names' at pp. 46–51; see also pp. 30–8). Accordingly, he needs a different ground from the one I have rehearsed here (which he cites as 'the most important argument Russell gave', p. 52) for refusing to count pure definite descriptions as referring expressions. For Evans himself, singular terms need not be 'Russellian'. But since Evans argues that for *Frege* singular terms were 'one and all Russellian' (1982: 47), I shall suppress this complication here, and work with Russell's original criterion, which is at any rate a criterion for being a Russellian referring expression. If Evans is right about Frege, the conception that underlies Russell's grouping is the only one that matters for him. ³ See, e.g., Russell 1917.

John McDowell

descriptions. The theory's characteristic form comes to figure also in Russell's account of the thoughts that speakers have in mind when they utter sentences containing what would ordinarily be regarded as singular referring expressions, if the objects in question are outside the range of the speakers' acquaintance, on Russell's restricted conception of acquaintance. This goes for ordinary proper names, say of people other than oneself, and ordinary uses of demonstratives, as opposed to the peculiar demonstratives that Russell identifies as logically proper names. In these cases too we have to suppose that speakers direct their thoughts at objects—if there are objects such that the truth or falsity of the thoughts turns on how it is with them—by exploiting specifications of the objects.

This extension of the apparatus would chime neatly with the original motivation for the Theory of Descriptions. As long as we are confined to objects of acquaintance on Russell's official conception, it would be hard to make sense of a risk that someone might utter a sentence intending to express a genuinely subject-predicate thought, but turn out to have expressed no thought because the belief that there was a suitable object was mistaken. It would be natural to think that one cannot be mistaken about the existence of what is given in one's immediate consciousness. But outside the range of one's acquaintance, there is room for scepticism-though perhaps only a hyperbolic Cartesian scepticism in some cases-about the existence of objects one might want to be able to direct one's thoughts at. Here, as in the original application, extending the Theory of Descriptions would provide thoughts immune to the fate that Russell cannot tolerate in the original case, of turning out to be only illusions of thoughts if the belief that suitable objects exist was mistaken. The worst that could happen is that they would be thoughts all right, but false ones.

It has to be acknowledged that this motivation—avoidance of illusions of thoughts—is not at centre stage when Russell recommends extending the apparatus of the Theory of Descriptions to utterances in which, say, ordinary proper names figure.⁴ What Russell urges is not that if one tries to frame subject–predicate thoughts about

⁴ This implies that the way Evans depicts Russell, at 1982: 44-5, needs qualification.

Evans's Frege

objects outside the range of one's acquaintance, one runs the risk that the supposed thoughts turn out to be illusory. The attempt to express such thoughts is ruled out directly by the fact that one is not acquainted with their objects; Russell does not appeal to the fact that belief in the existence of their objects is chancy.

In fact, though he thinks that such thoughts are beyond the reach of our minds, Russell nevertheless exploits them, somewhat strangely, in his picture of how we communicate with the relevant kinds of sentences. If I say something 'about Bismarck', perhaps using that ordinary name, the thought in my mind must have the sort of shape that the Theory of Descriptions provides for, since I am not acquainted with Bismarck. But in speaking as I do, Russell holds, I 'describe' a subject-predicate proposition about Bismarckone that only Bismarck himself could get his mind around, since only Bismarck himself was acquainted with the object in question. Someone other than Bismarck who understands me similarly thinks a descriptive thought, with a similar relation to that singular proposition, which neither of us can actually think. It is because each of us thinks a proposition that 'describes' that same singular proposition, not thinkable by either of us, that we are in communication with each other.5

This comports badly with supposing that Russell's motivation for extending the Theory of Descriptions even includes the wish to avoid invoking thoughts of a kind such that their very existence might turn out to be an illusion. If Bismarck is a hoax on the part of historians, the picture that Russell thinks I must have of how I can communicate by talking 'about Bismarck' involves my purporting to 'describe' a thought that by Russell's lights would not exist on that hypothesis.

However, this fact about Russell's thinking need not matter for a broad-brush picture of one way in which it was widely received, which is all that my present purposes require.

The Theory of Descriptions makes straightforward sense of a way in which thoughts can be targeted on objects. Many philosophers were captivated by this. And they took encouragement from Russell in supposing that this way of directing thoughts at objects need not be restricted to cases where the specification by

⁵ See Russell 1912: 31. There is a fine discussion of this part of Russell's thinking in Mark Sainsbury, 'Russell on names and communication', in his 2002: 85–101.

John McDowell

means of which a thought singles out its object is made explicit in the relevant utterances. There was an intelligible tendency to suppose that the Theory of Descriptions captured the very idea of a thought's or an utterance's being directed at an object. Thoughts that single out objects do so by way of specifications.⁶

As I stressed, for Russell himself the idea that a description can single out an object as what a thought concerns, in the sense that the truth or falsity of the thought turns on how it is with that object, marks only a superficial match with the semantical character of logically proper names. In Russell's view, logically proper names enable the expression of thoughts that are about objects in a deeper sense. It is understandable that this should have tended to be forgotten, given how Russell's treatment restricts, almost to vanishing point, our repertoire of the kind of thought we can express with the help of logically proper names. But it remains the case that the extended Theory of Descriptions does not capture Russell's view of genuinely subject–predicate form. In this general 'descriptivism', what it is plausible to identify as Russell's own conception of singular reference goes missing.

3. How does this Russellian picture relate to Frege's distinction between sense and reference, as it applies to singular terms?⁷ First we need a quick sketch of the point of Frege's distinction.

What difference does the presence of a singular term make to the capacity of sentences containing it, perhaps as uttered on suitable occasions, to express thoughts?⁸ Part of the answer is

⁶ For an excellent example of this kind of approach, see John R. Searle's (1958) treatment of proper names, which is notable for its ingenuity about how to arrive at the content of the specification that, according to Searle, displays how a name singles out its object. Searle conceives his topic as how ordinary proper names refer; his Russell-inspired treatment is not responsive to the Russellian thought that might be put by saying that 'reference' by specification is not genuine reference— not what can be done by exploiting logically proper names. For a conception of this kind applied to the object-directedness of thought in general, not restricted to thought expressed with the help of proper names, see Searle 1983.

⁷ In this formulation, I have reverted to 'reference' for Frege's *Bedeutung*— 'Meaning' in Evans's exposition, following the more recent translations. See Sainsbury 2002: 225. I shall mostly stay with the German term. My formulation is designed to allow for the fact that Frege's distinction applies to meaningful expressions in general, not just singular terms; the application to whole sentences, in particular, will come up later.

 $\frac{8}{5}$ The allusion to occasions of utterance is strictly needed to accommodate indexicality and similar phenomena. But my concerns here are at a level of abstractness

Evans's Frege

that, at least in primary occurrence, and at least if it does refer to something, a singular term serves to indicate which object the thought expressed by a sentence concerns, in the sense that the truth or falsity of the thought turns on how it is with that object.

But if that were the whole of the answer, we would not be able to distinguish the contributions to the expression of thoughts made by pairs of singular terms that refer to the same object. We would not be able to distinguish the thoughts expressible by sentences that are alike except in that where one has one such term, the other has the other. And Frege's idea is that at least sometimes we ought to want to make such a distinction. There is no need for it if the one term differs from the other 'only as object'-that is, presumably, if they are merely notational variants (1892: 57). But there can be cases in which a rational subject who understands both terms can, say, believe the thought she understands as expressed by one sentence from such a pair while disbelieving the thought she understands as expressed by the other. To suppose that she both believes and disbelieves the same thing would bring her rationality into question, though ex hypothesi her combination of attitudes is consistent with her being rational. We can smoothly maintain the hypothesis of rationality if it is not the same thing that she believes and disbelieves. Her understanding of the two sentences associates them with different thoughts. And, since the sentences that express the different thoughts are alike apart from the singular terms they contain, it must be the difference in singular terms that accounts for the difference in thought expressed. It must be that though the singular terms have the same *Bedeutung*, they differ in their contributions to the thoughts expressible by sentences that contain them. That is what a difference in sense (Sinn) is, on the part of subsentential expressions. (A difference in Sinn on the part of whole sentences just is a difference in thoughts expressed.) Thoughts about the same object can differ in how the object figures in them, how it is presented by singular terms that refer to it in expressions of the thoughts.9

at which this can pretty much be ignored, and I shall continue to talk often of sentences as (capable of) expressing thoughts.

⁹ This explanation of the point of saying that coreferring singular terms can differ in sense follows Frege's exposition in the letter to Jourdain that Evans cites (1982: 14–15) (the 'Afla'-'Ateb' case).

John McDowell

On this account, the controlling aim of Frege's introduction of Sinn is to provide a conception of thoughts-possible contents of propositional attitudes and speech acts-that, in conjunction with a repertoire of concepts of kinds of propositional attitudes partly explained in terms of rational relations between them (for instance, that rationality precludes believing and disbelieving, or withholding judgement with respect to, the same thing), yields descriptions of ways in which minds are laid out such that the descriptions put a subject's rationality in question only when the subject's rationality is indeed open to question. If we find ourselves shaping up to crediting a subject with, say, believing and disbelieving the same thing, even though we can find no fault with the subject's rationality,¹⁰ we are to go back and find a way to distinguish what she believes from what she disbelieves. This is what dictates that we find, or postulate, a semantically relevant respect in which coreferring singular terms can differ.

We might say that Frege's introduction of *Sinn* reflects an idea along these lines: the very idea of a configuration in a mind needs to be seen in the context of the concept of rationality.¹¹

It can be illuminating to draw a connection between this way of spelling out what Frege wants to do with the notion of *Sinn* and his attacks on psychologism about logic. His target there is a conception of logical laws as laws of thought, understood on the model of laws of motion. On this conception, the topic of logic is how minds actually proceed in certain transitions from beliefs to beliefs. (Perhaps how minds normally proceed, but psychologism would understand the invocation of normality statistically, not normatively.) Frege objects that if we conceive our topic as a certain

¹⁰ Perhaps a failure to bring the two states together for rational assessment counts as a failure of rationality, in which case the proviso excludes such cases. In any case, they are not what we need to consider in order to understand Frege's point. If it dawns on me that I have contradictory beliefs, what dawns on me is precisely that I believe and disbelieve the same thing, and now rationality requires me to make an adjustment. Frege's point applies to cases where rationality requires no adjustment in attitudes even in a subject who is consciously adverting to both the attitudes in question. If I do not know that Paderewski (the statesman) and Paderewski (the pianist) were the same person, there is nothing irrational about my, say, regarding it as unlikely that Paderewski (the pianist) ever engaged in politics, even though I know a bit about Paderewski (the statesman).

¹¹ Compare Donald Davidson's well-known invocation of the idea of 'a constitutive ideal of rationality' in Davidson 1970. region of facts about how minds work, we shall not get logic into view at all. Logic deals with some of the ways in which minds *must* work, with 'must' expressing requirements of reason.

We can put the point here in parallel with the point about the concept of sense. It is not that the concept of deductive inference is alien to the very idea of moves that minds in fact make. The concept of deductive inference is indeed a resource for describing transitions from one layout of mind to another. But the concept makes sense only in the context of an idea of requirements of reason. Just so with the concept of, say, what someone believes. It is a resource for describing certain configurations of minds. But here too, the concept makes sense only in the context of the concept of rationality. By virtue of its connection with concepts of propositional attitudes, it seems right to say that the concept of sense belongs to psychology. But if we say this, we must conceive psychology otherwise than psychologistically.¹²

What I have said so far does no more than locate the notion of Sinn in terms of a role it is to play. Thoughts-senses expressible by whole sentences-are to be propositional-attitude contents, individuated in accordance with what Evans (1982: 18-19) calls 'the Intuitive Criterion of Difference'. Senses of subsentential expressions are to be individuated in a way that allows them to be seen as the contributions made by the expressions to the thoughts expressible by sentences they occur in. And of course this is not yet to say anything specific about what the sense of this or that referring expression might be.¹³ But that is nothing to complain about. When we locate the concept of sense in terms of its role, we equip ourselves with a general frame in which to place the detailed accounts of the senses of different types of referring expressions that Evans goes on to offer. Indeed, we need something at this level of abstraction in order to appreciate exactly what those detailed accounts are supposed to be accounts of.

4. What we have so far is that singular terms with the same *Bedeutung* can differ in sense, and the difference in sense is a difference in how objects figure in thoughts expressible by sentences

¹² A psychologistic psychology of propositional attitudes could be described as a 'physiology of the understanding', echoing Kant's remark about Locke in his *Critique of Pure Reason*, Aix.

¹³ As Evans says (1982: 18), 'Frege never said much about particular ways of thinking of objects'.

containing the singular terms—a difference in 'way of being given' (*Art des Gegebensein*), mode of presentation.¹⁴

In one place Frege illustrates this idea by suggesting that Aristotle might figure, in some thoughts expressible by using the name 'Aristotle', as the pupil of Plato and teacher of Alexander (1892: 58). The words of Frege's gloss here have the form that Russell discusses under the head of 'definite descriptions'. And that may have helped to encourage a widespread assimilation of what Frege is aiming at, when he credits singular terms with *Sinn* as well as *Bedeutung*, to the neo-Russellian 'descriptivism' that I sketched earlier.

On this reading, Frege's idea that singular terms have *Sinn* anticipates the Theory of Descriptions in its generalized application. Frege and Russell—but, in line with the reception of Russell that I described earlier, Russell deprived of his conception of genuinely subject–predicate form—figure together, for instance, as progenitors of the target of Saul Kripke's influential attack on descriptive conceptions of singular reference (Kripke 1980).¹⁵

In the aftermath of the revolt against 'descriptivism', in which Kripke's work is a landmark, much theorizing about singular

¹⁴ Sainsbury proposes a pared-down Fregeanism that discards the idea of mode of presentation as a gloss on the attribution of *Sinn* to singular terms (2002: 2–3). Part of his reason is the fact, exploited by Evans, that the idea is hard to reconcile with taking in stride singular terms with *Sinn* but no *Bedeutung*, which is central to Fregean thinking on Sainsbury's reading. But whether the thesis that there can be *Sinn* without *Bedeutung* merits that centrality is just what is going to be at issue here. It is hard to see how the basic Fregean conception, as expounded for instance in connection with the 'Afla'-'Ateb' case, could do without an idea of different ways in which an object can figure in thoughts, and 'mode of presentation' ('way of being given') need do no more than express that idea. It need not encourage the pressure towards reductivism that Sainsbury, rightly to my mind, resists.

¹⁵ Remarks like the one about Aristotle are certainly not sufficient ground for concluding that Frege belongs in Kripke's target area. The remark is about how Aristotle figures in certain thoughts (not about how the reference of the name 'Aristotle' is fixed). But even though it is not just about reference fixing, it need not imply that, if Aristotle had died in infancy and someone else studied with Plato and taught Alexander, those very thoughts would still have been thinkable, but would have had that other person as their topic; or that if no one had studied with Plato and taught Alexander, those very thoughts would still have been thinkable, but would have had no one as their topic. Frege's remark is about a way in which, thanks to certain assumed facts, *Aristotle* can figure in some thoughts. It is consistent with supposing that in those envisaged alternative circumstances, there could not have been the thoughts we are considering, thoughts in which Aristotle figures as the student of Plato and teacher of Alexander.

thought and its expression came to focus on certain contextual relations, typically of a causal character, in which objects can stand to episodes of thought and speech.¹⁶ In a newly dominant conception of how singular reference works, directing thoughts at objects by exploiting such relations replaced targeting thoughts on objects as those that conform to specifications. This made room for a kind of recovery of the forgotten parts of Russell's thinking. The relations with objects that are prominent in the new thinking about reference take on something like the role of Russell's concept of acquaintance. But genuine subject-predicate form does not now need to be subject to the tight restrictions that were imposed by the requirement of acquaintance as Russell applied it. Not that Russell is usually conceived in this way, as a forerunner of the newly dominant conception. The forgotten parts of his thinking tend to stay forgotten, and he figures only as a proponent of the 'descriptivism' that the new thinking aims to supplant.

With Frege assimilated into the general 'descriptivism' that the newly dominant conception defines itself against, the new orthodoxy is conceived as rejecting Frege's apparatus of Sinn and Bedeutung, or at least as requiring substantial modification in its employment. This has fateful consequences for how philosophers conceive the connection between how thought and speech are directed at objects, on the one hand, and rationality, which provides the frame within which Frege's notion of Sinn functions, on the other. At least in the sorts of case that fuel the rejection of 'descriptivism', vestigial versions of Frege's notion of Sinn come to figure, if at all, only in characterizing configurations in minds thatin a supposedly required divergence from Frege's idea that Sinn determines Bedeutung-at most partly determine which objects thoughts concern, needing extraneous help in singling out objects from causal relations between thinkers and objects, now conceived as obtaining outside the sphere of a subject's rationality.¹⁷

¹⁶ Here we should note Evans's 1982: ch. 3, where he protests against the way in which the generalized causal conception of singular reference was supposed to be encouraged by Kripke's own suggestions about how uses of proper names trace back, through the continuation of practices of using them, to their bearers. But the genesis of the general conception that came to replace 'descriptivism' as the dominant conception of singular thought and speech need not matter for my purposes here.

¹⁷ For a splendidly clear expression of a view of this kind, keeping the Fregean terminology but abandoning the thesis that *Sinn* determines *Bedeutung*, see McGinn 1982.

John McDowell

5. Now the main interest of Evans's exploitation of Frege lies in its rejecting this picture.

Russell formulates his conception of genuinely subject-predicate form by speaking of propositions in which objects themselves figure as constituents. Propositions so conceived would be individuated by their objects, so that there could not be two such propositions in which the same property is attributed to the same object. So in Russell's version of the conception, there is no room for Fregean *Sinn*. If this is what it takes for thoughts to be object-dependent, thoughts individuated by Fregean *Sinn* would not be objectdependent.

But Evans points out that what is essential to Frege's notion of *Sinn* has no such implication. Frege's controlling idea, as I said, is that if a rational subject can, say, both accept and withhold acceptance from what we would otherwise need to conceive as a single thought, we must find two different thoughts to be the contents of the two attitudes. Nothing prevents this idea from being applied to thoughts that are 'Russellian', in the sense of being dependent for their being available to be thought or expressed on the existence of their objects. So far as singular thoughts are concerned, Frege's aim is to provide for thoughts about objects to be individuated more finely than by the objects they are about, as in Russell's conception. That formulation uses a neutral notion of a thought's being about an object. Nothing prevents its being applied to thoughts that are about objects in the 'Russellian' sense that they depend on the objects for their existence.

From this angle, it looks like a mere mistake on Russell's part to think that genuinely subject-predicate propositions alike in their predicative material must be identical if they concern the same object. This idea may be encouraged, in Russell's own thinking, by the way his doctrine of acquaintance restricts the range of genuinely subject-predicate propositions. Within the restricted range, it is hard to see how there could be pairs of cases like those Frege exploits to argue that we need a finer individuation of thoughts. But once Russell's restriction is lifted, as in the conception of reference that has supplanted 'descriptivism' as the dominant position, it becomes clearer that object-dependence in thoughts is not, of itself, alien to the Fregean framework. There is nothing to prevent us from contemplating rational subjects who combine beliefs and disbeliefs in whose content the same predication is made of the same object—the sort of combination that recommends the Fregean apparatus—even though the contents that Fregean considerations require us to differentiate are contents whose being thinkable at all require the existence of the object. A proposition can be object-dependent even though its identity is determined not by the *Bedeutung* but by the *Sinn* of a singular term used in giving expression to it.

The *Sinn* of a referring expression is the way its *Bedeutung*, the object referred to, figures in thoughts expressible by using it—the way one thinks of the object in thinking such a thought. There can be more than one such way of thinking of a single object, even if each such way of thinking could not be in a thinker's repertoire if the object did not exist. So Fregean fineness of grain simply does not imply object-independence.¹⁸

This opens the way to a satisfying synthesis, which constitutes the context in which Evans considers, for instance, perceptually demonstrative thoughts. The reaction against 'descriptivism' embodies insights about how certain contextual, in particular causal, relations between subjects and objects—which of course have to be actual to stand in those relations to subjects—matter for the directedness of thought at objects. And Evans's point shows that these insights can, after all, be fully integrated with a Fregean stress on rationality, as the frame within which thought about thought and its expression belongs. It need no longer seem that the insights belong with a

¹⁸ Evans actually suggests (1982: 22) that if we gloss Frege's notion of the Sinn of a referring expression in terms of a way of thinking of an object, it is hard to interpret that except as positively requiring that such a way of thinking is object-dependent. And there is something to be said for that claim. (This is why Sainsbury wants to drop the notion of modes of presentation; see n. 14 above.) But the present point is just that nothing in the idea of ways of thinking of objects, conceived as individuated on Fregean principles, excludes their being object-dependent. Can Sinn in general be glossed in terms of the idea of a way of thinking of the associated Bedeutung? This may seem to be threatened by the obvious fact that it cannot be a requirement for understanding a sentence that one know its truth-value. But the 'obvious fact' is just that to understand a sentence one need not know that its truthvalue is true, or that it is false, as the case may be. That does not prevent us from saying that entertaining the Sinn of, say, the sentence 'Kant wrote three Critiques' is a way of thinking of a truth-value (in fact the truth-value true, though one need not know that is a correct identification of the truth-value one is thinking of). In entertaining the Sinn of the sentence, one is thinking of the truth-value as the truth-value thereof that Kant wrote three Critiques. For this locution, see Furth 1968. For the point, made without that locution, see Evans 1982: 17, including the discussion of Dummett in n. 17.

'Photograph Model' of singular thought—a conception according to which what determines which object a thought concerns is at least partly external to its subject's rationality.¹⁹ We can allow for cases in which reference is partly constituted by the obtaining of the sorts of contextual relations that figure in the new, supposedly anti-Fregean way of thinking, without after all needing to separate such reference from the characteristically Fregean topic, configurations within the minds of subjects understood in such a way that the very idea of such configurations belongs in a framework determined by the requirements of rationality.²⁰

Consider uses of demonstrative expressions that single out objects by exploiting their salient availability to perception. Describing the topic like this brings out how the determinate directedness at objects of the thoughts that can be expressed with the help of such expressions depends on contextual, and partly causal, relations between subjects and objects. But even so, 'thoughts' in that formulation can be understood in Frege's way. Fregean fineness of grain, held in place by considerations involving rationality, does not need to be conceived as confined to some inner realm, constitutively independent of those real relations to objects. The real relations to objects do not need to be conceived as extra factors, over and above configurations in that supposed inner realm, in the determination of which objects are spoken of in such utterances and thought of in understanding them. Demonstrative senses can be fully Fregean senses that, precisely because they are partly constituted by real relations to actual objects, reach all the way to the objects.

6. As I remarked, Frege's distinction between *Sinn* and *Bedeutung* is not confined to singular terms. He applies it also to predicates

¹⁹ On the 'Photograph Model', see Evans 1982: 73–9 and ch. 4.

²⁰ The synthesis that Evans enables needs to be separated from the response to the 'new theory of reference' exemplified with great clarity by Searle (1983). Searle's strategy is to incorporate mention of the causal relations that the new thinking makes much of into the specifications by means of which he continues to argue that singular thought singles out its objects—so that the characteristic concerns of the new thinking are swallowed up in a sophisticated version of 'descriptivism'. In Evans's synthesis, it is—as in the new thinking itself—not the concept of those relations that carries thoughts to their objects, but the relations themselves. That is the connection with object-dependence; if there is no object, there is no relation. But there is no longer the suggestion, which Searle admirably reacts against, that determining which object a thought concerns is at most partly taken care of by how a subject's mind is configured.

and, strikingly, to whole sentences. The *Sinn* of a sentence, perhaps on an occasion of utterance, is the thought it can express. Its *Bedeutung* is its truth-value, true or false as the case may be.

Readers are sometimes surprised by this. Surely, they think, the *Bedeutung* of a sentence ought to be a concatenation of the *Bedeutungen* of its significant parts. That is thought to capture an intuitive notion of a state of affairs or situation. So we have a supposed improvement on Frege, in which the *Bedeutung* of a sentence is a situation—conceived as a concatenation of objects and, perhaps, properties—that is actual or not according to whether or not the sentence could be truly asserted. There is a helpfully explicit example of this kind of 'correction' to Frege in Jon Barwise and John Perry's 1983.²¹

Barwise and Perry assume that 'reference' expresses a more or less intuitive semantical idea, and that Frege's use of the concept of Bedeutung is a more or less inept attempt to capture it. What controls the supposedly intuitive idea is a principle of 'the Priority of External Significance', which stands opposed to conceptions that find the significance of an expression in its connection to something in subjects' minds (Barwise and Perry 1983: 42). Frege acknowledges that in the Bedeutung of a sentence, as he conceives it, all specificity is eliminated. All true sentences have the same Bedeutung, the truth-value true, and all false sentences have the same Bedeutung, the truth-value false (Frege 1892: 65). Barwise and Perry take this as an acknowledgement that Frege's apparatus as he uses it cannot accommodate structure in the external significance of language, which they take to be what is supposed to be secured by talk of Bedeutung. Their different account of what is 'referred to' by sentences is supposed to supply sentences with a kind of external significance in which specific structure is not eliminated.

Evidently they assume that Frege cannot preserve specific structure in external significance at the level of *Sinn*. At the level of *Sinn* the specific structure that differentiates the significance of one true sentence from another, or that of one false sentence from another, is not lost. Barwise and Perry must suppose that at the level of *Sinn*

²¹ Especially pp. 20–6. Properties, on any intuitive conception, are no better a fit for what Frege conceives as *Bedeutungen* of predicative expressions than states of affairs, on this kind of conception, are for what he conceives as the *Bedeutungen* of sentences, and the point is connected. But I shall not elaborate it here.

the external significance of language, its directedness at objective reality, is no longer in view.

But this is simply wrong about Sinn. The Sinn expressed in an assertoric utterance is what one says in making the utterance. What one says is, schematically, that things are thus-and-so, and that things are thus-and-so is what is the case, if one's assertion is true. And something that is the case is, in a quite intuitive way of speaking, a state of affairs. So an intuitive notion of states of affairs is perfectly available to Frege, but at the level of Sinn rather than Bedeutung.²² Talking about the significance of sentences at the level of Sinn preserves structural specificity, while in no way retreating from the fact that meaningful utterances are (apart from the special case in which mental realities form the topic) directed at the extra-mental world. Barwise and Perry suggest that Frege is confused in choosing truthvalues over situations as *Bedeutungen* of sentences (1983: 21). But the Priority of External Significance yields no ground for this extraordinary suggestion. The purpose for which they think he should have opted for situations as the Bedeutungen of sentences is served by Sinn as Frege conceives it. The point of crediting sentences with Bedeutung lies elsewhere, in the needs of a semantical account of logical validity. And for that purpose truth-value is just what is required.²³

²² Something that is the case, a fact, is something that can be truly said, or thought, to be the case. A fact is a true thought, in the sense in which a thought is the Sinn expressible by a sentence, perhaps on an occasion. Some may find it wrong to call the conception of states of affairs that I am insisting is available to Frege intuitive; it may seem less than intuitive to say that Hesperus being visible over there (imagine an occasion of utterance) is a different state of affairs from Phosphorus being visible over there. But whatever the status of this conflicting intuition, it is irrelevant to my point against Barwise and Perry. States of affairs individuated according to the Sinne of sentences usable to affirm them are finer-grained than proponents of this intuition would like, but fineness of grain in significance has no tendency to imply that the significance is not external. (This comes out particularly vividly when we see that fineness of grain is consistent with object-dependence in thoughts.) So even if my Fregean states of affairs are too fine-grained for some people's tastes, they are perfectly in line with Barwise and Perry's Principle of External Significance. There is widespread confusion about this, and I am inclined to suspect that it underlies the idea that states of affairs would need to be coarsely individuated.

²³ It is a mistake to approach Frege's general picture of *Bedeutung* in the way Barwise and Perry do, as extrapolating, to other sorts of expression, an intuitive notion of how singular terms relate to the objects they enable us to talk about. It would be better to start by appreciating the point of identifying the *Bedeutung* of a sentence with its truth-value, and to see how what is in fact an approximation to the intuitive notion—though it would be wrong to make too much of this—is yielded by

All this can be said without any special attention to singular terms. Advancing from a thought to its truth-value, which is what we do in judgement on Frege's account, cannot be a step that starts from a configuration that is not world-directed and moves to a stance that is world-directed, though with a loss of structured significance. A thought is already to the effect that things are thusand-so. It does not acquire its bearing on the world when someone affirms it inwardly in judgement or outwardly in assertion. And when we do focus on singular terms, it needs no detail of interpretation to see that for Frege having an object in mind can only be entertaining a thought partly determined by a singular Sinn. There is no need for a further step—advancing from a thought to a truthvalue-in order to arrive at a position in which one's mind is directed towards the associated Bedeutung. One's mind is already directed towards the associated *Bedeutung* just by virtue of entertaining a thought determined by the relevant Sinn. But the point becomes especially vivid with Evans's insight that, consistently with Frege's basic principles, a singular Sinn can actually be object-dependent. Evans's version of Fregean thinking makes it especially clear that a move from Bedeutung to Sinn is not a withdrawal from directedness at extra-mental reality.

The point of Fregean *Sinn* is, as I have said, to provide for a conception of contents that fits smoothly with a conception of what rationality requires of a thinker. Now it needs to be acknowledged that the considerations that are operative here do not make the Fregean apparatus compulsory. We need finer-grained contents if the attitudes we are allowed to work with, in the sorts of case Frege exploits, are restricted to, say, belief, disbelief, and suspended judgement. But we can preserve the rationality of the subjects in such cases even if we equip their attitudes with coarse-grained contents, contents individuated by the objects they concern, as long as we compensate by enriching our inventory of attitudes. Where we find ourselves shaping up to saying that a rational subject both believes and disbelieves the same thing, we can go on identifying the contents of the two attitudes, and protect the subject's rationality by saying that she believes a certain coarse-grained proposition under

a notion of *Bedeutung* for singular terms that is controlled by the requirements of a story according to which, roughly speaking, a singular term's having the *Bedeutung* it has consists in the difference it makes to the *Bedeutung* of a sentence that contains it. See Carl 1994: 116.

John McDowell

one guise and disbelieves it under another.²⁴ This might be seen as a sketch for a notational variant of Frege's proposal.

Some people may be tempted to say that it cannot be a mere notational variant. It is a substantive improvement, because by allowing us to exploit propositions individuated by the objects that are their topics (like Russell, one might say 'having those objects as constituents'), it respects something along the lines of Barwise and Perry's principle of the Priority of External Significance. But what I have urged in this section implies that this supposed ground for preference is illusory. And this becomes especially clear when we take on board Evans's point that a singular *Sinn* can be object-dependent.

7. Frege often says that expressions can lack *Bedeutung* and still have *Sinn*. Singular terms that lack *Bedeutung* are not thereby shown to be without *Sinn*. And sentences containing singular terms without *Bedeutung* themselves have no *Bedeutung*, truth-value, though that does not disqualify them from having *Sinn*, from expressing thoughts. On the face of it, this is incompatible with supposing that genuine subject–predicate form is Russellian. And of course Evans does not deny that Frege says those things. He does not claim to find Russellian doctrine explicitly espoused in Fregean texts, at least after the introduction of *Sinn*. The claim is rather that the best way to work out what Frege means by attributing *Sinn* to singular terms is to align the Fregean apparatus with Russell's conception of genuine subject–predicate form.²⁵

This belongs to a genre of readings in which something a philosopher says is set aside on the ground that doing so yields a better version of his thinking.²⁶ In the simplest variety of such readings, the selection of what to discard reflects only the reader's view of what constitutes a philosophical improvement. In this spirit, someone who thinks that the best account of how thought is directed at objects is given by the 'descriptivism' that finds its inspiration in the Theory of Descriptions might omit, from a reading of

²⁴ For an elaboration of a position on these lines, see Salmon 1986.

²⁵ In Evans's own view, genuine subject–predicate form is not restricted to sentences containing 'Russellian' referring expressions (see Evans 1982: 30–3, 46–51). But he thinks that Frege's conception of singular terms is at bottom Russellian. See n. 2 above, and the text below.

²⁶ For another instance of a 'Fregeanism' of this general kind, in which we allow ourselves to improve on Frege's own presentation of his thought, see Sainsbury 2002.

Russellian thinking about singular reference, the idea of logically proper names, and the conception of genuinely subject-predicate form that belongs with it. The only question that such a reading of Russell could raise would be whether it really opened into a superior account of singular reference. There would be an admitted violence to Russell himself, excused on the ground that it led to better philosophy.

But this is not the sort of line that Evans takes on Frege. Evans insists that Frege's conception of *Sinn* can be understood to fit object-dependent thoughts, and that it stands firm independently of speculation about Frege's own attitude to such an idea. This is worth stressing. In the end it is the idea of object-dependent *Sinn* that matters, not whether it can be attributed to Frege himself. But I want to consider Evans's reading of Frege, and not just the use that Evans makes of Frege.

The ground for going further, and putting the idea of objectdependent singular *Sinn* into a reading of Frege, is not a thesis about genuinely subject-predicate form that is merely imported from outside Frege's thinking. Evans claims that, in spite of Frege's remarks about *Sinn* without *Bedeutung*, the idea of object-dependence as a mark of genuinely subject-predicate thoughts is not alien to Frege's own thinking. There are features of Frege's own thinking that point in this direction.

Before he introduced the idea of *Sinn*, Frege expressed a Russellian line about singular thoughts.²⁷ However, it does not seem to have occurred to him to query the 'traditional' classification of singular terms, which includes definite descriptions. Russell was going to find the 'traditional' classification incompatible with taking singular thoughts to be object-dependent, and he was going to conclude that definite descriptions do not belong in a more principled category of singular terms. It is plausible that the considerations that were going to sway Russell made it uncomfortable for Frege to combine the 'traditional' classification with the Russellian view of singular thoughts, as expressed in those early texts. So it

 27 See the passages from Frege (1969: 174, 160), cited by Evans (1982: 12). (It seems clear that the 'Seventeen Key Sentences' cannot postdate the introduction of *Sinn*, as in the dating by Scholz that Evans mentions as one that would be better for his reading of Frege. Frege's standard line after the introduction of *Sinn* is the one I am considering, that lack of *Bedeutung* on the part of a singular term does not deprive sentences containing it of *Sinn*.)

will have relieved a tension when, after the introduction of *Sinn*, he found himself able to countenance thoughts expressed by sentences containing singular terms that have no *Bedeutung*. But Evans suggests that this apparent relief of a tension can have satisfied Frege only because he did not completely think through what he was doing.

The idea of a thought without a truth-value, which this doctrine commits him to, is problematic by lights that should have been Frege's own. Judging, in Frege's account, is advancing from a thought to the truth-value true. Such an advance is correctly undertaken if the thought is true, incorrect if not. This may seem to allow for cases in which a judgement is incorrect because the thought involved fails to be true, not by being false but by having no truth-value—because an expression of it would contain an expression without *Bedeutung*. But can we really recognize what happens in such a case as judging? Judging is judging something to be so. Supposing there is no condition such that if it is met—if things are indeed so—the judgement is true, how can what we are dealing with be a case of judging? But if there is such a condition, it is either met or not,²⁸ and we are back to thoughts with one or other of the truth-values, true or false.²⁹

²⁸ Vagueness might require a qualification to this. But it would be quite implausible to suggest that emptiness in a singular term induces vagueness in atomic sentences containing it. The difficulty I am raising arises in connection with sentences that concatenate empty singular terms with non-vague predicates. So we can focus on that case, and ignore any extra messiness that vagueness might bring.

²⁹ Sainsbury (2002: 24) enables himself to talk of conditions for something to be the bearer of a name even in a case in which the name has no bearer ('For all x, "Vulcan" refers to x iff x is Vulcan') by employing a free logic, in which atomic sentences with empty names are false. This enables him to see conditions for sentences containing such names to be true ("Vulcan orbits the sun" is true iff Vulcan orbits the sun') as on a par with conditions for sentences containing non-empty names to be true. But this move abandons the link, which is surely central to Fregean thinking, between lack of Bedeutung on the part of a singular term and lack of Bedeutung on the part of (atomic) sentences containing it. Sainsbury's construction precisely equips an atomic sentence containing an empty name with a Bedeutung, the truthvalue false. Frege, for whom such sentences were without Bedeutung, could not have defended the ascription of Sinn to empty singular terms, and consequently to atomic sentences containing them, like this. See Evans 1982: 24-5 on how free logic is alien to Frege's thought. (It is not alien to Evans's own thought; see n. 2 above. For a discussion of how Evans might employ free logic in a semantics suitable to his conception of referring expressions, see Mark Sainsbury's contribution to this volume (Ch. 2).)

As Frege notes, people can be confused into supposing that an expression such as 'the least rapidly convergent series' has a Bedeutung (1892: 58). Such people will endorse sentences containing such expressions. They will take themselves to be able to make judgements expressible by such sentences. But Frege cannot suppose that there is a way things might be such that, if things were that way, such a supposed judgement would be true. If there were, there would be no alternative to saying that things are not that way, so the thought is false. So how does the thought supposedly expressed by such a sentence relate to what such a person is doing when she takes herself to be making such a judgement? In this kind of case, specifying the supposed thought cannot be saving what the person judges, in the sense of specifying how things are if the person is judging truly. No doubt a person in this situation judges that a certain sentence expresses a true thought. But we cannot get down from that metalinguistic level to a specific thought from which the person advances to the truth-value true.³⁰

Perhaps the idea would be that some thoughts are judgeable the ones that do have truth-values—and some not. But what is the unity in the concept of a thought, if thoughts are supposed to come in these two varieties—one such that to specify one of its members is to specify how things are according to it, the other not? If it does not belong to a thought as such to be judgeable, Frege's conception of judgement as advancing from a thought to a truth-value lacks the straightforwardness it seemed to have. And we still have the difficulty of saying what happens when someone accepts a sentence that, on this way of talking, expresses a thought that is not suited

³⁰ This is a version of Evans's argument at 1982: 24. Carl (1994: 24) says that Evans identifies thinking with judging (forming a belief), thus losing contact with the indispensable Fregean idea of something from which one needs to advance in order to make a judgement. (See also 1994: 180 n. 26.) But what Evans identifies with forming a belief is of course not merely entertaining a thought, but *accepting* one; Carl's criticism rests on a crude misreading. Carl goes on to suggest, if I understand him, that Frege's mature conception of judgement, as an advance from a thought to a truth-value, positively requires the idea that sentential *Sinne*, thoughts, may or may not be associated with sentential *Bedeutungen*, truth-values. (See, e.g., p. 117.) But this seems obviously wrong. The thought from which one advances, when one judges truly or falsely, must already have its truth-value—true or false according to whether the judgement is true or false. It does not acquire its truth-value by means of the advance. Carl fails to engage with the difficulty of applying Frege's conception when someone thinks she can make such an advance in connection with a sentence that has no *Bedeutung*.

John McDowell

for judgement because it lacks a truth-value. It is not a case of accepting a thought, in the sense in which specifying a thought that someone accepts is specifying how things are thought by her to be. But surely that is what accepting a thought is, not just one interpretation of 'accepting a thought'.

We can find a clue to how Frege thinks about all this in the fact that he is happy to describe such cases in terms of unwittingly straying into the sphere of fiction.³¹ Of course, if 'fiction' is just a label for the supposed realm of thoughts that lack truth-value, this takes us no further. But in one place in the Nachlass where Frege talks head-on about fiction, he shows himself tempted by a certain conception of, say, story-telling, which indeed has its attractions. According to this conception, a story-teller makes a use of language modelled on, and imitative of, the making of factual claims, so as to give the appearance of making assertions without actually doing so. And Frege suggests that when he talks of expressing thoughts in fictional discourse, this relates to the thoughts that figure in assertions and judgements only as talk of story-tellers as making assertions would relate to assertions that are expressive of judgement.³² The cleanest way to capture what Frege seems to be driving at here is to say that there is really only one kind of thoughts, those such that to specify one is to specify how things are according to it, so that it is true if things are that way, and false otherwise. Where he allows himself to talk as if there is another kind of thought, whose members lack truth-value, what he really has in mind-at least at some level of his thinking, which never quite surfaces into explicitness-is a kind of case in which someone merely goes through the motions of expressing a thought in the only sense Frege is really committed to, according to which it is part of the very idea of a thought that there is such a thing as how things are according to it. Merely going through the motions is something people do intentionally in fiction, strictly so called, and unintentionally in the sort of case we are considering.³³

³¹ See the passage from 'Der Gedanke' that Evans quotes (1982: 28). This belongs with the fact that Frege standardly illustrates his *Sinn*-without-*Bedeutung* claims with examples from fiction, ordinarily so called.

³² See the passage from Frege (1969) that Evans cites (1982: 29).

³³ Commentators have made great efforts to resist the drift of this passage from Frege. David Bell (1990*a*: 273) claims that what Frege calls *Scheingedanken* are thoughts that 'only aim to convey appearance'. Sainsbury acknowledges that *Scheingedanken* figure here alongside *Scheinbehauptungen*, and that *Scheinbehauptungen* are performances that give the appearance of being assertions Evans's Frege

This suggests that when Frege exploits the separation of *Sinn* from *Bedeutung*, thought from truth-value, to alleviate the discomfort that will have resulted from combining his early 'Russellian' conception of subject–predicate form with the 'traditional' grouping of singular terms, this is not the complete breach with his early 'Russellian' way of thinking that it might seem, and perhaps seemed even to Frege himself. In exploiting the idea of an unwitting lapse into fiction, he has confused himself. A conception that would be most cleanly put by talking of a mere appearance of *Sinn* has presented itself to him as if it allowed a kind of genuine *Sinn* that, in connection with indicative sentences, cuts loose from possession of a truth-value. If we pay attention to the details of how Frege thinks about fiction, we can find, even in the places where he explicitly rejects the 'Russellian' line, traces of a continuing pull towards it, which he may have concealed even from himself.

As Evans puts it, Frege's remarks about fiction suggest that

we may gloss those passages in which Frege says that a sentence containing an empty singular term may express a thought as follows. Yes: a sentence containing an empty singular term may have a sense, in that it does not necessarily have to be likened to a sentence containing a nonsense-word. But no: it does not *really* have a sense of the kind possessed by ordinary atomic sentences, because it does not function properly, it is only *as if* it functions properly. 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 might speak of apparent, rather than mock or pretend, thoughts. (Evans 1982: 30)³⁴

As Evans acknowledges, this reading 'does not present Frege in a very good light' (1982: 28). That may seem to be the beginning of a principle-of-charity case against it.³⁵

(*Behauptungen*) but are not (2002: 11–12). But he strains to avoid the parallel interpretation of *Scheingedanken*. He suggests, in effect, that the prefix *Schein* here means 'not to be taken seriously', and whereas 'assertions' that are not to be taken seriously are not really assertions, thoughts that are not to be taken seriously can be really thoughts. But, as against both Bell and Sainsbury, it seems plain that the force of the prefix *Schein* is something like '(merely) apparent'.

³⁴ 'Mock thoughts' is the wording in the translation Evans cites; but 'apparent thoughts' (which does not embody the implication of pretence) would be a perfectly good translation of Frege's *Scheingedanken*.

³⁵ For the suggestion that Evans's Frege is not the 'profound and powerful thinker' he is usually taken to be, see Bell 1990*a*: 276.

But when Evans suggests that Frege's early 'Russellianism' is submerged, rather than definitively discarded, by his allowing himself Sinn without Bedeutung, it does not constitute a response merely to elaborate the characteristic Fregean claim to be concerned only with sentences usable in 'science', only in reasoning directed at achieving knowledge.³⁶ Evans does not miss or ignore that Fregean intention. The trouble he fixes on comes from Frege's proceeding as if a concept for which he gives a proper explanation only in the context of the 'scientific' use of language, the concept of thoughts, is straightforwardly available for talking about cases in which people falsely suppose that sentences are suitable for 'scientific' use.³⁷ And there is the discussion of fiction that Evans exploits, which strongly suggests that the underlying conception, never brought to explicitness by Frege, is that the availability of the concept in those cases is a mere appearance, part of the content of the false supposition.

If Evans's Frege is Frege, of course it would have been better if he had made the submerged thought explicit, to himself and to his readers. Evans's Frege is a thinker one of whose insights he himself

³⁶ That is what Carl (1994) does. Bell too takes a version of this line. He scores some telling points against Evans's tendency to assimilate Frege's semantical interests to those of, say, Davidson (though it does not inspire confidence when Bell substitutes 'explicit' for 'implicit' in quoting Evans's claim—which I agree is over the top—that Frege's mature conception of reference is the semantics that is implicit in his earlier works (1990a: 269)). But in the end the credentials of Evans's suggestion of a submerged 'Russellianism' are not affected. The alternative picture of Frege's programme that Bell outlines (1990a: 276–7) does not remove the difficulty that Evans finds in Frege's official *Sinn*-without-*Bedeutung* doctrine, and Evans's suggestion about what underlies the doctrine loses no plausibility by being put in the context of Bell's more accurate account of Frege's general purposes.

³⁷ The explanation is that to determine a thought is to determine a condition for its truth. Defending the idea that singular terms can have *Sinn* without *Bedeutung*, Sainsbury (2002: 208) suggests that the natural counterpart to the idea of determining a condition for a sentence to be true is the idea of determining a condition for a term to refer—a condition that may or may not be satisfied, without threat to its affording a specification of the term's *Sinn*. But a condition for a sentence to be true is not a condition for it to refer *überhaupt*, but a condition for it to have one rather than the other of the *Bedeutungen* that sentences can have. Moreover, the connection of sentential *Sinn* with conditions—for truth in particular—is appropriate precisely because of the special linguistic character of sentences: they enable speech acts whose content is *that things are thus-and-so*, and specifying an instance of that form is specifying a condition for a sentence to be true. It is of course not to be expected that there should be a parallel to this for subsentential expressions. Sainsbury's analogy is not natural at all.

Evans's Frege

only gropes at. But this imputation of unclarity about his own thinking cannot tell against the reading, on principle-of-charity grounds, unless not making it would put Frege in a better light than crediting him with at least groping towards an insight. This depends on the magnitude of the insight, and it would be hard to overestimate that. I gestured towards this earlier by saying that Evans's Frege enables a synthesis between acknowledging that contextual relations between subjects and objects matter for determining the contents of thoughts, on the one hand, and giving full weight to the idea that thinking is an exercise of rationality, on the other. This can be seen as a substantial contribution to a project that goes back at least to Kant, and that is beset with difficulties in the intellectual environment of modern philosophy: integrating our rational powers with our natural situatedness in the world.

Names in Free Logical Truth Theory R. M. SAINSBURY

It is...an immediate consequence of recognizing names like 'Julius' in a language that classical logic must be modified.

Evans 1982: 36

I

Evans believed that natural languages do or could contain more than one kind of proper name: Russellian names, whose determining characteristic is that to suppose that they have no bearer is to suppose that they have no significance, and descriptive names, whose determining characteristic is that they are introduced by means of a definite description which fixes their reference and which, at least in the early stages, constitutes their 'content'. In contrast to a Russellian name, a descriptive name may be intelligible while lacking a bearer. From the point of view of systematic semantic theory, Evans thought that a truth theory could do justice to these different kinds of names.¹ A Russellian name would be given an axiom like

1. 'Hesperus' refers to Hesperus,

whereas a descriptive name would be given an axiom like

2. for all x ('Julius' refers to x iff x =Julius) (Evans 1979: 184; 1982: 50).

¹ His preferred form of semantic theory, what he calls 'interpretational semantics' (Evans 1976), has a somewhat richer structure, but this makes no difference to the present discussion.

We can imagine 'Julius' to have been introduced by a stipulation of the following kind:

Let us use 'Julius' to refer to whoever invented the zip (Evans 1979: 181).

There might have been no such person: the zip might have been invented by a committee or might have been a natural phenomenon. These possibilities are not inconsistent with the intelligibility of 'Julius'. Axiom (2) has to be understood within negative free logic, whose main features are that every simple sentence containing a name with no bearer is false, and that the quantifier rules are restricted.² Evans cites Burge (1974) as a source for free logical truth theories.³ In classical logic, (2) entails that Julius exists (there is an x such that x = Julius), but this inference is not valid within negative free logic. This allows (2) to be true even if 'Julius' has no bearer. In that case, 'x = Julius' is false of everything, so if the biconditional is true of everything, "Julius" refers to x' must be false of everything, which is as it should be. The truth theory as a whole has a single logic, so axioms like (1) will coexist with axioms like (2) within the free logical structure. Even within free logic the truth of (1) ensures that Hesperus exists, for (1) is a simple sentence, one whose truth requires that its names have bearers. So embedding (1) within a free logic does not weaken its logical powers. Both kinds of axiom can lead to homophonic theorems such as

- 3. 'Hesperus is visible' is true iff Hesperus is visible
- 4. 'Julius is tall' is true iff Julius is tall.⁴

In the free logical framework, neither theorem entails that the name has a bearer. Given the restrictions on existential generalization, we can no more infer from (3) that something is Hesperus than we can infer from (4) that something is Julius. So although the free logical framework does not impair the logical powers of axioms for

 2 A 'simple' sentence is one consisting of a non-complex *n*-place predicate concatenated with *n* referring expressions. If all referring expressions were simple, simple sentences would be atomic sentences. Free logicians wish to avoid excluding complex referring expressions (e.g. compound names like 'Plato and Aristotle' and definite descriptions), so they must make room for the possibility of non-atomic simple sentences.

³ Evans's commitment to the viability of free logic appears elsewhere, e.g. in his discussion of E-type pronouns (Evans 1977: 124–5) and in his discussion of the contingent a priori (Evans 1979: 195).

⁴ For details of free logical derivations, see Larson and Segal 1995.

Russellian names, it does impair the logical powers of the relevant theorems.

There is a reason for thinking that this fact prevents a truth theory from doing justice to Evans's idea. He is committed to the following views: there are or could be both descriptive and Russellian names in a language; these are semantically very different; and interpretation relies upon the appropriate T-theorems. These views are in tension with the fact that the T-theorems do not differentiate between Russellian and descriptive names. In \$\$=-3 I develop this objection and show how, on Evans's behalf, it can be met. In \$4 I endeavour to explain the origins of Evans's view that no names are both non-descriptive and non-Russellian.

Π

Setting indexicality to one side, a truth theory 'serves as' a theory of meaning if following this procedure leads to the correct interpretation of an arbitrary declarative utterance *u*: first, identify a sentential type, s, to which u belongs, then deduce a T-theorem of the form 's is true iff p'; finally, conclude that the speaker, in uttering u, said that p. If you encounter an utterance of 'Snow is white', the procedure directs you to identify the sentence uttered (namely, 'Snow is white'), use the axioms of your truth theory to derive the T-theorem "Snow is white" is true iff snow is white', and derive the lemma that the utterance in question is true iff snow is white. Given what you know about the truth theory (for example, about how it was grounded in the evidence of the speaker's usage), you are then entitled to the final conclusion that the speaker said that snow is white. This is your interpretation of the utterance. Not all true truth theories will lead to correct interpretations. For example, a truth theory which uses the following axiom.

5. 'Hesperus' refers to Phosphorus

rather than (1) will be no less true than one which uses (1), but it would enable one to derive the T-theorem '"Hesperus is visible" is true iff Phosphorus is visible', which, though true, would wrongly endorse the conclusion that one who has uttered 'Hesperus is visible' has said that Phosphorus is visible. (Evans would certainly accept the characterization of this conclusion as wrong, though no

doubt this is controversial.) In a form of words of which Evans approves: though neither (1) nor (5) *states* the sense of 'Hesperus', (1) *shows* it, whereas (5) does not.

In using a truth theory as a theory of meaning, interpretation arises from what is on the right-hand side of the biconditional of a T-theorem. In this classic picture, which Evans endorses, no distinctions of scope regarding names will make for a general difference in the interpretation of Russellian as opposed to descriptive names. For simple sentences, there is no scope difference for either kind of name. In complex sentences, for example ones containing negation, a user of a descriptive name could legitimately contrast, for example,

6. [Julius] Julius is not tall

with

7. not [Julius] Julius is tall

(using the scope-indicating device Evans offers, borrowed from Principia Mathematica⁵). (6) is true only if Julius exists, whereas (7) would be true if 'Julius is tall' is false because Julius does not exist. In either case, the conclusion reached by applying the theory in interpretation is not one that entails the existence of Julius. In the ordinary kind of T-theorem, (6) will feature as the right-hand side of a biconditional. If there is no such person as Julius, it will be false, so, assuming the truth of the T-theory, the left-hand side is false, which is as it should be. We do not have the resources to generate a T-sentence which is true only if a name contained in it refers, so we cannot generate a T-sentence with a profile appropriate to a Russellian name. Likewise, from the fact that the speaker said that [Julius] Julius is not tall, we cannot infer that Julius exists. If a name's Russellian status is marked in object language logical forms by the award of widest scope, then, for such names, we will find only object language sentences like (6), but the requirements of interpretation will not enable us to infer that, for example, Hesperus exists. Although this conclusion is available in the T-theory as a whole, thanks to axioms like (1), it is not made available by the only theorems in the theory that are used in interpretation, the

⁵ Compare *Principia Mathematica* (Russell and Whitehead 1910–13), 69. If A is a formula, so is $[n_1 \dots n_k]A$, for any names, $n_1 \dots n_k$. Occurrences of n_i within A are bound by the scope-marker. The truth of $[n_1 \dots n_k]A$ requires that every name bound by the scope-marker have a bearer; failing that, the sentence is false.

R. M. Sainsbury

T-theorems. The existence assumption characteristic of a Russellian name does not surface in interpretation. If an utterance is interpreted as containing a name having wide scope, requiring it to have a bearer if the utterance is to be true, this does not reveal whether this is because the name is Russellian or because it is descriptive but used with wide scope; if the latter is the case, the interpretation can be correct (and the T-theorem true) even if the name is empty.

Someone of Evans's persuasion may well believe that Russellian names sustain a principle which one might call 'exportation from within indirect contexts'. The principle would license the inference from 'Jack said that Hesperus is visible' to 'Hesperus is such that Jack said that it [Hesperus] is visible'. There is no immediate reason why this principle should not hold even for contexts for which there is no guarantee that substitution of coreferring names preserves truth. Within the classical methodology of truth theories, however, one cannot make use of the principle to mark a difference between the roles played by Russellian and by descriptive names. The pattern of reasoning

```
someone said that [n] \dots n \dots; so [n] someone said that \dots n \dots
```

is not generally valid if descriptive names are in the language, since the premiss does not require the name to have a bearer, whereas the conclusion does; and whatever kind of name is used in the object language, the standard way of using a truth theory in interpretation delivers no more than the premiss.

If there were nothing more to say, the right conclusion would be, I think, that the weaker logic required by descriptive names has in effect emasculated Russellian names, making their role in interpretation indistinguishable from that of descriptive names with wide scope. This was far from the position that Evans wished to hold.

III

In the classical methodology of truth theories the content made available for interpretation is the complete and self-standing content occupying the p position in the schema 's is true iff p'. This simple idea needs to be refined in order to deal with demonstratives. Once this refinement is made, for reasons independent of the present discussion, one can see how a similar refinement could make room for a systematic interpretive and T-theoretic difference between Russellian and descriptive names.

Davidson has suggested that the form of theorems for sentences containing such expressions as 'that book' would be along the following lines:

'That book was stolen' is true as (potentially) spoken by p at t if and only if the book demonstrated by p at t is stolen prior to t. (Davidson 1967: 34)

An instance is:

'That book was stolen' is true as (potentially) spoken by Davidson at noon on 02/02/02 if and only if the book demonstrated by Davidson at noon on 02/02/02 is stolen prior to noon on 02/02/02.

Davidson comments that 'there is no suggestion that "the book demonstrated by the speaker" can be substituted ubiquitously for "that book" salva veritate' (Davidson 1967: 34). But something more damaging to his project is the case: the instance cannot be used in interpretation. Intuitively, it is never correct to report an utterance by Davidson of 'That book was stolen' as his having said that the book demonstrated by Davidson at noon on 02/02/02 is stolen prior to noon on 02/02/02. What Davidson actually said by those words does not involve the concept of demonstration, or the conceptual apparatus used to specify dates. I think that the only way to achieve a correct description of what is said in such a case involves two parts. First, the scene is set by some such remark as 'Pointing to a copy of Word and Object at noon on 02/02/02 ...'. This gives the necessary background for the interpretation of the demonstrative phrase and the tense of the verb. Next, the content is specified relative to that background: Davidson said that it [that book] had been stolen. Putting the two parts together we have the following intuitively correct report:

Pointing to a copy of *Word and Object* at noon on 02/02/02, Davidson said that it had been stolen.

The reporter can use any devices and concepts he wishes in order to set the scene, even ones, as in this case a date and a book title, that were not overtly exploited by the speaker; these then bind the use of pronouns or tenses in the report, ensuring that they have the correct referent, without suggesting that the original speaker himself

R. M. Sainsbury

exploited them. The upshot is that there may be no self-standing sentence, corresponding to the position occupied by p in 's is true iff p', that the reporter can use to express the content of the speech he correctly reports. Success is achieved in the more complex twopart way.⁶ By the same token, the truth-theoretic methodology must be modified. T-theorems usable by an interpreter will need to quantify over utterances rather than merely sentences, and should make room for the two-part form, as in this schema:

8. In uttering *u* the speaker referred to some object *x* and time *t*, and *u* is true iff...x dots f(t).

User instructions say something like: find out to what objects and times the speaker referred, and construct a report by first referring to these in your own preferred way, and, applying (8), make those referring expressions govern what replaces 'x' and 'f(t)' as they occur after 'iff'.⁷

We can use a structurally similar device to secure a distinction, within this refined truth-theoretic methodology, between Russellian and descriptive names. The idea is to have two kinds of T-theorem, and associate each with different instructions for those using a T-theory as a theory of meaning. For Russellian names, the T-theorems might follow the pattern:

9. [Hesperus] 'Hesperus is visible' is true iff Hesperus is visible,

whereas those for descriptive names would follow the weaker pattern:

10. 'Julius is tall' is true iff [Julius] Julius is tall.

User instructions for truth theories would be sensitive to this difference of scope. If the theorem contains a name n with widest scope, as in (9), then the instruction would be to use the (translation of the) name in reporting what a speaker said, using a report along the following lines: '[n] the utterer said that...n...', where what fills in the dots comes from what is to the right of the biconditional in the theorem. If the theorem does not contain a name with widest scope, as in (10), then the user instructions should tell us to accept as a conclusion simply: the utterer said that $[n] \dots n \dots$. The idea is

⁶ Cf. Sainsbury 2002: 137-49.

 $^{^7}$ The treatment of tense is obviously grossly oversimplified, though in ways irrelevant to the present discussion.

to mark a Russellian name by awarding it wide scope not merely in the object language sentence, but in the whole of any T-theorem containing a Russellian name, and thus in the proper expression of a metalanguage report of what was said in the object language. By contrast, descriptive names always take less than widest scope in the T-theorems, however wide their scope may be in the object language sentence the theorem addresses. The Russellian/descriptive contrast is represented by wide/narrow T-theoretic scope. The above discussion of demonstratives is supposed to suggest that there is no special structural pleading here: it may not be possible to contain within the 'p' position of a classical truth theory the full content of every utterance. Interpretation may require a two-part structure, in which the scene is first set by introducing some entities by expressions upon which the specification of the content is dependent.

If this solution can be made to work, the virility of Russellian names is restored. The existence-requiring axioms now transfer this existence requirement to the T-theorems. In considering how the solution could be implemented within Evans's approach, we must remember that, according to him, there is a single relation of reference that is absolute as opposed to world-relative. If there are both Russellian and descriptive names, both are relata of this single absolute relation of reference. The upshot is, as I will show, that the representation of the Russellian/descriptive distinction in terms of T-sentential scope requires further enrichment of the T-theory.

The compositional axiom for simple sentences states in general how their truth conditions depend upon the contribution of their parts. Evans formulates it as follows:

(P) If S is an atomic sentence in which the *n*-place concept expression R is combined with *n* singular terms t_1, \ldots then S is true iff <the referent of t_1, \ldots , the referent of $t_n >$ satisfies R. (Evans 1982: 49; cf. 1979: 184)⁸

He comments: '(P) simultaneously and implicitly defines reference and satisfaction in terms of truth' (Evans 1979: 184; 1982: 49). Even if there is more that can be said to characterize these notions (in terms of how they connect with speakers' intentions, for example),

⁸ Talk of atomic sentences excludes by *fiat* complex singular terms (which may contain logical connectives and other symbols). In the subsequent reformulation, I tacitly drop the requirement.

R. M. Sainsbury

most people will agree that reference and satisfaction have to make (P) true. As part of the metalanguage, (P) itself is governed by negative free logic. Its definite descriptions ('the referent of t_i ', etc.) can be understood either as complex referring expressions or as Russellian quantifier phrases, and in either case a question of scope could arise. Using the referential option, the definite descriptions could take wide scope relative to the biconditional:

[the referent of t_1]... [the referent of t_n] if *S* is a sentence in which the *n*-place concept expression *R* is combined with *n* singular terms t_1, \ldots, t_n , then *S* is true iff < the referent of $t_1 \ldots$, the referent of $t_n >$ satisfies *R*.

Given that, according to Evans, 'Julius' refers to its bearer, if any, 'in exactly the same sense as that in which a Russellian name refers to its bearer' (Evans 1982: 50), (P), if correct, must hold for both Russellian and descriptive names. The wide scope version clearly cannot hold for an empty descriptive name. Hence the definite descriptions in (P) must always take narrow scope. A more accurate version of (P) is:

(P') If S is a sentence in which the *n*-place concept expression R is combined with *n* singular terms t_1, \ldots, t_n then S is true iff [<the referent of t_1, \ldots , the referent of $t_n >$] <the referent of $t_1 > \ldots$, the referent of $t_n >$] <the referent of $t_n >$ satisfies R.

With this kind of principle governing compositionality, one cannot derive the envisaged kind of wide scope theorems for Russellian names. There is simply nothing in the T-theory to legitimate the derivation of a theorem with a name taking wide scope over 'is true iff'.

In keeping with the significance of scope at the T-sentential level, the present proposal can reflect the difference between an axiom for a Russellian name and one for a descriptive name in terms of scope. With the scope indicators in place, (2) above becomes:

11. for all x ('Julius' refers to x iff [Julius] x = Julius) (cf. Evans 1982: 50).

Given the natural free logical rules for manipulating scope indicators, the following is logically equivalent to (1):

12. [Hesperus] for all x ('Hesperus' refers to x iff x = Hesperus).

74

To use these features of the axioms to generate the required scope distinctions among the T-sentences requires some further rule of inference, perhaps one formulated along these lines:

13. if 's is true iff p' is a theorem, then so is $[n_1 \dots n_k]$ s is true iff p', where $n_1 \dots n_k$ are (the translations into the metalanguage of) names contained in 's' which are given widest scope in their axioms.

If there are no names in 's' which are given widest scope in their axioms—that is, if 's' contains no names or only descriptive ones—then the scope marker $[n_1 \dots n_k]$ has no effect. The special rule of inference (13) ensures that only Russellian names allow interpreters to interpret on the pattern: [Hesperus] the speaker said that Hesperus was...; or, more idiomatically: concerning Hesperus, the speaker said that it was.... Descriptive names would always take narrower scope than the T-sentential biconditional, and so would permit at most interpretations like: the speaker said that [Julius] Julius was....

This rule of inference (13) is clearly not a valid logical rule in the usual sense, since it identifies a suitable premiss not merely in terms of how it is built up out of logical operators and schemata, but also in terms of an occurrence of the non-logical and non-schematic 'is true'. The classical idea behind using a truth theory as a theory of meaning is to have axioms that state all the semantic information. This completeness is supposed to be guaranteed by the fact that theorems which suffice for interpretation follow 'by logic alone' from the axioms (cf. Evans 1975: 26, where 'deductive entailment' introduces the relevant notion). It is no minor adjustment to the methodology of truth theories to allow non-logical rules of inference. However, it is an adjustment which is arguably kin to one needed to accommodate indexicals. A semantic theory could not hope to incorporate sufficient information for interpreting an utterance containing an indexical. An interpreter must bring to bear not only the kind of information which could be recorded in a semantic theory, but also such empirical details as who the utterer was, what the time of the utterance was, and what objects were referred to in the utterance. This information could not be stated in any detail in a semantic theory. Even in the simplest kind of case, like the use of 'I', for which a semantic theory might supply the general information that an utterance of 'I' refers to the utterer, an interpreter must still have independent knowledge of who the utterer is in order to arrive at an interpretation, and this knowledge is neither supplied by the semantic theory, nor purely logical. This is more conspicuous in other cases—for example 'that', where a large range of cognitive skills, sensitive to various contextual factors, may be required for an interpreter to decide what its referent is. There are generalities to be had, of course, but they could not have the detail needed to resolve all particular cases. That a semantic theory should provide all the information needed for interpretation is best seen as a kind of limiting ideal; it is not one to which we can reasonably aspire for natural languages.

If this much is accepted, then the rule of inference (13) can claim a right to appear in a truth theory, despite its non-logical character. It is there to steer the interpretation of a proper subset of the names in a language in an appropriate direction. The basic fact about each such name is recorded in its axiom, in which there is a wide scope metalanguage translation for it; (13) shows how this fact is to impinge upon interpretive practice. There is thus a case for the view that there is no irresoluble tension between suitably refined truththeoretic aspirations and Evans's idea that a language could contain both Russellian and descriptive names, the latter enforcing a free logic upon both object language and metalanguage. The difference shows up in the way in which the theory reflects the fact that, for Russellian names, the existence of a bearer is guaranteed. Let us say that an 'interpretive' theorem is one that may be truly prefixed by something like 'the following sentence states a fact which is guaranteed by meaning or logic alone'.9 In a correct semantic theory, all the axioms are interpretive in this sense. Whether or not all the theorems are, depends on the logic of the truth theory.10 Existence claims for Russellian names will be interpretive. On reasonable versions of free logic, from something of the form '[n]...' one can infer: ' $\exists x \ x = n$ '. The axiomatic status of (12) will ensure the theoremhood (within the free logical truth theory) of ' $\exists x \ x =$ Hesperus', and this will count as interpretive. It is hard to see how there could be corresponding theorems for (even

⁹ This formulation is intended to secure acceptance from those who do not think that logical axioms are guaranteed by the meanings of the logical constants.

¹⁰ The usual assumption is that it will not hold in general, since there will normally be theorems 'p iff q', interpretive theorems 's is true iff p' and uninterpretive ones 's is true iff q'. However, Larson and Segal (1995) have proposed that, given a sufficiently weak version of free logic, every theorem can be interpretive.

non-empty) descriptive names, but if there were, they would not be interpretive.

In summary, with suitable adjustments, truth theory can make room for the two different kinds of name, Russellian and descriptive, and accord them suitably distinct roles in interpretation. My view is that natural languages do not and could not contain either descriptive or Russellian names. This cannot be argued here.¹¹ Rather, in the section that follows, I will try to uncover the source of Evans's contrary opinion.

IV

A common view has been that whereas Evans was right to recognize Russellian names, he was wrong to make room for descriptive ones. A name may be introduced by means of a description, but this does not guarantee that its subsequent use is correct only if it conforms to the description. Evans was well aware of this. For example, he writes:

For present purposes, it is not necessary to concern ourselves with the situation that would arise if the name [Julius] became associated with other predicates as a result of discoveries made using the stipulation [that 'Julius' is to be used for whoever invented the zip]. We need only consider the simple case—the initial period during which the name is unquestionably a 'one-criterion' name. (Evans 1979: 181)

Evans is aware that once the name passes into general currency, there can be no guarantee that the initial reference-fixing description will continue to play its original role. Information accumulates around the name. There may be no mechanism that ensures that some of it will retain the privileged position that it must have in order for it to be a descriptive name. A name like 'Deep Throat' may retain its connection with some description like 'the source of Woodward and Bernstein's Watergate-related information', perhaps in virtue of the fact that all uses of the name defer to Woodward and Bernstein's. Even in such a case, the connection might come loose. It is a coherent supposition that Deep Throat did not provide all the information that Woodward and Bernstein reported. One could envisage developments of this suspicion properly

¹¹ Sainsbury (2005) tries to justify this claim.

R. M. Sainsbury

describable as the suspicion that none of their information came from Deep Throat. This does not appear to be a manifest contradiction, as it would be if 'Deep Throat' retained its meaning and its envisaged descriptive connection. Likewise, given the passage of time, it will be in order for someone to speculate without contradiction that perhaps Julius did not really invent the zip, but merely posed as the inventor; and then for people to speculate that he did not even do that, and that the attribution to him of the invention was a result of a historian's mistranscription.¹² Nothing the person who introduced the name can do could prevent such developments. Were they to occur, we would have a name which is neither descriptive nor Russellian: not descriptive, since there is no description determining which object is the referent of the name (cf. Evans 1979: 180); and not Russellian, since neither the way in which the name was introduced, nor anything in its subsequent history, could guarantee that it had a bearer. This category of names, the nondescriptive and non-Russellian, is, I believe, precisely the one to which most natural language names should be assigned. They are often introduced by description, in a way that does not guarantee that they have a bearer; yet what it is irresistibly natural to classify as the same name, with its meaning unchanged, may cease to be associated with the original description, whether as determiner of reference or as determiner of what an interpreter needs to know.

Rather than defend the claim that most natural language names are neither Russellian nor descriptive, I wish to speculate about why, among so many varieties of singular term that Evans discusses, he makes no room for the category of non-descriptive and non-Russellian. One explanation that applies to many theorists does not apply to Evans: in the grip of the assumption that classical logic supplies all the necessary semantic categories for a proper description of natural language, singular terms can correspond either to individual constants, and so be Russellian names, or to definite descriptions (understood in Russell's way), and there are no other options. Negative free logic breaks this classical strait-jacket, making room for non-descriptive yet non-Russellian singular terms. Since Evans was familiar with, and sympathetic to, negative

¹² This view of the relation between names and descriptions resembles one of two apparently opposed views discussed by Searle (1958), whose main feature is, I think, supposed to be retained in his eventual synthesis.

free logic, this line of thought provides no explanation for his failure to recognize this semantic category.

A better explanation attends to the means which Evans thought were available for explaining how various uses of a given name are unified. They could be unified by an object to which all uses relate, or by a description which all uses invoke; no third option is recognized, a restriction which in turn (I will argue) is to be accounted for by a residual internalism in Evans's thought. The demand for unification is at work in his discussion of a remark by David Wiggins:

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. (Wiggins 1975: 11)

Evans comments:

The view expressed by Wiggins leads to the claim that ordinary proper names are Russellian.... There must be something that unifies them [*sc*. the 'countless different descriptions'], and this can only be the fact that they are all ways of identifying the same object....But, given the close semantical connection between a descriptive name and a description, no such problem [of unification] arises. (Evans 1979: 181)

Evans considers an alternative way in which uses might be unified: by the *source*, in the 'journalistic sense', of the invoked information. He points out that even if having a source in common provides a sufficient condition for understanding, it does not provide a necessary one: in this sense of 'source', a hearer may correctly interpret a speaker even if there is no single source responsible for both speaker's and hearer's information. The reader is left to conclude that there are no appropriate further ways of providing necessary and sufficient unifying conditions for understanding, so that the Russellian status of all non-descriptive names is secured.

As Evans was aware, there are alternatives: for example, those based on the idea that, at least once all mundane issues are set aside (for example, that the hearer has heard the speaker correctly), it is necessary and sufficient for understanding that the use made of the name by the speaker should belong to the same name-using practice as the use brought to bear by the hearer in the act of putative understanding. There is more than one way to say what it is for two uses of a name to belong to the same name-using practice.

R. M. Sainsbury

Evans in another context discusses the Kripkean 'Causal Theory', the view that 'the denotation of a name in a community is...to be found by tracing a causal chain of reference preserving links back to some item' (Evans 1973: 10).¹³ Bearing in mind that the same name (as normally individuated) can be used of many different things, we can re-express this as a view about the individuation of name-using practices: a use belongs to a single practice, P, iff either it is the event which introduced the name of the referent of the practice, and so inaugurated that practice, or else descends from that introduction by a series of links which preserve reference. Although this Kripkean position is consonant with the view that names are Russellian, it can be modified into a view which allows non-descriptive names to be empty by removing mention of the referent: a use belongs to a single practice, P, iff either it is the event which introduced the name, and so inaugurated that practice, or else it descends from that introduction by a series of links which would preserve reference (if any). If some account of this sort is correct, the various uses of a name in which various different descriptions are brought to bear can be unified without reliance upon an object to which all relate: they all belong to a single nameusing practice. We will see that Evans's residual internalism plays a part in his rejection of views of this kind.

In the earlier paper (1973), though not in the book (1982), Evans offers arguments against such theories. For example, he says, 'Change of denotation is...decisive against the Causal Theory of Names' (Evans 1973: 11). I have suggested elsewhere (Sainsbury 2002: 220–2) that this is not in fact decisive: an alternative, and indeed preferable, description of such cases is not that there is a single practice with first one and then another referent, but that there are successive practices each with a distinct referent. His other attacks on the Causal Theory are similarly inconclusive.

¹³ The usual caveat about the supposition that this is a theory that Kripke did or would endorse needs to be entered. Evans's formulation, perhaps unintentionally, suggests that the Causal Theory would offer a way of in practice telling which object the bearer of a name is. An alternative view is that the theory might offer an account of what makes something the referent of a name without suggesting that in practice we should find out which object the referent is by applying the theory. The fact that we use associated descriptions as evidence, in determining to which practice a use of a name belongs, may wrongly have given encouragement to descriptive theories of what names mean. For example, a Kripkean can endorse Evans's view that 'There is something absurd in supposing that the intended referent of some perfectly ordinary use of a name...could be some item utterly isolated (causally) from the user's community and culture' (Evans 1973: 12). Introducing a name that in fact has a bearer involves a causal relation between the bearer and the introducer. The result is that a use standing in the appropriate causal relation to a normal (non-fictional) successful introduction is thereby a use causally related to the referent.

If Evans's detailed arguments against such theories were decisive, one would have a satisfying explanation for his supposition that the categories of Russellian and descriptive names are non-empty and exhaustive. I think, however, that an explanation can be found at a deeper level. Despite his advocacy of some aspects of a generally externalist picture of the mind—in particular, his insistence that, however it might seem to the subject, no thought is expressed by the use of a sentence containing a name with no bearer—internalism retains some hold on him. This is manifest in his discussion of the steel balls. Considering the case in which the subject, through amnesia, has forgotten everything about one of two similar steel balls he saw earlier, Evans writes:

There is no question of his recognizing the ball; and there is nothing else he can do which will show that his thought is really about one of the two balls (about *that ball*), rather than about the other. The supposed thought—the supposed surplus over the *ex hypothesi* non-individuating descriptive thought—is apparently not connected to anything. (Evans 1982: 115)

The general line is that if the subject were thinking about one of the balls, there should be something within the inner aspects of a subject's thought, and not merely the external causal fact, which makes this so. Indeed, Evans commits himself to something stronger: there needs to be something that the subject can *do* which will *show* which ball he is thinking about. These are clearly internalist views. The general principle is that one who can refer in thought can do something that shows that he does, rather as, for some internalists, one who is justified can show that she is. Here, perhaps, we find the deep root of Evans's interpretation of Russell's Principle as a highly demanding one. Whereas everyone can accept some interpretation of the mantra that one who refers must know to whom or what she refers, in Evans's hands the principle comes to require from the

R. M. Sainsbury

thinker the capacity to distinguish the referent from all other things. Reference is to be determined by facts 'within' the speaker, and so in some sense accessible to the speaker, rather than by the mere impact of causal forces, or mere immersion in a name-using practice.

Evans is happy to allow that information possessed by the thinker can help determine the object of her thought. There is some indication that he views a thinker as automatically in possession of the relevant properties of information possessed. He discusses the following example:

Do you remember that bird we saw years ago? I wonder whether it was shot. (Evans 1982: 308)

Evans says that the hearer does not fully understand the speaker 'until he remembers the bird-until the right information is retrieved'. But how can remembering the bird be glossed in terms of retrieving the 'right' information? Evans would agree that it is not enough to entertain merely predicative information, expressible as, for example, 'was crimson', however detailed this is, and even if it happens to be true of the right bird. The information is 'right' only in virtue of being causally derived from the right object, that is, the bird in question. This means that talk of the right information does nothing to advance the project of saying what it is to remember the bird. Evans's gloss suggests that, momentarily, he supposed otherwise; which in turn suggests, I believe, that he was tempted to try to internalize the external fact of causation in terms of internal states of information possession. Had Evans been happy with information possession characterized externalistically, the 'right' information could be described simply as information originating in that bird.

The internalist aspect is only lightly camouflaged by the importance that Evans assigns to the origin, rather than the fit, of information associated with a name in determining the referent. Since origin is a causal notion, externalists may think that justice is being done to their position. But since the relevant information is supposed to be accessible to the thinker, there is a sense in which something accessible to the thinker plays an instrumental part in determining her referent. A pure externalist, by contrast, will be happy to see the determination as fixed by facts that are not independently internally accessible—for example, by immersion in a practice. The causal relation will make the referent an object of thought for the thinker, but not necessarily in virtue of any further fact about the non-relational organization of the thinker's mind.

To point this out is not, as such, to provide a reason for thinking that Evans was misguided to retain these internalist strictures. On the other hand, if they lead to the view that every name is either Russellian or descriptive, a dispassionate look at the actual facts of our linguistic practice will, I believe, give us reason to reject them.¹⁴

¹⁴ My thanks to the Leverhulme Trust for the grant of a Senior Research Fellowship, during the tenure of which this paper was written, and to an anonymous referee for the Press and, especially, José Luis Bermúdez for helpful comments on the penultimate draft.

Plural Terms: Another Variety of Reference?

IAN RUMFITT

Ι

Things which appear similar to ordinary observation, and which behave similarly for ordinary purposes, are frequently called by the same name. When people acquire the methods for more detailed observation, and an interest in the construction of theories, many of these groupings have to be revised. Whales do not really belong with the fish they superficially resemble, since the similarity of form and behaviour conceals radical differences of structure and function. Now it may be said that our intuitive semantic classifications stand in need of similar revision. 'What do you mean?', 'Who are you *talking about*?', 'That's not what you *said*', 'That's not *true*'. These are the rough-and-ready semantic concepts of the market-place, which have been used and refined by many people-philosophers, grammarians, teachers-who have been obliged to reflect on the operations of their own language. And, despite this refinement, they continue to reflect those similarities in linguistic form and function which are particularly striking. It is the aim of this book to examine whether one such grouping of intuitive semantics—that of singular terms or referring expressions^{*} has a place within a developed semantic theory of natural language, and if so, what place it has. (Evans 1982: 1)

Thus Gareth Evans, in the very first paragraph of *The Varieties of Reference*; but one of his claims here ought to give us pause. In the last sentence quoted, Evans appears to identify the categories of 'singular term' and 'referring expression', and in a footnote appended at the asterisk he remarks that 'these two phrases will be used interchangeably throughout' the book (ibid.). His own initial

characterization of the intuitive notion of a referring expression, however, may lead one to doubt whether this identification can be correct. Evans follows Strawson in characterizing the 'referring task' as that of 'forestalling the question "What (who, which one) are you talking about?" '(ibid. 2). He also proposes that when 't is a referring expression understood in the same way in its occurrences in both of the sentences "t is F" and "t is G", then it follows logically from the truth of these sentences that there is something which is both F and G' (ibid.). No doubt these principles help to explain why, as Evans puts it, 'the class of referring expressions has traditionally been taken to include proper names; definite descriptions ("the tallest man in the world"); demonstrative terms ("this man", "that woman"); and some pronouns' (ibid. 1). For in the sentence 'That woman is smoking', the demonstrative term may be said to forestall the question 'Who are you talking about?'; and when the two occurrences of this term in 'That woman is smoking' and 'That woman is threatening me' are understood in the same way, these sentences jointly entail that someone is both smoking and threatening me. By the same token, however, these intuitive principles suggest that the category of referring expressions should include *plural* as well as singular terms. For in the sentence 'Those women are smoking', the plural demonstrative term may also be said to forestall the question 'Who are you talking about?' And when the two occurrences of this term in 'Those women are smoking' and 'Those women are plotting to kill us' are understood in the same way, these sentences jointly entail, not that someone is both smoking and plotting to kill us, but that some people are. The point extends to the pronouns that Evans wanted to classify as referring expressions. He defended the view that 'in the sentence "John owns a donkey and beats it", the pronoun "it" has the function of designating [i.e. referring to] the object (if any) that verifies the antecedent clause containing the existential quantifier' (Evans 1977: 118-19). But his arguments for this view equally support the claim that in the sentence 'John owns some sheep and Harry vaccinates them', the pronoun 'them' has the function of designating the objects (if any) that verify the antecedent clause (cf. ibid. 104).

This last point suggests that Evans, had he lived to complete *The Varieties of Reference*, might well have revised his initial identification of referring expressions with singular terms. Indeed, in a footnote

Ian Rumfitt

to his discussion of definite descriptions, he confesses: 'it is one of the deficiencies of this book that I have not included work on plural reference' (Evans 1982: 60 n.). However that may be, the work as it stands leaves an important matter of unfinished business. The Varieties of Reference constitutes an extended argument for according to the category of referring expressions an important place within a developed semantic theory—so long as we recognize that the category contains sub-varieties which function in significantly different ways. As Evans would have been the first to insist, however, the notion's utility does not mean that any words that pass the intuitive tests will be treated as referring expressions by a developed semantic theory. He held, indeed, that the development of such a theory exposes deep differences between proper names and singular indexicals and demonstratives, on the one hand, and singular definite descriptions on the other (Evans 1982: 53-60). In 'The man who broke the bank at Monte Carlo was tall', the definite description may be said to forestall the question 'Who are you talking about?'. And it follows from this sentence together with 'The man who broke the bank at Monte Carlo was lucky' that someone was both tall and lucky. All the same, Evans argued, when we undertake the primary semantical task of explaining how expressions of diverse sorts help to determine the truth conditions of sentences containing them, we shall conclude that 'it is better to treat "The φ " as a quantifier than as a referring expression' (Evans 1982: 59).

In this essay, I address the corresponding question about plural terms. That is to say, I consider whether plural terms are best treated as referring expressions. By a *plural term*, I shall mean a plural analogue of one of the singular terms that Evans took to be referring expressions. Thus the category is understood to comprise compounds of singular proper names, such as 'William and Mary'; collective names, such as 'the Channel Islands'; plural indexicals and demonstratives, such as 'we' and 'those men'; and compounds of two or more singular or plural terms, such as 'my colleagues and I'. What I mean by a 'referring expression' will become clearer as we go on. But the basic idea is that a positive answer to my question amounts to two claims. First, a plural term such as 'the Channel Islands' *designates* Jersey, Guernsey, Alderney, etc. in just the sense in which the proper name 'Jersey' designates the island of Jersey. Second, the plural term's designating those islands accounts

Plural Terms

for the truth conditions of sentences containing the term in just the way that the proper name's designating that island accounts for the truth conditions of sentences that contain the name. In taking congruity in accounting for truth conditions to be the mark of a semantical kind, I am following Evans.^T

I shall defend a positive answer to my question. In a word, I shall be taking a 'referentialist' position on plural terms. A full defence of either answer would require a fully developed semantic theory for English, and this I do not have. But I hope to make a positive answer plausible, not only by rebutting objections to it, but also by saving the referentialist position from false friends who have propounded inadequate versions of it. Much of the paper consists, indeed, in ground clearing. As will be seen, even prominent theorists in this area do not always see clearly what separates them from their rivals. Helping them do so may advance the current debate about plurals even if further developments in semantic theory should ultimately falsify a positive answer to my question.

I shall not try to decide whether plural descriptions are best grouped semantically with compound names, plural indexicals, and demonstratives etc., or whether (as Evans would have preferred) the plural definite article should be construed as a determiner, so that 'The men who broke the bank at Monte Carlo had known one another since their school days' is grouped with 'Five men at the party had known one another since their school days'. (For a version of this latter view, see Sharvy 1980.) Many of the relevant considerations here are analogues of familiar moves in the corresponding debate about singular descriptions. Thus some philosophers follow Donnellan (1966) in discerning a distinctively referential use of the singular description in 'Smith's murderer is insane' when that sentence is uttered in the presence of a man commonly assumed to have murdered Smith. They will no doubt say the same about the plural description in 'Smith's murderers are insane' when the latter sentence is uttered in the presence of two or more men commonly assumed to have murdered Smith. Mv official agnosticism on this question will not stop me from using plural descriptions as examples of plural terms, however. Although he denied that singular definite descriptions were referring expressions, Evans held that it was always possible to introduce a singular

¹ See again his reasons for not classifying definite descriptions as referring expressions: Evans 1982: 53–60.

Ian Rumfitt

referring expression (a 'descriptive name') to stand for any object identified by description (1982: 31-2, 49-51). His arguments for this view are persuasive, and they extend to show that one may always introduce a plural referring expression (which we may call a 'descriptive collective name') to stand for any objects identified by description. Readers who deny that plural descriptions are referring expressions are, then, invited to replace examples of them in the sequel by a suitable descriptive collective name.

Π

Some philosophers have defended a referentialist position on plural terms by treating them as semantically singular. A prominent example is Tyler Burge. Discussing the sentence 'The stars that presently make up the Pleiades galactic cluster occupy an area that measures 700 cubic light years', Burge suggests that

the most natural tack is to treat the subject term...as a singular term and give the relational predicate the reading 'occupies'. We have precedent (for example, in our dealings with quantified sentences) for counting the difference between singular and plural forms of natural-language predicate expressions irrelevant as far as our formal representation of those expressions is concerned, and for locating the effect of the plural elsewhere in the sentence. Thus, the expression 'occupy' in 'All men occupy space' is formally represented by the same predicate as is 'occupies' in 'Burge occupies space'. (Burge 1977: 98)

There are certainly some cases in which a plural term is best treated as semantically singular. In the sentence 'The Channel Islands, the Scillies, and the Cairngorms are the three most popular tourist destinations in Britain', the first three words are naturally taken to refer to the whole archipelago: it would betray a misunderstanding of the intended sense to object that the sentence must be false because there are more than three Channel Islands.² For reasons to be given shortly, I also think that Burge is right to seek a treatment according to which the singular and plural forms of a verb like 'occupy' share a sense. Such a treatment, though, does not entail that there is some one thing that a plural term designates, and I begin by arguing that in general no such thing is to be found.

² I owe this example, and the point it illustrates, to Kevin Mulligan.

As Burge observes, a salient candidate for such a thing is a set. 'The stars that presently make up the Pleiades galactic cluster' may be taken to designate the set of such stars-i.e. the set to which an object belongs if and only if it is a star now in the Pleiades cluster. This view, however, faces an immediate difficulty. In the ordinary sense of 'occupy', a set is not the sort of thing that occupies space. And while it may be possible to explain an extraordinary sense in which a set may be said to occupy space, Burge is right to deny that there is any 'intuitive basis for thinking that the stars occupy the area in a different sense ("set-occupies") from the sense in which the cluster [a single physical object] occupies it' (1977: 99). One might, perhaps, try to unify the senses here by saving that a single object occupies space just when the corresponding singleton set-occupies it. But even if one could swallow this, another problem looms. One may speak in the plural of sets themselves, as when one says, 'The sets that are not members of themselves include the empty set and the singleton whose only member is the Pleiades cluster'. But this plural description cannot be understood to designate the set of all sets that are not members of themselves, for Russell's paradox shows that there is no such set (cf. Boolos 1984: 66). A philosopher who believes in proper classes might take this plural description to designate the proper class of all sets that are not members of themselves. But this device offers no solution to the underlying problem. For one may speak in the plural of proper classes as well as of sets, as when one uses the expression 'The proper classes of all impure sets and of all ordinals'. This expression cannot be understood to designate the proper class to which the proper class of all impure sets and the proper class of all ordinals belong; for the mark of a proper class is precisely that it does not belong to anything. Neither sets, nor set-like things, will do as *designata* of plural terms (cf. Lewis 1991: 67).

A second line of thought takes a plural term to designate a suitable mereological fusion (Goodman 1966: 51-2). Now the expression 'the stars that make up the Pleiades cluster and the prime factors of 30' is an intelligible plural term. So if there is to be a fusion corresponding to every such term, we shall need to assume a fairly generous principle of mereological composition.³ But even if

³ See, though, Lewis 1991: 79–81 for a defence of the following principle of unrestricted composition: 'Whenever there are some things, no matter how many or how unrelated or how disparate in character they may be, they have a mereological fusion' (ibid. 7).

Ian Rumfitt

such a principle is granted, a problem remains. For even when it is given that there is such a thing as the mereological fusion of the As, it will not always be possible to find a non-plural predication about that fusion which has the same truth conditions as 'The As are F'. For consider the statement 'The bags of sugar in the larder are evenly distributed between its shelves'. Let us suppose, for the sake of argument, that there is such an object as the fusion of these bags of sugar. On that supposition, the statement 'The fusion of the sugar in the larder is evenly distributed between its shelves' makes perfectly good sense. All the same, its truth conditions are not those of the original plural predication. If there are two shelves in the larder, one of which supports a two-pound bag of sugar, while the other supports two one-pound bags, then the statement about the fusion will be true, while the plural predication will be false. Of course, once it is granted that the fusion of the sugar exists, our plural predication's truth conditions will be shared by those of 'The fusion of the sugar in the larder is such that the bags of sugar that make it up are evenly distributed between its shelves'. Here, though, the reference to the fusion is clearly an idle wheel; the statement has the right truth conditions only because it invokes the subordinate *plural* predication 'the bags of sugar that make it up are evenly distributed between its shelves'. So, even if accounts of statements involving plural terms that postulate fusions can escape falsity, they do not constitute progress. This observation generalizes a point that Frege made about attributions of number. Mill had taken a cardinal number to be a 'property belonging to the agglomeration of things'; more exactly, he had taken it to be 'the characteristic manner in which the agglomeration is made up of, and may be separated into, parts' (Mill 1843: book III, ch. xxiv, §5). To this proposal, Frege objected that 'the definite article in the phrase "the characteristic manner" is a mistake right away; for there are various manners in which an agglomeration can be separated into parts, and we cannot say that one alone would be characteristic' (Frege 1884: §24, p. 30). So it is in our example. There are various ways in which the sugar in the larder might be separated into bags, so an even distribution of the sugar is consistent with an uneven distribution of the bags.

Burge himself recognizes difficulties with both these suggestions, and proposes taking plural terms to refer to what he calls 'aggregates'. We need not go into the details of his proposal, however, to

Plural Terms

appreciate a general problem for any theory that takes the reference of an arbitrary plural term to be a single object. The problem in question is a version of Russell's paradox, and is due to Alex Oliver and Timothy Smiley (2001), who build on prior work by Barry Schein (1993) and by Boolos. For suppose that the single item that a plural term is taken to designate is called a *collection*. And suppose we say, whenever an object x is one of the Fs, that x is a *constituent* of the collection of Fs; on the view being considered, constituency is a dyadic relation between objects. Now consider four men: as they might be, Moore, McTaggart, Whitehead, and Russell. Then, as Oliver and Smiley observe, these four men give rise to six pairs:

Six collections, then, but only four men; so *either* two collections must be identical to the same man, *or* one of them is not identical to any of the men. But the first alternative would make two of our collections identical, so that 'Moore and McTaggart wrote *Principia Mathematica*' would come out true, or some similar equally absurd result. So the second alternative must obtain, and for argument's sake we may take it to be the collection of Whitehead and Russell, which is not identical to any of the men. In particular it is neither Whitehead nor Russell, and therefore not a constituent of itself. (Oliver and Smiley 2001: 304)

It follows that the statement 'The collection of Whitehead and Russell is one of the things which are not constituents of themselves' is true. On the view being considered, then, it would be true to say 'The collection of Whitehead and Russell is a constituent of the collection of things which are not constituents of themselves'. But this statement cannot be true, for there is no such collection: it would be a constituent of itself if and only if it is not. In this *reductio*, 'collection' is no more than a place-holder for the kind of object that is taken to be the reference of a plural term. Such a term, then, cannot be treated as though it were semantically singular.

This argument is not quite knock-down. One way to block the threatened paradox would be to insist that in the problematical claim, 'The collection of Whitehead and Russell is a constituent of the collection of things which are not constituents of themselves', the range of 'things' cannot include the collection of things which are not constituents of themselves. This insistence might itself be grounded in two ways: first, by denying that there is any such object as the collection of things which are not constituents of themselves, perhaps for reasons akin to those which a Zermelo-Fraenkel

Ian Rumfitt

set-theorist has for denying that there is a set of things which are not members of themselves; second, by denying that it is legitimate to quantify over all the objects that there are. All the same, the cardinality considerations on which Smiley and Oliver rely generalize to establish an important conditional conclusion: namely, that *if* it is possible to quantify over absolutely everything, then *either*

- (1) there are some things, the Fs, for which there is no singular surrogate
- or
- (2) there are some things, the *F*s, and some things, the *G*s, such that the *F*s are not the *G*s, but the singular surrogate for the *F*s is the same as the singular surrogate for the *G*s.

Option (1) will obtain if the singular surrogate for some things is taken to be a set or a class; option (2) will obtain if it is taken to be a mereological fusion. Neither option will be found at all comfortable by a theorist who aspires to give a general semantic theory for plural terms by postulating that such a term refers to a suitable singular surrogate. So if it is possible to quantify over absolutely everything, then a plural term cannot be treated as though it were semantically singular.⁴

III

How, in that case, should a referentialist treat plural terms?

We may begin by remarking that some of the predicates that attach to a plural subject can attach to a singular subject without any change of sense. When we say 'William and Mary reigned over England', we place William and Mary in the same relation to England in which we place Louis XIV to France when we say 'Louis XIV reigned over France'. What he did to France by himself between 1643 and 1715, they did to England together between 1689 and 1695. How can we be sure that these occurrences of 'reign' are univocal? One powerful reason has to do with WHquestions. If I ask 'Who reigned over England in the 1690s?', I shall

⁴ I cannot defend the antecedent of this conditional here. For powerful arguments in its favour, however, see Cartwright 1994. In this paragraph, I am much indebted to correspondence with Agustin Rayo.

Plural Terms

be told: 'William and Mary did so jointly until Mary's death in 1695, whereupon William did so alone until his death in 1701.' It is hard to credit that 'reigned' in the question has to be taken in two senses, depending on whether it expects a plural or a singular answer.

The version of referentialism that I espouse rests on the idea that the key semantical predicates, 'designate' and 'satisfy', are such expressions. They, too, can attach to a plural or a singular subject without change of sense. When we have a singular referring expression, such as the proper name 'Louis XIV', we may speak of that expression's designating a certain object. Equivalently, we may say that the object in question, the Sun King, is designated by the name 'Louis XIV'. And just as the verb 'reign' has a common sense in the statements 'Louis reigned over France' and 'William and Mary reigned over England', so the verb 'designate' has a common sense in the statements 'The Sun King is designated by the name "Louis XIV"' and 'William and Mary are designated by the name "William and Mary"'. What he did to France by himself, they did to England together. And what the name 'Louis XIV' does to him by himself, the compound name 'William and Mary' does to them together. On this view, then, the designations of plural terms will be stated by such principles as:

'William and Mary' designates William and Mary.

'The Channel Islands' designates the Channel Islands.

'Those men' (as used in such-and-such a context) designates Smith and Jones.

In understanding these principles, two points are crucial. First, the expressions that follow the verb 'designates' are themselves plural terms. To say that the compound name 'William and Mary' designates William and Mary is not to place the name in a relation to any one thing. Rather, the name designates *them.*⁵ Second, the various predicates in the form '*NN* designates . . .' do not distribute. That is, even though we are given that 'The Channel Islands' designates the

⁵ The otherwise congenial theory of plural reference developed by David Bell (1990b: 63–9) from suggestions of Husserl occludes this point with its talk of 'William and Mary' designating the aggregate of William and Mary. Bell is clear that 'the aggregate of William and Mary' is not a singular term and must be understood to be co-designative with 'William and Mary' (ibid. 64). But if so, talk of aggregates explains nothing and merely brings into play an expression which an unwary reader might misconstrue as singular.

Channel Islands, and that Jersey is one of the Channel Islands, we cannot infer that 'The Channel Islands' designates Jersey. Similarly, it does not follow from the fact that the compound name 'William and Mary' designates William and Mary that it designates William.

The like is true of the semantical verb 'satisfies'. A single object satisfies the predicate 'reigned over France' if and only if it reigned over France. And some objects satisfy the predicate 'reigned over England' if and only if they reigned over England. Just as the occurrences of 'reigned' in these sentences are univocal, so too are the occurrences of 'satisfy'.

We have here the rudiments of a truth theory for English that vindicates the classification of singular and plural terms as two kinds of referring expression. For we may state the semantical contribution of the names 'Louis XIV' and 'William and Mary' by saying

1. The proper name 'Louis XIV' designates Louis XIV

and

2. The compound name 'William and Mary' designates William and Mary.

We may also give the semantical contributions of 'reigned over France' and 'reigned over England' by the conjunctive axioms

- 3a. Whatever it may be, it satisfies 'reigned over France' if and only if it reigned over France.
- 3b. Whatever they may be, they satisfy 'reigned over France' if and only if they reigned over France.

and

- 4a. Whatever it may be, it satisfies 'reigned over England' if and only if it reigned over England.
- 4b. Whatever they may be, they satisfy 'reigned over England' if and only if they reigned over England.

We also have the general principle

T. A sentence formed by combining a singular or plural term with a predicable is true if and only if what the term designates satisfies, or satisfy, the predicable. Given these semantical axioms, we may derive the truth conditions of 'Louis XIV reigned over France' as follows:

- 1. 'Louis XIV reigned over France' is true if and only if what 'Louis XIV' designates satisfies 'reigned over France' (from axiom T).
- 2. 'Louis XIV' designates Louis XIV (axiom 1).
- 3. 'Louis XIV reigned over France' is true if and only if Louis XIV satisfies 'reigned over France' (from lines 1 and 2).
- 4. Louis XIV satisfies 'reigned over France' if and only if Louis XIV reigned over France (instance of axiom 3a).
- 5. 'Louis XIV reigned over France' is true if and only if Louis XIV reigned over France (from lines 3 and 4).

We may similarly derive the truth conditions of 'William and Mary reigned over England':

- 1. 'William and Mary reigned over England' is true if and only if what 'William and Mary' designates satisfy 'reigned over England' (from axiom T).
- 2. 'William and Mary' designates William and Mary (axiom 2).
- 3. 'William and Mary reigned over England' is true if and only if William and Mary satisfy 'reigned over England' (from lines 1 and 2).
- 4. William and Mary satisfy 'reigned over England' if and only if William and Mary reigned over England (instance of axiom 4b).
- 5. 'William and Mary reigned over England' is true if and only if William and Mary reigned over England (from lines 3 and 4).

I said earlier that the mark of a semantical kind was congruity in accounting for truth conditions. The case for classifying proper names and compound names as varieties of referring expression is that these derivations of truth conditions manifestly run parallel, and that there is no equivocation over the semantical terms used in them. The same argument extends to vindicate the classification of our other varieties of plural term as referring expressions.

Simple as it is, this argument may not be persuasive. The suggested semantic theory takes plurality seriously, in that it uses plural terms and plural forms of semantic predicates in giving the semantic axioms for the object language's plural terms and the plural forms of its predicates. Those axioms, however, will not yield any T-theorems without a logic, and we need an explicit statement of the logical principles that

Ian Rumfitt

are tacitly assumed in the derivation. The reason why we need this is not that there is any doubt about the derivation's correctness. The claim that William and Mary satisfy 'reigned over England' if and only if they reigned over England (line 4 in the second derivation) certainly follows from the axiom that whatever (or whoever) they may be, they satisfy 'reigned over England' if and only if they reigned over England. The worry, rather, is that an explicit statement of the logical principles regulating inferences involving plural quantifiers and their attendant pronouns might well expose a difference in sense between singular and plural instances of 'designate' or 'satisfy'. Such a difference would undermine the claim that singular and plural terms are both species of designator. No doubt a difference in sense is prima facie implausible. It is no easier to believe that the question 'Who satisfies the predicate "wrote The Iliad"?' is equivocal, depending on whether it expects a singular or a plural answer, than it is to believe the same about the corresponding question in the material mode, viz. 'Who wrote The Iliad?'. But in advance of an explicit statement of the relevant logical principles, we cannot be sure that they do not force one to acknowledge such a difference.

I accept the challenge of vindicating the logical univocity of the semantical predicates. Most of the rest of the paper is devoted to meeting this challenge. First, however, it will help to consider the position of a philosopher who thinks that any account along the lines I have sketched is quite misguided, so that pursuing these matters further would be a waste of time.

IV

The philosopher I have in mind is Michael Dummett. 'Fregean semantics', he writes, 'undermines the superficial similarity between singular and plural...a plural noun-phrase, even when preceded by the definite article, cannot be functioning analogously to a singular term.' Rather, 'it is only as referring to a concept that a plural phrase can be understood, because only a concept-word admits a plural. But to say that it refers to a concept is to say that, under a correct analysis, the phrase is seen to figure predicatively. Thus "All whales are mammals", correctly analysed, has the form "If anything *is a whale*, it is a mammal", and "The Kaiser's carriage is drawn by four horses" the form "There are four objects each of which *is a horse that draws the Kaiser's carriage*"' (Dummett 1991a: 93). In the

Plural Terms

passage from which I have been quoting, the particular version of referentialism that Dummett primarily has in mind is one that takes a plural term to designate a single composite object. But his position must rule out any theory which posits a common sense for the verb 'designate' in the statements 'The singular term "William" designates William' and 'The plural term "William and Mary" designates William and Mary', for the proper name 'William' (as it is used at the end of the first statement) is certainly not a Fregean concept-word.

What is Dummett's ground for saying that we can only understand a plural phrase as referring to a concept? All he says in the passage just quoted is that 'only a concept-word admits a plural', but this reason is not impressive. In the plural phrase 'these men', 'men' is indeed the plural of the concept-word 'man'; equally, though, 'these' is the plural of the demonstrative pronoun 'this'. We are, after all, trying to reach an understanding of the semantic function of entire phrases like 'these men' rather than of their component nominals or concept-words. I think a better indication of Dummett's reasons is suggested by the allusion to Frege's doctrine of numerical adjectives at the end of the passage I quoted.

At the core of that doctrine is a perception that numerical adjectives do not have the semantic function that a casual inspection of their grammatical distribution might lead one to expect. Grammatically, the sentence 'Five horses are in that field' seems to belong with 'White horses are in that field'. Frege's theory of number, though, starts from the insight that such grammatical congruity as there may be between 'white' and 'five' conceals a profound semantic difference. In saving that white horses are in that field, one is saying that there exist at least two objects each of which (a) is white, (b) is a horse, and (c) is in that field. Thus the adjective 'white' serves to effect a classification of objects-or at least, of middle-sized material objects. On Frege's view, indeed, absolutely any object must either answer to the adjective 'white' or not. (This is the principle that he calls *tertium non datur*.) In saving that five horses are in that field, however, one is not saving that there exist at least two objects each of which (a) is five, (b) is a horse, and (c) is in that field. The difference arises, Frege thinks, because it is not the role of the adjective 'five' to effect a classification of objects. It is not that some objects answer to the adjective 'five' while others do not-though of course some objects may answer to the distinct adjective 'has five members'. Rather, 'five' serves to effect a classification of modes of classifying objects.

The expression 'horse in that field', does serve to classify objects: some objects (we may suppose) answer to it, while others do not. But in saying 'Five horses are in that field', one is classifying that mode of classification as having five instances. That is to say, one is effecting a classification which places that mode of classification alongside such other modes of classification as 'being a Platonic solid', 'being a movement of the Pastoral Symphony', and so forth. This is what Frege means when he says that 'five' is a second-order adjective. It classifies classifications of objects.

Dummett thinks that this account of the meaning of numerical adjectives was a great insight on Frege's part, and so do I. Why, though, might it lead one to think that plural terms are predicative? Well, having come so far, one may be tempted to go a step further and reason as follows. A numerical adjective may attach directly to a plural term. Thus, leading my party into a restaurant, I may say to the head waiter, 'We are five'. According to Frege's theory, 'five' is a second-order adjective. So any subject to which it attaches ought to be a first-order adjective. The grammatical subject of the sentence 'We are five', though, is the plural term 'we'. So it may come to look as though that expression must be construed as having a predicative semantic function if we are to respect the insights of Frege's theory of numerical adjectives.

Whether or not this argument is Dummett's, it is interesting. But it is not, I think, persuasive. It rests on two claims. First, that the occurrence of 'five' in 'We are five' has exactly the same sense as the occurrence of 'five' in 'Five horses are in the field'. Second, that the occurrence of 'are' in 'We are five' is die bloße Kopula, a grammatical tie between subject and predicate that contains nothing of semantic significance. These claims, however, are hard to sustain. Just as we may say that five horses are in that field, so we may say that five of us are. But the 'of' here is semantically significant. In saving that five of us are in the field, we are ascribing the property of having five instances to what Frege would have called the firstlevel concept of being one of us in that field. Now there is no doubt that 'one of us' is a first-level predicate, an expression that effects a classification of objects. (Mrs Thatcher was said to be intensely interested in the classification that it effects of persons.) But it does not follow that the component plural term 'us' must be understood as figuring predicatively. Again, the 'are' in 'We are five' does not seem to be purely copulative. It seems to have the sense of 'We

number five', i.e. 'There are five of us'. Frege's theory of numerical adjectives, then, does not commit one to treating a plural term as predicative.

Unpersuasive as this argument may be, it helps to bring out the issue between Dummett and the referentialist. On the referential view of plural terms, a speaker who understands the predicate 'one of us' will do so by virtue of knowing who we are, and by virtue of understanding quite generally what it is for a thing to be one of some things. On Dummett's view, this is back to front. If we understand a statement containing a plural term, that will be because we understand an analysis of the statement in which the plural term has given way to various predicative expressions. How is this issue to be resolved?

So far from its being necessary to understand statements involving plural terms by way of such analyses as Dummett proposes, they do not in general seem to be possible. Perhaps 'All whales are mammals' may be analysed as meaning 'Anything that is a whale is a mammal'. But what about even a simple statement involving a non-distributing predicable, such as the statement 'William and Mary reigned over England together'? William and Mary are the things, each of which is one of William and Mary, so that the statement is equivalent to 'The things, each of which is one of William and Mary, reigned over England together'. But we are still left with a plural definite description. Following Richard Sharvy (1980), one might Russell this description away in favour of plural quantifiers and attendant pronouns. For, as he observes, 'The As are F' is equivalent to 'Some As are such that all As are some of them, and they (i.e. those As) are F'. But the plural pronoun in the last clause has not been replaced by a predicative expression, so this equivalence (even if it could be accorded the status of an analysis) does not give Dummett what he wants. Of course, one might say that 'William and Mary reigned over England together' is equivalent to 'Each one of William and Mary is one of those who together reigned over England'. As we shall see, this trick has its uses, but it would be absurd to claim that it provides any analysis of a statement containing a plural. For the putative analysans contains a plural relative clause ('who together reigned over England'). If we lack an understanding of how the elements of a complete plural statement come together to determine its meaning, we also lack such an understanding of a plural relative clause. If there is a method

of analysis under which any plural term may be seen to figure predicatively, Dummett has yet to tell us what it might be.

V

George Boolos was famous for suggesting a regimentation of arguments involving plurals in which plural quantifiers and dependent pronouns were represented as second-order quantifiers and associated predicate variables (Boolos 1984, 1985*a*, 1985*b*). It may seem that his suggestion is just as hostile to a referentialist view of plurals as is Dummett's theory. Some of Boolos's critics certainly imagine so. As I wish to explain, however, things are not so simple. When Boolos's theory is purged of elements that are demonstrably unsound, what remains is a theory with which a referentialist can be comfortable, and for which he may even be grateful.

We may think of Boolos as having tried to broker a marriage of convenience, contrived to dissolve the embarrassment that its contracting parties would face were they to remain single. If plural terms and plural quantifiers cannot be dispensed with in favour of singular surrogates, we shall eventually need rigorously formulated rules for evaluating arguments that involve them. Since the rules of axiomatic second-order logic may be formulated as rigorously as anyone could wish, a translation scheme, such as Boolos proposed, from vernacular statements involving plurals to formulae of secondorder formal languages at least makes a start on the problem of identifying valid arguments involving such statements. Boolos thought, though, that the other party to the contract stood to benefit too. Rigorously formulated as the rules of second-order logic may be, many logicians and philosophers have found it obscure what they are rules for; and the converse translation from formalism to vernacular provided (as Boolos liked to say) a 'reading' of second-order languages that could lighten their darkness. Under the translation he proposes, the second-order quantifiers are not understood as ranging over a domain of allegedly mysterious things such as Fregean concepts. Neither are they an odd way of quantifying over sets. Rather, they range over the same objects as the associated first-order variables, but they do so in a plural rather than a singular fashion.

Plural Terms

Oddly, some of Boolos's critics fail to see that he is assigning a novel interpretation to the devices of a familiar formalism so that one has to put out of mind more traditional ideas about what that formalism means. Beong-Uk Yi, for example, complains that

the scheme of second-order analysis meets difficulties in preserving logical relations among facts like the following:

- (9) Russell and Whitehead cooperate
- (10) Russell and Whitehead are the authors of Principia Mathematica
- (11) The authors of *Principia Mathematica* cooperate.

Consider the second-order facts that the scheme renders them as:

- (9a) cooperate (is identical with Russell or Whitehead)
- (10a) $\forall \alpha \ [\alpha \text{ is identical with Russell or Whitehead} \leftrightarrow \alpha \text{ is an author of$ *Principia Mathematica* $]}$
- (11a) cooperate (is an author of Principia Mathematica).

Although the properties indicated by the phrases 'is identical with Russell or Whitehead' and 'is an author of *Principia Mathematica*' are co-extensional (so, fact (10a) holds), they are not the same property; to be identical with Russell or Whitehead is not the same as to be an author of *Principia Mathematica*. If so, one of the properties may instantiate a second-order property—for example, the one indicated by '*cooperate*'—that the other does not. Thus (11a) is not a logical consequence of (9a) and (10a). But, clearly, (11) is a logical consequence of (9) and (10). (Yi 1999: 172–3)

So, Yi concludes, the regimentation of plural terms using a secondorder formal language misrepresents the structure even of the simple inference whose premisses are (9) and (10) and whose conclusion is (11). As a criticism of Boolos, though, this complaint misses the point. We may agree with everything that Yi says about properties. But who says that a second-order formula *has* to be understood as attributing something to a property? Boolos's idea was that it does not.

All the same, and at the risk of flogging a dead concept *horse*, there are reasons for deeming Boolos's suggestion inadequate as a general method of understanding higher-order formal languages. As a number of writers have pointed out,⁶ in extending his treatment to cover quantification over *n*-place relations, we shall need to quantify plurally over something like *n*-tuples. Since *n*-tuples are

⁶ See e.g. Rayo and Yablo 1999.

normally understood to be a species of set, this does not square with Boolos's ambition of giving a nominalistic reading of the second-order quantifiers, in which they are innocent of any ontological commitment not already carried by the first-order quantifiers. What is worse, the interpretation does not extend to cover third- or higher-order logics. We can all understand what is meant by saying that a speaker is quantifying plurally over some objects. But what is meant by saying that he is quantifying plurally plurally over objects? The only sense I can make of this is as saying that one is quantifying plurally over objects with many members. But again, that brings in ontological commitments not incurred by quantification over the first-order domain.

In both of these respects, Boolos's way of making sense of secondorder formal languages compares unfavourably with a method that Dummett had proposed some years earlier (Dummett 1973: 216-17). The crucial infelicity in Frege's talk of concepts, or *Begriffe*, is that the expression 'is a concept' behaves grammatically like a first-level predicate, whereas its sense is supposed to be that of a higher-order expression. Thus the words 'Julius Caesar is not a concept' form a well constructed sentence, even though-given the constraints of the Fregean hierarchy of levels-they ought not to make any sense. As Dummett showed, however, we may evade this infelicity while remaining true to the essence of Frege's conception of higher-order discourse. For instead of saying that the secondorder variables range over first-level concepts, we may say that they range over things which any object either is or is not. In this formulation, the last two occurrences of the word 'is' are to be understood as being copulative. And while we can make sense of the sentence 'Julius Caesar is a thing that any object either is or is not', we can make sense of it only by switching to the unintended reading on which the last two occurrences of 'is' express identity. Unlike Boolos's, this explanation extends to cover both relational concepts (Frege's Beziehungen) and concepts of third and higher orders. I am a former pupil of Dummett's, and my wife is a former pupil of Strawson's. So there is something that any object either is or is not to any other object, and I am that thing to Dummett, and my wife is that thing to Strawson. This illustrates how Dummett's scheme can extend without strain to cover quantification over relations. Again, we may describe the role of a monadic third-order variable by saying that it ranges over things which any thing that any object either

Plural Terms

is or is not itself either is or is not. No doubt, in explaining the role of variables of the fourth and higher orders, it will soon become irresistible to introduce a potentially misleading expression like 'first-level concept' as an abbreviation for 'thing that any object either is or is not'. It is clear, though, that the device for constructing such explanations may in principle be iterated as often as we wish.

This last consideration may be accorded little weight, on the ground that logics of third and higher orders are esoteric systems of purely formal interest. But they are more important than that. In his paper 'Reading the Begriffsschrift' (1985b), Boolos showed how under his translation scheme, instances of the second-order comprehension schema could be read as saving such things as: 'Either there is no object such that ... it ... or there are some objects such that an arbitrary object is one of them iff...it' But suppose we are trying to read, not the Begriffsschrift, but Part III of Grundgesetze, where Frege expounds his theory of real numbers as those which give a measure, i.e. which answer such questions as 'How tall?' and 'How long?'. Basic to that theory is the notion of a Größengebiet, or quantitative domain, which Frege conceived as comprising some relations over a field of objects. Thus the quantitative domain for tallness will comprise such relations as ' \mathcal{E} is the same height as n', ' ξ is 1 cm taller than η ', as these relations obtain between (for example) people. The members of a quantitative domain, then, will lie at the first level of the Fregean hierarchy of levels, counting the objects as the ground or zero'th level. Although the treatment of reals in Grundgesetze is incomplete, there is every indication that the next stage in their construction would have been to consider ratios of members of quantitative domains. From a formal point of view, Frege treats the ratio of *a* to *b* as a relation between *a* and *b*, so that ratios of members of a quantitative domain will lie at level two of the hierarchy. One may then define an equivalence relation between these ratios, and introduce real numbers by abstraction with respect to this equivalence relation. This will be analogous to the way Frege introduces cardinal numbers by abstraction with respect to the equivalence relation of equinumerosity, in the axiom that Boolos liked to call 'Hume's Principle'. But the analogy should not blind us to the fact that we are one level higher in the Fregean hierarchy. Because ratios belong at the second level, any equivalence relation between ratios belongs at the third level, and a third-order quantifier will be needed to bind variables that range over ratios.

The abstraction principle for the reals, then, will be a third-order abstraction principle, so we shall need to be able to 'read' thirdorder quantifiers in order to understand it.

It may be protested that Frege must have had in mind some other way of defining the real numbers, for he does not trouble to introduce any third-order quantifiers into the logical apparatus of Grundgesetze (Frege 1903). The observation is correct, and even in the extant part of his theory of real numbers, we see him preparing the ground for the evasion of third-order logic by using a relation's extension (what he calls a Relation) as a proxy for the relation proper (the Beziehung). This device for bringing the analysis down a level, however, is grounded on the fatal Basic Law V, and a lesson of the paradoxes is that there will not in general be an object that can stand proxy for a relation. What Frege sketches, then, is an essentially third-order theory of the real numbers that is brought down-in more than one sense, alas—by the application of Basic Law V. Boolos would have been the last man to deny the philosophical interest of reconstructing that third-order theory with a view to seeing whether it provides a satisfactory account of the real numbers. But his approach to higher-order logic does not provide us with the conceptual resources needed to understand it.

VI

I conclude that Boolos's approach to the problem of making sense of higher-order languages was unsuccessful, and a step back from Dummett's. That conclusion, however, does not refute the other side of his project, which was to exploit the syntactic resources of second-order formal languages to provide regimentations of vernacular statements and inferences involving plurals. Indeed, as I now explain, we enjoy greater freedom of action in exploiting those resources once we abandon the attempt to find a translation scheme which will validate all the principles even of monadic axiomatic second-order logic.

The requirement that it should validate all those principles clearly puts the basic idea of Boolos's translation scheme under strain. That idea was to have a second-order variable translate a plural pronoun such as 'they', just as a first-order variable translates a singular pronoun such as 'it'; and then to read the second-order quantifiers Plural Terms

as quantifying plurally over the first-order domain. As Boolos recognized, however, the theorems of axiomatic second-order logic include principles whose truth depends on empty first-level concepts' being admissible values of second-order variables, so that special provisos and saving clauses are needed to cope with this case. For example, an instance of the second-order comprehension schema is:

(C) $\exists X \forall x (Xx \leftrightarrow x \neq x)$.

If one reads this formula in Dummett's style, it says: 'There is something that any object either is or is not, and an object is that thing if and only if it is distinct from itself.' When so read, the formula expresses a truth. If, however, we were to read it as saying 'There are some objects such that any object is one of them if and only if it is distinct from itself', it comes out false. For if there are some objects such that any object is one of them if and only if it is distinct from itself, then at least one object must be distinct from itself. But no object is distinct from itself.

Boolos was well aware of this problem, and he solved it by complicating his translation scheme with clauses that provide for the empty case. Once we have abandoned the enterprise of seeking a translation that will validate all the theorems of axiomatic second-order logic, however, we can tinker with the formal system instead of complicating the scheme of translation. One way of doing this is suggested by Prawitz's work. In his textbook on natural deduction, Prawitz presented two versions of the language of second-order logic (Prawitz 1965: 63–6). In the second-version, the second-order variables do not directly replace predicative expressions such as ' $x = a \lor x = b'$, but instead replace terms formed by applying a variable-binding operator λ to predicative expressions, so as to yield in this case ' λx : $x = a \lor x = b'$. Prawitz gives harmonious introduction and elimination rules for the λ -operator which in the simplest cases reduce to:

$$\frac{\lambda \text{-intro:}}{\frac{A_t^x}{(\lambda x:A)t}} \quad \frac{\lambda \text{-elim:}}{\frac{\lambda x:A}{A_t^x}}$$

In these rules, t is an arbitrary term, and A_t^x is the result of replacing every free occurrence of the variable x in A with the singular term t.

With λ -terms in play, the rule for introducing the monadic secondorder quantifier may be formulated as follows:

$$\exists^{2}\text{-intro:} \quad \frac{A_{T}^{X}}{\exists X A}$$

where A_T^X is the result of replacing every free occurrence of the second-order variable X in A with the λ -term T. Now it is easy to prove that these rules yield the instance of comprehension mentioned above, viz.

 $\exists X \; \forall x \; (Xx \leftrightarrow x \neq x).$

Instead of labouring to find a plural rendering of that formula which makes it come out true, however, we may instead emend the λ -rules so as to block its derivation. Thus, we could have as the introduction rule:

$$\frac{\lambda \text{-intro: } A_{t}^{x} \exists x A}{(\lambda x: A)t}$$

and the corresponding two figures of elimination:

$$\frac{\lambda \text{-elim:}}{A_t^x} \quad \frac{(\lambda x: A)t}{\exists x A}$$

When the λ -operator is regulated by these new rules, the term ' λx : Fx' may be understood to be a term having the sense of 'the thing or things, each of which is F and of which every F is one', or, more briefly, 'the F-thing or things'. Correspondingly, ' $\exists X A$ ' may now be read as saying simply 'Some thing or things is or are . . .', where the lacuna is filled in by a translation of the formula A. ' $(\lambda x: Fx)$ y' will be a well-formed formula meaning 'y is either the F-thing or is one of the F-things'. This last provision preserves a striking feature of Boolos's own interpretation of the second-order formalism. Although he reads 'X' as 'they', and 'x' as 'it', 'Xx' remains a well-formed formula, and not simply a translation of the list of words 'they it'. It is well-formed because a user of the formal language who places a first-order variable directly after a second-order variable is understood to be saying 'it is one of them'. I argued earlier that 'is one of' was a semantically significant notion.

106

In Boolos's system, it is a semantically significant notion which—as someone almost put it—is said but not shown.

Our reading of ' λx : Fx' as 'the F-thing or things' shows that our translation scheme is not liable to an objection that has been levelled against Boolos's. I have already mentioned one of Yi's criticisms of Boolos's scheme of paraphrase. He also deems this scheme 'objectionable...because applying the scheme to sentences like "John and Carol carry a piano" or "A book is written by Russell and Whitehead" fails to preserve the fact that they share predicates with sentences like "John carries a piano" or "A book is written by Cicero"' (Yi 1999: 185). I agree with Yi that a satisfactory logic of plurals must respect the point that what Cicero did to Pro Milone by himself, Russell and Whitehead did to Principia Mathematica together. As we have seen (§III), such a logic is necessary if we are to assign the natural interpretation to WH-questions, such as 'Who wrote Principia Mathematica?' It is also needed if we are to respect the point that many singular statements exclude alternative statements involving plural terms. The claim 'I developed such-and-such an idea by myself' jostles for logical space with the claim 'We developed it together', and I expect I am not alone in knowing of academic friendships that have been broken by such jostling. Boolos's scheme of translation makes it mysterious how such claims could exclude one another. For he wants to regiment the singular claim using a first-level predicable and the plural claim using a secondlevel predicable. On the present proposal, there is no mystery. Each occurrence of 'reign' in 'William and Mary reigned together, and then William reigned by himself' will be regimented by a second-level predicate that takes a λ -term as argument. The first λ -term will have the form ' λx : $x = a \lor x = b$ '; the second will have the form ' λx : x = a'. To be sure, this latter λ -term is still not a singular term in the formal language. But it is far from clear that it ought to be. For when we extract the singular term 'William' from the vernacular sentence 'William reigned by himself', we are not left with the verb 'reigned' alone. Rather, we are left with the expression '... reigned by himself'.

In fact it is a merit, not a disadvantage, of this sort of theory that we can use the extraction of singular terms from higher-order formulae to represent logico-semantic distinctions between different sorts of verb and adjective. One important distinction is between verbs and adjectives that *distribute* over plural terms and those that

do not. Is it the case that they are *F* iff each of them is *F*? Where the answer is yes, we may introduce a first-level predicate letter to regiment singular occurrences of '*F*' and understand plural instances such as 'They are *F*' to mean 'Any object that is one of them is *F*'. Thus in the distributive case, we start with a first-level predicate letter '*F*¹'; a related second-level predicate letter '*F*²' may be explicitly defined by the principle: $F^2(X) \leftrightarrow \forall x \ (Xx \to F^Tx)$. Where the answer is no, so that the relevant verb or adjective does not distribute, we shall regiment the expression in question using a second-level predicate letter that will take a λ -term as argument whether the vernacular verb or adjective is in the singular or the plural. Thus, in regimenting the sentences 'Hercules is carrying the piano', and 'John and Carol are carrying the piano', we shall use formulae such as

(1) $C(\lambda x: x = a)$

(2)
$$C(\lambda x: x = b \lor x = c).$$

The letter 'C' is a monadic second-level predicate letter in Prawitz's alternative formulation of second-order logic. It takes a λ -term as argument to form a complete formula. Given our way of reading λ -terms, the formula (1) means 'The thing or things, each one of which is identical with Hercules, is or are carrying the piano'. Since Hercules is the only thing identical with Hercules, this has the sense of 'Hercules is carrying the piano by himself'. Applying our gloss for λ -terms to formula (2), we reach 'The thing or things, each one of which is identical either with John or Carol, is or are carrying the piano'. The fact that 'C' is common to both formulae reflects the fact that the vernacular sentences 'Hercules is carrying the piano', and 'John and Carol are carrying the piano' share a predicate.

Although the full regimentation of the singular sentence 'Hercules is carrying the piano' involves a second-level predicate, first-level predicates may be extracted from that regimentation. Accordingly, the first-order calculus may be used in validating certain arguments involving such a singular sentence, in accordance with Quine's maxim of shallow logical analysis. So, for example, we may extract the singular term 'a' from formula (1) to obtain the first-level predicate ' $C(\lambda x: x = \xi)$ '. This will have the sense of ' ξ is carrying the piano by himself'. Since the object *a* is one of the thing or things identical with *a*, formula (1) entails

(3) $\exists X (C(X) \land Xa)$

Similarly, since the object b is one of the thing or things identical with either b or c, formula (2) entails

(4) $\exists X (C(X) \land Xb).$

We may extract a common first-level predicate from (3) and (4): namely, ' $\exists X \ (C(X) \land X\xi)$ '. This predicate has the sense of ' ξ is carrying the piano, either by himself or together with another or others'.

A further merit of the theory is that it extends without strain to cover so-called multigrade predicates: those, such as '... have a common ancestor', whose argument place may be filled by a list of two or more names. Thus 'Tom, Dick, and Harry have a common ancestor' may be regimented in the form

 $C(\lambda x: x = a \lor x = b \lor x = c).$

Clearly, the disjunction used to form the λ -term may be extended to include as many singular terms as we please.

Although the extraction of first-level predicables may give some reassurance that the first-order calculus remains of some use in validating ordinary arguments involving verbs, adjectives, and singular quantifiers, the revised scheme of translation may seem to involve an alarming intrusion of second-order apparatus into places where one would not expect to find it. Against that, however, it is worth pointing out that the use of λ -terms provides an extraordinarily economical treatment of some inferences whose validation requires additional logical rules on alternative approaches. Yi's own logic of plurals provides a convenient comparison (Yi, 2002). His formal language for plurals introduces two new logical constants: a term-forming operator on terms '@' that is intended to capture the 'and' that we find in 'William and Mary', and a relational expression ' η ' that is intended to mean 'is one of or are some of'. And his logic of plurals duly includes axioms involving these constants. Thus axiom schemata (19) to (22) in his logic of plurals are as follows:

- (19) $\eta (\alpha, \beta @ \gamma) \leftrightarrow [\eta (\alpha, \beta) \lor \eta (\beta, \gamma)]$, where α is a singular term
- (20) $\eta \ (\alpha, \beta @\gamma) \leftrightarrow [\alpha = \beta \lor \eta \ (\beta, \gamma)]$, where α and β are singular terms

- (21) $\eta(\alpha, \beta) \rightarrow \alpha = \beta$, where α and β are singular terms.
- (22) $\exists \nu \eta (\nu, w)$, where w is a plural variable (Yi, 2002: 48).

Using our translation scheme with λ s, however, no new logical symbols are needed to express 'and' and 'is one of or are some of'. And the logical truths expressed by instances of Yi's schemata (19)–(22) are already derivable using our revised rules of λ -introduction and λ -elimination. Thus we may express '*a* is one of *b* and the *F*-thing or things' by writing simply '(λx : $x = b \vee Fx$)*a*'. The λ -rules suffice to establish the equivalence between this formula and '*a* = $b \vee (\lambda x: Fx)a'$. Of course, this translation scheme is going to appal those logicians who demand a symbol to correspond to each logically interesting notion. For Boolos's idea, in total contrast, was to express 'is one of' simply by placing a singular term next to a predicate (or in our case a λ -term). For myself, I confess to being attracted by the Zen-like economy he achieved by rendering a crucial logical notion completely invisible.

The crucial point for present purposes, though, is that there is nothing in the purified Boolosian scheme of regimentation to which a referentialist about plural terms need object. The scheme respects the intuition that what Louis did to France by himself, William and Mary did to England together. And it equally respects the claim that what the proper name 'Louis' does to Louis by himself, the compound name 'William and Mary' does to William and Mary together. When Boolos's ideas are refined and amended in the way that they need anyway to be refined and amended, they lead to a position with which a referentialist about plural terms need have no quarrel. Indeed, they can be used to formalize the derivations of truth conditions for 'Louis XIV reigned over France' and 'William and Mary reigned over England' that I gave at the end of section III. Thus the derivation of truth conditions for the first of these sentences may be formalized as follows, in which italicized expressions are second-level predicables:

- 1. True ('Louis XIV reigned over France') \leftrightarrow $\exists X [Des ('Louis XIV', X) \land Sats (X, 'reigned over France')]$
- 2. $\forall X [Des (`Louis XIV', X) \leftrightarrow \forall y (Xy \leftrightarrow \lambda x (x = l) y)]$
- 3. True [('Louis XIV reigned over France') \leftrightarrow *Sats* ($\lambda x \ (x = l)$, 'reigned over France')]
- 4. Sats $(\lambda x \ (x = l), \text{`reigned over France'}) \leftrightarrow$ Reigned over France $(\lambda x \ (x = l))$

5. True ('Louis XIV reigned over France') \leftrightarrow

Reigned over France $(\lambda x \ (x = l))$

Similarly, truth conditions for 'William and Mary reigned over England' may be derived as follows:

- 1. True ('William and Mary reigned over England') $\leftrightarrow \exists X \ [Des ('William and Mary', X) \land Sats (X, 'reigned over England')]$
- 2. $\forall X \ [Des (`William and Mary', X) \leftrightarrow \forall y \ (Xy \leftrightarrow \lambda x \ (x = w \lor x = m)y)]$
- 3. True [('William and Mary reigned over England') \leftrightarrow *Sats* ($\lambda x(x = w) \lor x = m$), 'reigned over England')]
- 4. Sats $(\lambda x(x = w \lor x = m)$ 'reigned over England') \leftrightarrow Reigned over England $(\lambda x(x = w \lor x = m))$
- 5. True ('William and Mary reigned over England') \leftrightarrow Reigned over England ($\lambda x(x = w \lor x = m)$)

These derivations make it manifest that singular and plural terms contribute to determining truth conditions in the same way, and thereby constitute a single semantical kind.

VII

Although I have been anxious to refute certain bad objections to it, I cannot claim that the Boolosian project of co-opting second-order syntax will provide the most illuminating way of accounting for the validity of arguments involving plurals. Before such a claim could be made, several large issues would need to be addressed. What exactly do we expect of a regimentation of a vernacular argument? Are there arguments involving plurals that are clearly invalid, but of whose invalidity a follower of Boolos can give no account because the standard second-order consequence relation cannot be axiomatized? These matters are delicate, and pursuing them would take me far beyond my present enterprise of defending a referentialist view of plural terms. When they are pursued to the end, limitations may emerge in the strategy of co-opting a second-order formal syntax for the purpose of formulating the rules of plural inference. I shall conclude, though, by showing how that strategy can at least guide the first few steps into a field which has not yet been fully mapped out—namely, the way plurals interact with modal operators.

A good way into this field is provided by an interesting paper by Phillip Bricker (1989), who considers such sentences as 'Some person in the room must win the prize'. As Bricker notes, this sentence exhibits at least a threefold ambiguity. There is the *de dicto* reading, on which it attributes necessity to the proposition that some person in the room shall win. And there is the familiar singular de re reading, on which the sentence says that some person in the room has the property that he will necessarily win the prize. In addition, though, there is a plural *de re* reading, where the sentence says of the people actually in the room that it must be the case that at least one of them wins the prize. To see the difference between this last reading and the other two, Bricker invites us to suppose that Tom, Dick, and Harry are the people in the room, and that the draw has been rigged by removing all tickets belonging to the other entrants. In that case, the sentence is true under the plural *de re* reading: the people in the room are such that at least one of them must win the prize. The other two readings, however, are false. Since it is a contingent property of any of Tom, Dick, and Harry that he is in the room, and since a person's presence in the room is irrelevant to the award of the prize, the dictum 'Some person in the room must win the prize' is not a necessary truth. Furthermore, the singular de re reading is also false, for there is no particular individual who is guaranteed to win. Of course, the familiar de dicto and de re readings do not exhaust the possibilities of analysis using ordinary singular quantifiers. Bricker, however, considers in some detail variants using actuality operators and the like, before concluding that none of these will work, and that in order to formalize the plural de re reading, one needs a formal language that contains a device of plural quantification. Thus he proposes that the plural *de re* reading may be captured as follows:

 $\exists X (\forall y (Xy \leftrightarrow Fy) \land \Box < \exists x > (Xx \land Gx))$

Here '*Fξ*' symbolizes ' ξ is a person in the room', '*Gξ*' symbolizes ' ξ wins the prize', ' \Box ' signifies the variety of necessity expressed by the 'must' of 'Some person in the room must win the prize', and the brackets around the existential quantifier indicate that it has the so-called 'outer' reading, so that the variable 'x' is free to range over objects that exist at any possible world.⁷

⁷ A regimentation using an actually operator 'A' will yield the right truth conditions in the present, very simple case: $\Box < \exists x > (AFx \land Gx)$. But no such regimentation will be available when modal operators are iterated. Thus suppose that the

I think that this symbolization is fine. We need, however, to address an apparent problem with it.

We may begin by noting that the symbolization will not yield the plural *de re* reading unless the variable 'X' is understood in such a way that the object or objects assigned to it are held fixed across different possible worlds: in any of the relevant possible worlds, it must be one of Tom, Dick, or Harry who wins the prize. Our formulation of this principle, however, needs to allow for the fact that the objects in question may fail to exist in certain possible worlds. In allowing for this in the ordinary, singular, quantified modal logic, it is customary to invoke a first-level predicate of existence; in what follows, it will not matter whether this predicate, which may be written ' $E\xi$ ', is taken to be primitive or is defined in the usual way as ' $\exists y \xi = y$ '. For once we have this predicate, we may define a corresponding plural existence predicate ' $E^2\varphi$ ' as follows:

 $E^2 \varphi \leftrightarrow \forall y (\varphi \ y \rightarrow Ey)$

With the plural existence predicate in play, we may formulate one of the principles that captures constancy in the interpretation of the second-order variables as follows:

(M)
$$Xx \rightarrow \Box (E^2X \rightarrow Xx)$$

The proposed formalization of the plural *de re* presumes that principle (M) is a validity of the modal logic of plurals. That is to say, it presumes that the following is a validity of the modal logic of plurals: if an object is one of them, it is necessarily one of them if they exist. I have formulated (M) using free variables, which are understood to range over all possible objects.

In reading formula (M), it is of the last importance not to be distracted by alternative readings of the second-order formalism. On Dummett's way of understanding second-order quantification,

situation is as described, save that (a) Tom, Dick, and Harry are not in the room, but could have been; (b) Laurel and Hardy are in fact in the room; (c) there is no possibility of rigging the draw so that one or other of Laurel and Hardy must win. In these circumstances, it will be true to say, 'It is possible that some person in the room must win the prize', and the modalization of the plural regimentation duly comes out true:

 $\Diamond \exists X (\forall y (Xy \leftrightarrow Fy) \land \Box < \exists x > (Xx \land Gx))$

However, the modalization of the regimentation using 'actually'—viz. $\Diamond \Box < \exists x > (AFx \land Gx))$ —is false.

for example, (M) has no plausibility whatever. On that reading, (M) says that if an object x is something that any object either is or is not, then it is impossible for x not to be that thing in a circumstance where every object that is that thing exists. Now any object, we may suppose, either is or is not in New Haven on 5 March 2004. Since I was in New Haven on 5 March 2004, accepting (M) (as Dummett would read it) would commit me to the impossibility of my not having been in New Haven then, in a circumstance where every object that was then in New Haven exists. So accepting (M) (as Dummett would read it) would involve denying most forms of contingency. This brings out how starkly our Boolosian reading of the second-order formalism diverges from more traditional interpretations when modal operators come into play. In fact, Aaron Sloman noted the corresponding divergence for functions nearly thirty years ago:

If we consider the expression 'the town in which x was born', then we may say that there is a function corresponding to it which correlates (some, but not all) persons with towns. Suppose that Aristotle's first pupil, whoever he was, was born in Athens. Then Athens is the value of the function for that man as argument. But that man might have been born elsewhere, for example if his mother had decided to go on holiday just before his birth. In that case a different town would have been the value for the same man as argument. But a value of *what*? A different town could not be the value of the same *function*, for then the set of ordered pairs would be different, and so, since a function just is a set of ordered pairs (or at any rate something satisfying extensional criteria of identity), it would be a different function. Hence, if, as seems quite natural, we wish to say that the same something might have had a different value for the same argument, we must regard the 'something' as not satisfying extensional criteria for identity. (Sloman 1965: 159)

Sloman's use of 'something' corresponds to the use Dummett makes of it in his formula 'something which every object either is or is not'. Indeed, on a high Fregean view, whereby concepts are functions, Dummett's use will be a special case of Sloman's. (M), then, exemplifies a point that I have been at pains to emphasize namely, how radically Boolos's way of reading the second-order quantifiers diverges from more traditional conceptions of them.

Formula (M) may be read as expressing the necessity, or internality, of the relation of being one of. And if the underlying propositional modal logic contains the rule of necessitation, and the schemata

(K)
$$\Box (\varphi \rightarrow \psi) \rightarrow (\Box \varphi \rightarrow \Box \psi)$$

and

(B) $\varphi \rightarrow \Box \diamondsuit \varphi$

or equivalently

(B) $\neg \varphi \rightarrow \Box \neg \Box \varphi$

then (M) entails a form of the necessity, or internality, of the relation of not being one of. The entailment is derived as follows:

$I. Xx \to \Box (E^2 X \to Xx)$	(M)
2. $\neg \Box (E^2 X \rightarrow X x) \rightarrow \neg X x$	1, contraposition
3. $\Box (\neg \Box (E^2 X \rightarrow X x) \rightarrow \neg X x)$	2, necessitation
$4. \ \Box \neg \Box (E^2 X \rightarrow X x) \rightarrow \Box \neg X x$	3, (K)
5. $\neg (E^2 X \rightarrow X x) \rightarrow \Box \neg \Box (E^2 X \rightarrow X x)$	(B)
$6. \neg (E^2 X \rightarrow X x) \rightarrow \Box \neg X x$	4, 5 predicate calculus
7. $(E^2X \land \neg Xx) \rightarrow \Box \neg Xx$	6, predicate calculus

This derivation runs parallel to Prior's well known derivation (Prior 1955: 206–7) of the necessity of distinctness, (D), from the necessity of identity, (I):

(I)
$$x = y \rightarrow \Box (Ex \rightarrow x = y)$$

(D)
$$(Ex \land \neg x = y) \rightarrow \Box \neg (x = y)$$

Principle (I) will be a theorem of any sensible quantified modal logic in which the quantifiers have their normal interpretations and in which the underlying propositional modal logic is normal. Even a free quantification theory will deliver:

8. $Ex \rightarrow x = x$

so that necessitation will yield

9. \Box (*Ex* \rightarrow *x* = *x*)

If the first-order variables have their ordinary interpretation, then we shall have

io. $x = y \rightarrow \Box (Ex \rightarrow x = y)$

which is (I). There are formally tractable systems of quantified modal logic that do not contain (D). The non-modal laws of identity also hold good if the equals sign is taken to express indiscernibility,

i.e. if the distinctness sign is taken to express discernibility; and two things can be discernible without being necessarily discernible. (See Stalnaker 1995: 22–3.) This formal point, however, does not cast doubt on the truth of (D) when '=' is given its intended interpretation. Timothy Williamson (1996) has given an alternative derivation of (D), which relies upon plausible principles concerning actuality, rather than on the Brouwerian schema (B). It, too, could be adapted to provide an alternative derivation of proposition (7) above.

For these reasons, I think that we should accept proposition (7), which I shall call principle (N). It says that if they exist, and if an object is not one of them, then it cannot be one of them. In other words, only an object that is actually one of them can possibly be one of them. Like (M), (N) is presupposed by our way of regimenting a plural *de re* statement. Having assigned Tom, Dick, and Harry to be values of the variable 'X', we need these to be the only possible objects that could serve as values under that interpretation.

Even the truth of (M) (on the emended Boolosian reading), however, may still seem to generate a paradox. Let ' $F\xi$ ' be a first-level predicate letter that regiments ' ξ is a Jerseyman in New Haven on 5 March 2004'. Then, it seems, we may instantiate the variable 'X' in (M) with the λ -term ' λy : Fy' to reach

11.
$$(\lambda y: Fy) x \rightarrow \Box (E^2 (\lambda y: Fy) \rightarrow (\lambda y: Fy)x)$$

Now let '*a*' be a singular term designating me. Then from (11) we may infer

12.
$$(\lambda y: Fy) a \rightarrow \Box (E^2 (\lambda y: Fy) \rightarrow (\lambda y: Fy)a)$$

which, given our rules for the λ -operator, in turn yields

13.
$$Fa \rightarrow \Box (E^2 (\lambda y: Fy) \rightarrow Fa)$$

Since I am a Jerseyman who was in New Haven on 5 March 2004, the antecedent of (13) is true. Yet the consequent of (13) is surely false. There are possible worlds in which every Jerseyman who is in New Haven on 5 March 2004 exists, but in which I was elsewhere on that date. The conditional (13), then, is false, despite its being, apparently, a consequence of (M). So how can (M) be true?

The derivation of (13) from (M) runs parallel to a familiar argument involving singular terms. Starting from

(I)
$$x = y \rightarrow \Box (Ex \rightarrow x = y)$$

116

we may instantiate the variable 'y' with the singular description 'lz: Fz' to obtain

14.
$$x = lz:Fz \rightarrow \Box (Ex \rightarrow x = lz:Fz)$$

Again, we may instantiate the variable 'x' with a term 'a' designating me to obtain

15.
$$a = lz:Fz \rightarrow \Box (Ea \rightarrow a = lz:Fz)$$

Suppose, though, that I was the only Jerseyman in New Haven on 5 March 2004. Then the antecedent of (15) is true. However, its consequent is surely false. It is possible for me to exist without having been the only Jerseyman then in New Haven.

While the derivation of (15) from (I) may have presented some kind of paradox forty years ago, few will finding it puzzling today. The generally accepted solution is that (I) is true, but that (14) does not follow from it. Within a modal context, we may not instantiate a first-order variable with an arbitrary singular term. Rather, such instantiations are legitimate only if the instantiating term is a rigid designator, and on the interpretation specified the singular definite description '*lz: Fz*' is not rigid. A rigid designator designates the same object with respect to every possible world (or, perhaps, with respect to every possible world where that object exists). But '*lz: Fz*' designates various objects depending on who (if anyone) is uniquely a Jerseyman in New Haven on the relevant date.

I suggest that a comparable solution dispels any suggestion of paradox in the plural case. On this view, (M) is true, but (11) does not follow from it. Within a modal context, we may not instantiate a second-order variable with an arbitrary λ -term. Rather, such instantiations are legitimate only if the instantiating λ -term is rigid. A λ -term is deemed to be rigid if it designates the same object or objects with respect to every possible world (or, with respect to every possible world where those objects exist). This condition will be met just in case the same object or objects satisfy its component predicate with respect to every possible world (or, with respect to every possible world where those objects exist). On the specified interpretation for 'F', then, the λ -term ' λz : Fz' is not rigid. The objects it designates vary from world to world.

But are there any rigid plural terms? We can prove that there are some. For we can prove that when two singular terms are rigid, the compound name formed by concatenating them using 'and' is also

rigid. Thus, suppose that the proper name 'William' (which we shall regiment using 'a') is rigid. And suppose too that the proper name 'Mary' (which we shall regiment using 'b') is rigid. Then we can prove that the compound name 'William and Mary' (which will be regimented using the λ -term ' λx : $x = a \lor x = b$ is rigid.

Before we can give the proof, we need a criterion for rigidity. The usual definition of the notion for singular terms—a term is rigid if it designates the same object with respect to every possible world in which that object exists—mentions the expression. But that definition corresponds to two feature of its use. In the first place, where a is a rigid singular term, we have that any object that is identical with a will be identical with a in any possible circumstance in which it exists. That is to say, if a is a rigid singular term, we must be able to assert the relevant instance of the schema:

 $(R \square a) \ x = a \rightarrow \Box (Ex \rightarrow x = a)$

We also have the principle that no possible object could be identical with a unless it is actually identical with a. That is to say, if a is a rigid singular term, we must be able to assert the relevant instance of the schema:

 $(R I b) \diamondsuit (x = a) \rightarrow x = a$

In these schemata, the quantifiers should be understood to range over all possible objects.

Similarly, where T is a rigid λ -term, we shall be able to assert the relevant instance of the schemata:

$$(R2a) Tx \to \Box (E^2(T) \to Tx))$$

and

 $(R_2b) \diamond Tx \rightarrow Tx$

Again, the quantifiers are understood to range over all possible objects.

Standard logical rules ensure that a compound name's rigidity is guaranteed by that of the component singular terms. Thus, suppose that the singular terms a and b are both rigid. Then the relevant instances of (Ria) entail the corresponding instance of (R2a), and so similarly for (Rib) and (R2b). The derivation in the first case runs as follows:

16.
$$x = a \rightarrow \Box (Ex \rightarrow x = a)$$
 instance of (*R1a*)
17. $x = b \rightarrow \Box (Ex \rightarrow x = b)$ instance of (*R1a*)

18.
$$x = a$$
supposition19. $\Box (Ex \rightarrow x = a)$ 16, 18, modus
ponens20. $(Ex \rightarrow x = a) \rightarrow (Ex \rightarrow x = a \lor x = b)$ predicate calculus21. $\Box [(Ex \rightarrow x = a) \rightarrow (Ex \rightarrow x = a \lor x = b)]$ 20, necessitation22. $\Box (Ex \rightarrow x = a \lor x = b)$ 19, 21 (K)23. $\Box (Ea \rightarrow x = a \lor x = b)$ 18, 22 Leibniz's Law24. $(Ea \rightarrow x = a \lor x = b) \rightarrow$
 $(Ea \land Eb \rightarrow x = a \lor x = b)$ predicate calculus25. $\Box [(Ea \rightarrow x = a \lor x = b) \rightarrow$
 $(Ea \land Eb \rightarrow x = a \lor x = b)]$ 24, necessitation26. $\Box (Ea \land Eb \rightarrow x = a \lor x = b)$ 23, 25 (K)27. $(Ea \land Eb \rightarrow x = a \lor x = b) \rightarrow$
 $(Ay: y = a \lor y = b) \rightarrow$
 $(Ay: y = a \lor y = b) x)$ predicate calculus28. $\Box [(Ea \land Eb \rightarrow x = a \lor x = b) \rightarrow$
 $(Ay: y = a \lor y = b) x)]$ 26, 28 (K)
 $(Ay: y = a \lor y = b) x)$ 29. $\Box (E^2 (Ay: y = a \lor y = b) x)$
 $(Ay: y = a \lor y = b) x)$ 26, 28 (K)
 $(Ay: y = a \lor y = b) x)$ 30. $x = a \rightarrow \Box (E^2 (Ay: y = a \lor y = b)) \rightarrow$
 $(Ay: y = a \lor y = b) x)$ 18, 29 discharging
supposition (18)

A parallel argument starting from (17) yields

31.
$$x = b \rightarrow \Box (E^2 (\lambda y; y = a \lor y = b) \rightarrow (\lambda y; y = a \lor y = b) x)$$

and thence

32. $x = a \lor x = b \rightarrow$	30, 31 predicate
$\Box \left(E^2 \left(\lambda y \colon y = a \lor y = b \right) \rightarrow \right.$	calculus
$(\lambda y: y = a \lor y = b) x)$	

33. $(\lambda y: y = a \lor y = b) x \rightarrow$ $\Box (E^2 (\lambda y: y = a \lor y = b) \rightarrow$ $(\lambda y: y = a \lor y = b) x)$ 32. λ -rules

Formula (33), however, is an instance of schema (*R2a*), with the schematic letter '*T*' replaced by ' λy ($y = a \lor y = b$)'. That λ -term, however, is our way of regimenting a compound name such as 'William and Mary'. So the derivation just given shows that, given standard principles of modal logic, two singular terms each of which satisfies schema (*R1a*) will combine to form a compound

name which satisfies schema (R_{2a}). A similar derivation shows the same for the *b*-schemata. Together, then, these derivations show that a compound name will inherit its rigidity from the rigidity of its component singular terms. A compound name formed from rigid singular designators designates the same object or objects with respect to every possible world.

Are our other varieties of plural term rigid? What, in particular, about a collective name, such as 'the Channel Islands'? This expression does not mean 'the islands in the English Channel'. The Isle of Wight is an island in the English Channel, but it is not one of the Channel Islands. Rather, the term refers to the islands in a certain archipelago off the western coast of Normandy. Despite its not being a plural definite description, however, it is not clear that it is rigid.

For let us consider—case (1)—a possible world in which all the Channel Islands save a smaller one—let it be Herm—are in their actual locations, but in which Herm lies fifty miles to the west of the rest of the archipelago. Then one might be tempted to describe this as a possible world in which (a) the Channel Islands exist, but in which (b) Herm is not one of them. It is too far from the others to qualify as a Channel Island. If this description were accepted, then while Herm is actually one of the Channel Islands, it would be possible for all of them to exist without its being one of them. That is to say, we should have a situation which provides a counter-example to the relevant instance of (R2a). The expression 'the Channel Islands', then, would not qualify as rigid.

Or again, let us consider—case (2)—a possible world in which all the Channel Islands exist, and are related to one another as they are actually related, but in which there is an additional island—call it 'Atlantis'—in the middle of the archipelago.⁸ Then one might be tempted to describe this as a possible world in which an object namely, Atlantis—is a Channel Island. We would then have a possible object, Atlantis, which is possibly a Channel Island, even though it is not actually a Channel Island. But if this description were accepted, then we should have a counter-example to the relevant instance of (R2b). Again, the expression 'the Channel Islands' would not qualify as rigid.

Let us also consider plural demonstratives. In remonstrating at your indiscretion in relaying to a crowd some gossip about Smith, I might say, 'You shouldn't have said that. If Smith hadn't been delayed,

⁸ I owe this example to Michael Dummett.

Plural Terms

he would have been one of those people'.⁹ Here—case (3)—we seem to have an actual person who is possibly one of those men, even though he is not actually one of them. If so, then we have a counter-example to the relevant instance of (R_2b) . So the plural demonstrative 'those people' seems not to be rigid.

But while these considerations have some initial plausibility, they are far from being conclusive.¹⁰

To the first consideration, it may be replied that it is essential to a geographical item such as an island that it should be, more or less, in its actual location. If this is so, then the proffered description of case (I) cannot be correct. An island that lies fifty miles to the west of the Channel Islands is not Herm, however closely it may resemble Herm. So the proper description of case (I) is that in it, Herm does not exist. But if that is the right description, then not all the Channel Islands exist, so case (I) provides no counter-example against the thesis that 'the Channel Islands' is rigid.

To the second consideration, it may be replied that the claim that Atlantis is possibly a Channel Island may be sustained only by surreptitiously treating 'the Channel Islands' as a definite description rather than a genuine plural term. For-the thought runs-the ground for claiming that Atlantis is possibly a Channel Island in case (2) can only be that it is an island which lies in the same archipelago as Jersey, Guernsey, Alderney, and the rest. Now perhaps our knowledge that the term is associated with a description of this kind is what enables us to latch on to its sense. But-the thought continues—it is a description which fixes the reference of the term, rather than one which gives its meaning. (The distinction is due to Kripke, of course. See his 1980: 53-60.) By which is meant: the reference is fixed, for all possible worlds, as comprising those islands which actually belong to that archipelago. Atlantis is no such island. So it is not a possible Channel Island. Of course, if the reference of the term 'the Channel Islands' were to be fixed afresh, in the imagined possible world, as comprising all and only those islands lying in the same archipelago as Jersey, Guernsey, Alderney, etc., then the statement 'Atlantis is one of the Channel Islands' would come out true. However-the thought concludes-this does not show that the statement 'Possibly, Atlantis is one of the

⁹ I owe this example to Dorothy Edgington.

¹⁰ In the next few paragraphs, I am much indebted to discussion with Robin Jeshion and Michael Nelson.

Channel Islands' is true, as we *actually* use the expression 'the Channel Islands'. But only the latter thesis threatens the claim that the term 'the Channel Islands', as we actually use it, is rigid.

This last distinction may also be invoked in replying to the third objection. Smith, the reply goes, could not be one of those people. For if he were to be one of those people, there would have to be a correct answer to the question, which one he was, and—given that he is not actually among them—there is no such answer. If we are tempted to suppose otherwise, that will be because we surreptitiously imagine a fresh use of the demonstrative, made in circumstances in which Smith has joined the throng. On this counterfactual use of the demonstrative, 'Smith is one of those people' would be true. But again this does not show that 'Possibly, Smith is one of those people' is true, as the plural demonstrative was actually used. That, however, is what one would need to show in order to show that 'those men' is not rigid.

For these reasons, I regard the thesis that plural terms are rigid as a plausible conjecture which has yet to be refuted. Further investigation would be needed to decide whether it really is true, but now that we have a framework for considering the question, we need not rush to judgement. Our assessment of the semantic theory of The Varieties of Reference, however, will depend heavily on the eventual answer. Evans observed that in treating paradigm referring expressions such as proper names, we need not relativize reference to a possible world. Were we to classify definite descriptions as referring expressions, however, we would need to relativize the relation of reference in that way. These observations constitute his main argument against treating descriptions as referring expressions: the grouping that results from this treatment does 'not correspond to any natural semantical kind' (Evans 1982: 57). It should now be clear that this conclusion is inadequately supported: Evans's first observation is made on the basis of too narrow a sample of putative paradigms. Plural terms are referring expressions. But it is not yet clear whether they are all rigid. If some of them are not rigid, then we shall need to relativize the relation of reference to accommodate them. And then Evans's argument against classifying descriptions as referring expressions would fall to the ground.

I must leave the issue of rigidity undecided here. But a moral of the discussion returns us to our starting-point. Plural terms are

Plural Terms

referring expressions, and we shall not attain a satisfactory theory of the varieties even of singular reference unless we widen our gaze and recognize that reference may be effected by them as well as by singular terms.¹¹

¹¹ I wrote a first draft of this paper in the summer of 1997, and have read versions of it to the Philosophical Society and to the Tuesday Group at Oxford, to a conference on plurals at the School of Advanced Study at London University, to workshops on modality at the universities of Geneva and St Andrews, and to a colloquium at Yale.

Abandoning Coreference

KEN SAFIR

It seems that when the term "coreference" is used, whether in linguistics or in philosophy, there is often presumed to be a consensus about what it is, or at least about what it is in the context where the term is introduced. I don't think the term deserves to have much use at all, insofar as it disguises more interesting linguistic and pragmatic relations between nominal forms in natural language. My preoccupation with these relations issues in part from some of the central concerns and distinctions introduced in Evans (1980), an essay that has had wide and, in some cases, a very deep influence on how a variety of reference issues have been addressed in modern linguistics. As a linguist, my interest is in the way in which natural language shapes what we know, and from that perspective I want to understand how natural language sets boundary conditions on how linguistic forms can be used to achieve readings that pick out the same entity more than once.

From this perspective, it is perhaps useful to distinguish two notions of coreference at the outset: one with a fairly limited use and one that has a use that I believe turns out to be misleading. The first notion of "coreference" simply means that two terms pick out the same individual. If John says "I like Ralph" and Jill says "I like Sam" and Alice knows that Ralph and Sam are one and the same, then the terms "Ralph" and "Sam" corefer for Alice because they pick out the same entity in discourse, though neither John nor Jill

My thanks to José Luis Bermúdez for giving me the opportunity to comment on the work of a philosopher I have long admired (but never met). I would also like to thank both José Luis Bermúdez and Stephen Neale for some very useful comments that helped me figure out how to present what I had in mind to say. The usual disclaimers apply.

has any intention that they should do so. Insofar as this meets the description of a situation where two terms pick out the same individual, this is a fairly uninteresting case of *coreference*.

What I am told philosophers more typically mean by "coreference" I will consistently call intended coreference. Intended coreference occurs when there is an intention to use two or more terms to refer to a single entity. It is my position that linguists shouldn't have much to say about intended coreference, first because I don't think that any form of intended reference is a matter for linguists at all, and second because the identity relation that can hold between two linguistic forms is flexible in a way that most notions of coreference do not, or cannot, address. For this reason, I will usually refer to identity relations by a presentationally neutral term, coconstrual, by which I mean some sort of identity relation or other between nominals. Once the issues are clarified by separating speaker intentions from coconstrual, the variety of coconstrual relations, some of which are crucially distinguished by Evans, then provide evidence against the utility of every usage of (intended) coreference that I am familiar with.

I.O WHAT APPEARS TO BE AT STAKE

So much of Evans's perspective on matters of reference is tied to what he took to be linguistic consequences of intentions to refer that it might seem odd that I would use a piece of his work to establish the view that intended reference is irrelevant to linguistics. Evans, however, writes as a linguist as much as a philosopher in "Pronouns", and as a linguist, I intend to take him to task even as I take from him my main line of argument.^I

Evans (1980), which is specifically intended for a linguistics audience, is presented, at least in part,² as a reasoned objection to a sea change in linguistic theory, one introduced by Lasnik (1976), by which the goal of explaining patterns of coconstrual, particularly pronouns coconstrued with their antecedents, shifted from rules

² Of course, this is not the main point of Evans's article, which is to defend a particular view of the semantics of pronouns based on a Fregean approach to interpretation.

¹ Evans (1977: 115) explicitly states: "I am interested in the quantifiers and pronouns that occur in the English natural language (and many others, I bet). I am not interested in the quantifiers and devices of back references which exist in logically possible languages which we might speak but do not." Few linguists would describe their interest in natural language, as opposed to artificial languages, differently.

Ken Safir

that license coconstruals to rules that block them. Except for forms like *himself* and *each other*, terms that must be bound in specific syntactic configurations, Lasnik treats all other coconstrual as unenforced, a pragmatic matter, based on what the speaker would know of the context of the conversation, world knowledge, previous discourse, and the like. Evans was protesting against this change by arguing that some linguistic coconstruals of pronouns with their antecedents are enforced by rules of grammar.

Evans also argues that the form of blocked coconstrual that Lasnik posits—namely, noncoreference—is the wrong characterization of the failed coconstrual at issue. Evans observes that intended coreference is perfectly possible for many syntactic contexts where Lasnik's principle would predict noncoreference, once appropriate contexts are invoked. Rather, Evans establishes that what is most typically blocked by linguistic rules is dependent reference (or as I will eventually prefer to call them, *dependent identity* readings), although, as we shall see, this is not the whole story.

Once we have a clear idea of what theories of coreference are supposed to apply to, we can then undermine the term *intended coreference* itself. I argue that if intended coreference involves picking out the same referent, the same extension in the world of discourse, as some previous mention, then it does not describe the class of coconstruals that existing theories of (intended) coreference address, or that they should address. To make this case I will argue (in section 5) that there exist dependent identity readings which are not coreference readings.

In what follows, I will try to make my case against the usefulness of both *intended reference*, and hence *intended coreference*, with minimal appeal to technical notions in linguistics, and I think that this is largely possible. However, I do rely, as Evans did, on the assumption that my readers can sustain a certain peculiar appetite for anaphora puzzles.

1.1 ENFORCED COREFERENCE VERSUS PRAGMATIC Coreference

As the matter is put in Evans (1980), the issue is what linguistically competent average folk can or must know about the relationship between a term that designates or describes what a pronoun refers to and the pronoun itself. In earlier versions of generative grammar, particularly before the mid-1970s, it was assumed that pronouns and reflexives were introduced by rule, replacing tokens identical in form to their antecedents and *thereby* identical in referential value with their antecedents (Lees and Klima (1963) and Langacker (1969)). Thus (1a) would be transformed into (1b) based on the identity of form and referential value assumed to hold between the first mention of *Alice* and the second (where coconstrual is marked by italics).

- 1a. *Alice thinks that Alice is smart.
- 1b. Alice thinks that she is smart.

The reason why (1a) could not be used to mean (1b) was that the rule of pronominalization in this syntactic configuration (where the first *Alice* c-commands the second one—I define c-command later) is obligatory, such that (1a) would always be transformed into (1b), as long as we are talking about the same Alice. Similar transformations would ensure that more locally anteceded forms would be transformed into reflexives.

2a. **Marcus* loves *Marcus*.2b. *Marcus* loves *himself*.

In order for a pronoun or a reflexive to be coreferent (using the term of the time) with its antecedent, the pronoun would have to be introduced by the rules of pronominalization or reflexivization respectively.

Part of the problem of looking at pronominalization in this way, however, is that all of the relationships in (3a-c) allow (intended) coreference where the use of a name in place of the pronoun is also possible.

- 3a. *He/Freddy* is a terrible liar. It's amazing! *He/Freddy* will say anything!
- 3b. Even his/Freddy's mother thinks Freddy/he is lying.
- 3c. When it finally came out that *Freddy/he* had lied to all of *his* wives, *Freddy's/his* current wife decided to turn states evidence.

In other words, pronominalization is optional in all of the examples in (3) and in all of the combinations, although pronominalization is not optional in (1b). It is even possible for the pronoun to precede the name, unless the aforementioned c-command holds, as it would, for these examples, only with respect to the position of *his* in (3c) (which could not be replaced by *Freddy's* because *Freddy/he* c-commands *his*). This difference represented a weakness in enforced coreference theories, in that pronouns seem to allow (intended) coreference optionally wherever pronominalization was not treated as obligatory.³

Another problem with introducing pronouns by rule, first pointed out by Postal (1971), is that some pronouns don't have linguistic antecedents. If we are perfect strangers and we witness a car accident (e.g. driver's car hits parked car), I can turn to you and begin a conversation as in (4).

4. I'm glad I'm not his insurance agent!

To introduce this pronoun by linguistic rule would require positing an abstract unpronounced linguistic antecedent, "the guy who caused the accident" or such like, and then introducing *his* as a pronominalization of the longer unpronounced description by virtue of identity with the unpronounced one. Alternatively, some pronouns have to be introduced into sentences *independently* of the pronominalization rule.

Lasnik (1976) seizes on the problem with cases like (4) to point out that if pronouns have to be introduced in some cases with independently determined reference, what is to prevent all pronouns from being introduced in this way, where coreference arises *by accident*, such that *she* in (5) just happens to pick out the same referent as that picked out by *Alice*?

5. *She thinks that Alice is smart.

Under unexceptional intonation and discourse conditions for examples like (5) (we shall return to exceptional ones), *she* and *Alice* cannot be coconstrued, and, so Lasnik reasons, something must ensure that *Alice* and *she* be prevented from coreferring. Thus Lasnik proposes a *noncoreference* rule, which I simplify as in (6) (which was later installed as Principle C in Chomsky's (1981) Binding Theory).

³ Hornstein (2001) and Kayne (2002) have recently revived the idea that coconstrual should be achieved by syntactic movement operations, but they extend this to every case of coconstrual (even cross-sententially, according to Kayne). I refute this proposal in Safir (2003). 6. Lasnik's Noncoreference Principle: a name cannot be c-commanded by a coreferent NP.

Lasnik's Principle ensures that *Alice* and *she* cannot corefer in (5) regardless of how these tokens are introduced into the sentence structure. A few clarifications with respect to (6) are in order, however. By "a name" what is meant here is any referring expression such as a proper name, a demonstrative, or a definite description, where pronouns are not taken to be names in this sense. C-command is a particular sort of relation between parts of the hierarchical architecture of sentences. This architecture is usually represented with brackets or tree diagrams, where the existence of each node must be justified as part of the linguistic knowledge of adult native speakers. C-command can be simplified for our purposes as follows:

7. A c-commands B if B is a sister to A or B is dominated by a sister of A.

When two syntactic pieces (phrases) are joined into a larger one, then the two phrases joined are sisters, and the new larger phrase they form is said to dominate its two parts. Thus, in the schematic sort of structure provided in (8), everything is gathered into a single unit, the unembedded sentence S_1 , and the complement to the verb *believe* in this instance further embeds another sentence, S_2 . The nominals that we need to talk about in terms of coconstrual are NPs in this system (noun phrases) and I label them as well, but I dispense with some of the other bracket labels (called "node" labels in equivalent tree diagrams) for now.

[S₁ [_{NP}She] [believes [S₂ [_{NP}Tom] [loves [_{NP} [_{NP}Brianna's] [picture of [_{NP}her]]]]]

The subject of S_{I} is *she*, and the sister to *she* is everything else: e.g. *believes Tom loves Brianna's picture of her*. Thus *she* c-commands everything contained in its sister. Similarly, *Tom* c-commands everything contained in its sister: namely, *loves Brianna's picture of her*. Moreover, *Brianna's* c-commands *picture of her*, since these two constituents are sisters (i.e. c-command is more general than subject and predicate, but it is not useful to give further examples here). Given the way I have represented (8), there is no coconstrual at all, since nothing is italicized, so each NP presumably picks out a

Ken Safir

different value. Although we can determine what c-command relations hold in (8), Lasnik's (6) only comes into play if we assume that two or more of the NPs are coconstrued. Consider (9), for example, where *she* and *Brianna* are coconstrued.

9. *[S₁ [_{NP}She] [believes [S₂ [_{NP}Tom] [loves [_{NP} [_{NP}Brianna's] [picture of [_{NP}her]]]]]]

Since *she* c-commands everything else, including *Brianna's*, the coconstrual should fail, be marked noncoreferent in Lasnik's system. Being marked as noncoreferent (assigned a different index, in that system) would mean that no accident of reference could result in a situation where *she* and *Brianna's* could have the same referential value. Chomsky (1976*a*) introduces a similar noncoreference rule for pronouns too close to their antecedents, described in (10) which would rule out both the coconstrual of *her* with *Brianna's* in (11a) and the more garden variety case in (11b).

- 10. A pronoun cannot be c-commanded by coreferent NP in its local domain.
- 11a. *[S_1 [NPShe] [believes [S_2 [NPTom] [loves [NP [NPBrianna's] [picture of [NPher]]]]]
- 11b. *[S *Alice* [loves *her*]]

Since both *Alice* and *Brianna's* c-command the pronouns they are marked as coconstrued with, and since both antecedents are local enough (a matter I won't explore, but let us assume that the right notion of "local domain" can be stated), the pronoun is marked as disjoint in reference with *Alice* in (11b) and with *Brianna's* in (11a).

However, if only *her* were to be coconstrued with *she*, as in (12), nothing would block coreference (i.e. Lasnik's noncoreference rule in (6) would not block it and Chomsky's noncoreference rule for pronouns in (10) would not block it because the c-commanded pronoun is sufficiently far from its antecedent).

12. $[S_1 [NPShe]$ [believes $[S_2 [NPTom]$ [loves [NP [NPBrianna's] [picture of [NPher]]]]]]

Thus it is possible, though not required, for *she* and *her* to be intended coreferent, but no rule is required to ensure that coconstrual is enforced; rather, coconstrual of *she* and *her* is something that is made possible by the grammar, and the rest of the task of

130

assigning values to *she* and *her*, be they the same or different, is a matter for the context of the utterance (pragmatics), not for any rule of grammar. Moreover, all of the other cases where optional replacement of names with pronouns permits coconstrual, as in (3), require no special rule at all, since optional use of a pronoun in place of a name is now just the general sort of accident, unless (6) applies to block it.

Evans (1980) objects to two features of Lasnik's system. He objects to the notion that intended coreference is not enforced by linguistic rule (setting aside the reflexive cases, which everyone assumes are somehow linguistically enforced). He characterizes Lasnik's theory (and that of Chomsky (1976*a*, *b*)) as a "pragmatic theory of coreference" in that coconstruals between pronouns and other nominals are left to non-linguistic factors. Evans also objects to the notion that what is linguistically regulated is (intended) coreference rather than (intended) dependent reference. These concerns converge in his first objection, which targets the relation between coconstrual in (13a) and coconstrual in (13b).

13a. *Every man* loves *his* mother. 13b. *John* loves *his* mother.

In Lasnik's theory, Evans points out, it is a coincidence that the form *his* effects both of these coconstruals, because the relation between *John* and *his* in (13b) is potentially accidental coreference (pragmatically determined), while the coconstrual between *every man* and *his* in (13a) is bound reference, as it must be if it is a coconstrual, because *every man* is not referential.

Evans begins by arguing that the relation between *John* and *his* is not coreference, but asymmetric dependent reference, such that *his* asymmetrically depends on *John*. This asymmetric dependence is what also holds in (13a), where the antecedent is a quantifier. Failure of bound reference is what he argues is responsible for the obviation that Lasnik describes (or misdescribes, if Evans is right) as a noncoreference effect.

At the risk of setting some heads spinning with too much to take in too soon, I foreshadow what is to come. I agree with Evans that the right theory of linguistic coconstrual is one that primarily addresses dependencies and that both (13a) and (13b) are instances of dependency, but I agree with Lasnik (and disagree with Evans) that blocked coconstrual principles are broadly sufficient to

Ken Safir

predict the relevant coconstrual patterns permitted by linguistic competence; i.e. there are no linguistic rules requiring unique coconstrual of pronouns with their antecedents (although I shall set aside some candidate relations). In the cases where some form of coconstrual is blocked, I agree with Evans that noncoreference is not what is grammatically required. However, I agree with Lasnik that some cases of blocked dependency are not *only* failures of dependency. Rather, the result of (the descendants of) the noncoreference rules is *expected noncovaluation*, an expectation that, unlike the failure of dependent identity, can be overcome by a strong context. Thus coconstrual is always possible between two nominals, but it is not a linguistic matter to determine what coconstruals others intend. What linguistically competent individuals can be said to know is what varieties of covaluation and dependency are possible or unexpected if coconstrual is posited to hold.

1.2 DEPENDENT REFERENCE VERSUS COREFERENCE

The set of distinctions Evans makes between dependent reference and intended coreference may be his most profound contribution to the linguistic literature on anaphora. The distinctions he highlights had mostly been treated as between quantifier-bound anaphora and accidental coreference, but Evans's analysis of (13b)as involving a kind of bound anaphora that is parallel to that in (13a) had very important effects. There are two ways in which the bound anaphora effect for cases like (13b) is typically brought out in the literature.

One sort of illustrative example is based on the two interpretations of sentences like (14), one of which allows the inference in (14a) and the other of which allows the inference in (14b).

- 14. *Each female monkey* thinks that only *she* loves *her* mother.
- 14a. Each female monkey thinks the other monkeys don't love their respective mothers.
- 14b. If Mavis is a female monkey, then what Mavis believes is that Mavis's mother is not loved by any monkey other than Mavis.

One way of expressing this fact (there are some elaborations in the literature) is to say that (14a) is a reading where *her* depends directly on *only she*. The reading in (14b) is one where *her* is

directly dependent on *each female monkey*, rather than on *only she*, and *she* of *only she* is also directly dependent on *each female monkey*. However, it is clear that we do not want to say that either reading is one of coreference, since *each female monkey* is not referential (for discussion of such cases, see in particular, Heim (1993), Reinhart (1999), Safir (2004b: 22-3, 48)).⁴ In any event, what is at stake here is the pattern of dependency; what are *she* and *his* bound variables of?

Lasnik (1976) and later Reinhart (1983*a*: 150–6) note that bound readings can be preserved in ellipsis, for example.

15a. *Every monkey* loves *its* mother and every elephant does too.15b. *Marcus* loves *his* mother and Milton does too.

The elided portion of the second conjunct must meet a parallelism requirement with respect to the first conjunct, and in (15a) this requires, if *every monkey* and *its* are coconstrued in the first conjunct, that the interpretation of the second conjunct is that *every elephant* loves *its* mother. A similar reading, called a *sloppy reading*, is possible for (15b), whereby *Milton* loves *his* mother. What appears to be copied in the latter case, as Sag (1976) and Williams (1977) writing on ellipsis also point out, is that a parallel pattern of dependency can be preserved in (15b) just as it is in (15a).

Although Evans did not employ examples like (14) or (15) to make his point, examples like (14) show that patterns of dependency must be distinguished if we are to understand the patterns of coconstrual, and examples like (15) show that the patterns of dependency that hold of quantifiers may also hold of the relationship between a name and a pronoun coconstrued with it. Examples like (16) show that the relationships established for (14) may be recapitulated with proper name antecedents.

⁴ Reinhart (1999: 7) in particular avoids the use of the term 'coreference' in favor of 'covaluation', a view I will expand upon here, but she regards dependent coconstrual and covaluation as mutually exclusive.

i. Covaluation: α and β are covalued iff neither A-binds the other and they are assigned the same value.

My account differs in that I take covaluation to hold where dependent identity does, but I also take covaluation to hold in many cases where dependent identity does not. In this respect my treatment is more like that of Fiengo and May (1994), but for a critique of their theory see Safir (2004b: 14-18).

16. Angela thinks that only she loves her mother.

To extend Evans's point, it would be very odd to describe the reading where no one loves Angela's mother but Angela as one that involves accidental coreference between *Angela* and *her*, whereas the inference permitted for (14b) was necessarily handled by establishing a direct dependency between *every female monkey* and *her*.

2.0 HOW EVANS CUTS THE PIE

Evans (1980: 235-6) does not dispute the fact that some coconstruals arise outside sentence grammar by extralinguistic factors.

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 that is not specifically secured by any rule of the language—it is simply a manifestation of one speaker's general capacity to *make sense of* the acts (including the linguistic acts) of others. When the previous reference is within the same sentence as the pronoun (and subject to further conditions) the coreferential interpretation of the pronoun is secured, as one interpretation of the sentence, by a linguistic rule.

The linguistic rule in question is based on Fregean substitution. The bound relationship established by *every man loves his mother* arises for Evans by virtue of its relation to a sentence like *Marcus loves his mother*, where the pronoun is substituted for, hence the interpretation that Marcus loves his (Marcus's) mother.⁵ This is the rule that applies in sentence grammar wherever it can. The bound reading of *every man loves his mother* is built from

⁵ Although I often refer to "an interpretation of a sentence" or "coconstrual interpretations" in the course of my discussion, it is important to know that I do not mean an interpretation of a sentence in the way some philosophers and semanticists do, according to which all value assignments to pronouns have been made such that the proposition expressed by the sentence can be evaluated as true or false. The patterns of coconstrual I am speaking of concern interpretations of sentences that are permitted by formal grammar under certain assignments of value to nominal terms that would result in covaluations (whatever the value happens to be). The output of formal restrictions on coconstrual applying to a given sentence is not, then, a fully interpreted sentence, but a partially interpreted one. Although this point is not explicitly made in most linguistic discussions of anaphora, I believe it is widely assumed. I return to this issue at the end of 4.0.

134

a generalization of the same rule that applies to singular nominals like Marcus.6

However, the rule establishing dependent reference cannot always apply. Evans (1980: 358) introduces a rule that blocks dependent reference in certain circumstances, and though he puts it in positive language, the blocking rule plays a role similar to Lasnik's prohibition (which became Chomsky's Principle C).

17. A term can be referentially dependent on an NP iff it does not precede and c-command that NP.

I have restated Evans's principle, dropping out the precedence requirement (not crucial here) as the Independence Principle in Safir (2004a, b) (see also Higginbotham (1983)).

18. The Independence Principle: if X c-commands Y, then X cannot depend on Y.

The Independence Principle will block the readings that Lasnik is trying to exclude, since in (9), repeated below, the rule establishing dependent reference cannot apply-the potentially dependent term, she, cannot depend on Brianna's.

9. *[S₁ [NPShe] [believes [S₂ [NPTom] [loves [NP [NPBrianna's] [picture of [NPher]]]]]]

By contrast, coconstrual is possible between she and her in (19) because her is not blocked from depending on she.7

19. $[S_1 |_{NP}She]$ [believes $[S_2 |_{NP}Tom]$ [loves $[_{NP} |_{NP}Brianna's]$ [picture of [NP*her*]]]]]]

Evans also claims to resolve the transitivity of noncoreference problem (raised in Wasow (1979)) that Lasnik solves with his noncoreference rule. The issue arises where term A and term B are blocked from establishing coconstrual directly, but might plausibly establish coconstrual by using term C as a bridge. Thus failure to

⁷ Although *her* can depend on *she* in (19), notice that the Independence Principle would not permit she to depend on her because she c-commands her. This conclusion would appear to be harmless here, but I shall return to it once the consequences of blocked dependent readings are more clear.

⁶ Whether the dependency relation that arises between a singular term and a pronoun is or is not logically prior to that between a quantifier and a pronoun is not an issue that plays any role in my discussion. All that matters for my account is that the relations in question are dependent identity relations in both cases.

form coconstrual relations is not enough, Lasnik argues, to characterize what is blocked. Rather, a rule of noncoreference applying to A and B must insure that no coreference is achieved by transitivity with C when coconstrual between A and B fails to be established. Consider Lasnik's example, which I reproduce in (20a).

20a. *The woman *he* loved told *him* that *John* was a jerk.

20b. The woman *he* loved told us that *John* was a jerk.

20c. The woman *he* loved told *him* that we were all jerks.

The fact is that the pairwise coconstruals in (20b) and (20c) are both possible; so without positing any direct coconstrual relation between *him* and *John*, what prevents *him* to be coconstrued with *he* and *he* to be coconstrued with *John* in (20a)? Since Lasnik's noncoreference rule prohibits coreference where *him* c-commands *John* such that *him* and *John* bear different indices, it does not matter what else *John* is coreferent with as long as *John* cannot be coreferent with *him*. This blocks the (20a) interpretation.

Evans handles this case by arguing that the interpretation whereby *him* is dependent on *he* in (20a) is one where *he* has independent reference, but he simply does not consider it possible for *he* to depend on *John*. It is not at all obvious why this should be impossible, given (20b), unless Evans is assuming that the rule that establishes bound anaphora by substitution applies to every pronoun at once (which appears to be what he does assume). Alternatively, he might claim that if X depends on Y and Y depends on Z, then X depends on Z, violating the Independence Principle by transitivity of dependence for cases like (20a).

2.1 EVANS'S ARGUMENTS AGAINST A RULE OF NONCOREFERENCE

Now recall that Lasnik's theory predicts noncoreference between two terms if one term is a name and it is c-commanded by the other. Noncoreference, in the strongest sense that Lasnik endorses, says that whatever the reference one of the terms is determined to be, the other term does not overlap it in any way.⁸ Thus all of the

⁸ Although I reject both coreference and noncoreference as appropriate terms for characterizing the effect that Lasnik's principle is supposed to induce, one can argue on narrower grounds that disjoint reference is also not the appropriate

examples in (22) are predicted to result in noncoreference between the italicized terms.

- 22a. Oscar/he realizes that Oscar is incompetent.
- 22b. Oscar must love Oscar's mother.
- 22c. *Bill* thinks *Bill* is terrific.

Absent any context, our intuitions tell us that the c-commanding names or pronouns are not normally coconstrued with the names they c-command, but Evans shows that the right sort of context can render all of these examples acceptable.

- 23. Everyone has finally realized that *Oscar* is incompetent. *Even Oscar* has finally realized that *Oscar* is incompetent.
- 24. Look, fathead. If everyone loves *Oscar's* mother, then certainly *Oscar* must love *Oscar's* mother.
- 25. I know what John and *Bill* have in common. John thinks that *Bill* is terrific and *Bill* thinks that *Bill* is terrific.

The success of coconstrual in these cases indicates that noncoreference is not what the application of Lasnik's principle ensures. Evans considers the relative acceptability of the coconstruals in $(2_3)-(2_5)$ as evidence that dependent readings may be blocked, but not intended coreferent ones.

To summarize, where dependent identity coconstrual is blocked, intended coreference is still possible by nonlinguistic means. There is no rule of noncoreference, and indeed the noncoreference effect Lasnik speaks of is illusory.

3.0 WHAT'S WRONG WITH EVANS'S PICTURE

I have three major objections to Evans's attempt to derive all the patterns of coconstrual with only (a) a linguistic rule that enforces a dependency relation with the force of coconstrual and (b) a principle which blocks formation of dependency. First, no linguistic rule enforcing dependent readings is needed, because the possibility of dependent readings can be left as a default, once we understand what is at stake for the examples where Evans argues that a rule

relation. For arguments against disjoint reference, see Safir (2004a: 45–8) and references cited there.

is necessary. Secondly, not every failure of dependency produces the same sort of result. Failures of dependency that induce an *expectation of noncovaluation* must be distinguished from those that do not. Lastly, the distinction that Evans draws between dependency and scope is underappreciated. This latter distinction permits a cross-sentential treatment of the distribution of dependency relations that is more general, while it also extends the force of my second objection.

In 3.1 I begin by addressing the second objection, examining the cases normally characterized in the linguistic literature as falling under Principle B (Chomsky's noncoreference rule) and Principle C (Lasnik's noncoreference rule) of Chomsky's 1981 Binding Theory. As a matter of presentation, I assume in this section that my first objection is valid and that there is no rule of enforced coconstrual, although I will argue for this indirectly, since in the lines of argument I develop there is never any need for such a rule. In 3.2 I turn to the third objection, the difference between scope and dependency, which also provides partial justification for the first objection. I reserve one of the main arguments against an enforced coconstrual rule for section 4.

3.1 FAILURE OF DEPENDENCY AND THE EXPECTATION OF NONCOVALUATION

Although I think that Evans is correct to argue that coconstrual can succeed in (23)-(25), he sidesteps an important question about the cases that Lasnik's theory was meant to address. There is a reason that *He thinks that Bill is terrific* is taken by most speakers to involve two different individuals in the absence of a special context.⁹ In other words, the nature of the contexts that are required to permit coconstrual in (23)-(25) share a certain flavor that 'failure of dependency' does not capture on its own. Evans might respond that (23)-(25) are instances of intended coreference

⁹ Evans allows that names might be intended coreferent when one c-commands the other, but he is silent about the status of cases like *If everyone considers Oscar incompetent, then it can come as no surprise that even* he *considers* Oscar *incompetent.* These are less acceptable than the repeated name cases, but seem to permit of the same basic account.

which are simply not achieved by linguistic rule. However, if (23)-(25) can count as intended coreferent *by nonlinguistic means*, it seems that it is also necessary to say that there is a preference for establishing coconstrual by means of dependent readings (in Evans's theory, by means of his rule) *where possible*.¹⁰

But this begs the question of what exactly it is about these contexts that renders them successful in permitting intended coreferent readings. Notice that (23)–(25) permit the c-commander to be covalued with the c-commanded name just in case the c-commander is introduced as an *instantiation* of individuals who have a certain property, a property that presupposes that the identity of the individual involved is already familiar. Notice that the contexts provided in (23) and (24) already establish that Oscar is necessarily included in the set of people who love Oscar's mother or who think Oscar is incompetent. Now why should this make a difference for covaluation?

The use of *even* in (23) suggests that it is *unexpected* that the generalization in question should also hold of Oscar. *Even* is a word that adjusts expectations, such that amongst all the individuals who might be likely to have some property, *even* X picks one who is not very likely to have that property and asserts that X has that property nonetheless. This explains why the statement in (25) is ironic—it is an assertion that Bill shares a property that John does, but that Bill sharing that property is not what one would expect because it gives Bill's immodesty the air of an objective assessment, as if Bill were examining someone else and came to the conclusion that that person was terrific. The example loses this force if *Bill* is replaced by a pronoun as in (26), since the default interpretation in that case is one where *he* depends on c-commanding *Bill* for its identity value, rather than matching *Bill* in the first conjunct.¹¹

26. I know what John and *Bill* have in common. John thinks that *Bill* is terrific and *Bill* thinks that *he* is terrific.

¹⁰ This is a matter I will reintroduce as Preferred Covaluation in the next section, but interested readers should consult Safir 2004*a*: 25–34; 2004*b*: ch. 2 for references and discussion of some problems with previous formulations.

¹¹ A more technical discussion would discuss the role of Rule H, which is an additional condition on the pattern of dependencies introduced by Fox (1998). Rule H ensures dependence of X on the first c-commanding Y under coconstrual, unless Y is specified in some way that prevents X from depending on it. The role of Rule H is expanded and further justified in Safir 2004*b*.

A similar example, first noted by Higginbotham (1985: 570) and discussed by Fiengo and May (1994: 10) and Safir (2004*a*: 28), involves a discourse as follows.

- 27a. A: John is getting up to go.
- 27b. B: That's not John.
- 27c. A: Well, he's putting on John's coat.

The success of covaluation in (27c) arises because A assumes he knows who John is and that B is a fathead because he does not see that John is who they are looking at. The irony of A's statement in (27c) arises from the view that the individual putting on John's coat would have to be John, and that would go *against B's expectation* that John and the guy putting on his coat are not one and the same.

The point I am making here is that Evans has set out what it is for dependent reference to fail, but he has not set out why it is that failure of dependent reference results in an expectation of noncoreference. Put another way, Evans points out that intended coreference *can* succeed with a rich enough context, but he does not explain why a special context is required. So now let us provisionally restate the effect of the failure of dependence.

28. If X cannot depend on Y, then covaluation of X and Y is unexpected.

As Evans argued, (intended) coreference (or covaluation, the term I prefer and justify in section 5), is not what is blocked when dependency fails; rather, what is required when dependency fails; I believe, is that something must adjust our expectations in order for covaluation to be acceptable.

There are other cases where the expectation of noncovaluation is overcome. For example, in copular sentences like (29) the subject *Heracles* c-commands the rest of the sentence, including *Hercules*, and so *Heracles* and *Hercules* should not be allowed to corefer (a conclusion Lasnik recognized and embraced in his n. 6, p. 108, in the 1989 reprinted article).¹² Evans notes that Lasnik's commitment to the assertion that (29) is ungrammatical is a glaring example of what is wrong with his approach.

¹² With respect to copular constructions, Lasnik allows that *intended* coreference may be at play instead of just coreference, presumably because he is taking the position that *the morning star* is intended by the utterer to have a referent distinct from *the evening star* for the purposes of making the point that they are the same. This does not seem plausible to me either, if this is what is intended.

140

29. Heracles is Hercules.

Given (28), all that is claimed for these cases is that *Hercules* cannot be identity-dependent on *Heracles* for its antecedent, and therefore noncovaluation is expected. However, copular sentences address precisely the expectation of noncovaluation, since they assert that two terms not known to be covalued do indeed have the same value. Dependent reference is inappropriate in such cases, as can be seen for examples like (30).

30. Heracles is himself (today).

The use of a reflexive is very awkward as an identity statement, because its presuppositions clash—the reflexive established dependent identity whereas the copular asserts independent covaluation. Such sentences are not ungrammatical, but they are only fully acceptable with idiomatic readings, concerning, for example, Heracles' mental or physical disposition.

So one way of summing up what I have said so far in this subsection is that the effect Lasnik was speaking of is real, but Evans is correct to say that the effect in question is not one of grammatically enforced noncoreference. Rather, the effect of these cases where dependent identity readings are blocked is the creation of an expectation of noncovaluation.

I turn now to a more subtle point concerning contexts where dependent identity fails but no expectation of noncovaluation results. These should be the general case, given Evans's theory, but the contrast between these cases and the ones where expected noncovaluation does result serve to show that Evans's theory is insufficiently nuanced.

For example, there are some instances where Evans permits a relation of dependency to be formed, but there is evidence against the existence of the dependency in question. In some of these cases, where dependency cannot hold in spite of Evans's prediction, we fail to see the emergence of an expectation of noncovaluation.

Although there is some controversy about it, it appears that dependency is correctly licensed by Evans's account between *the president* and *his* in (31a), because verb phrase ellipsis permits a bound (sloppy) reading (see (15b)), one where the governor's wife is not expected to make sure of his (the governor's) infidelities. There is also a more favored *strict reading*, by which the governor's wife will not make light of the president's infidelities.

142

31. Although we were hopeful that *the president's* ambitious wife might make light of *his* infidelities, the **governor's** marriage is more traditional, so none of us expected **his** wife would.

However, where a pronoun embedded in the subject is coconstrued with a name that follows, as in the first conjunct of (32), only a strict reading is possible, one where the governor's wife makes light of the president's infidelities.

32. Although we were hopeful that *his* ambitious wife might make light of *the president's* infidelities, **the governor's** marriage is more traditional, so none of us expected **his** wife would.

There are two points to be made here. First, the success of the sloppy reading in (31) shows that dependent identity must succeed where *the president* does not c-command the pronoun in the first conjunct, or else the sloppy reading would not be parallel to it. In other words, *c-command is not crucial for dependent identity to hold* (consistent with Evans's theory, but not with some others—see below). Second, if dependent coconstrual fails in the elided phrase of (32), then dependent coconstrual did not hold for the first conjunct in (32). Yet the first conjunct in (32) does not induce a noncovaluation effect between *his* and *the president*—they can easily be coconstrued, even though there is no dependency relation between them and no instantiation context need be invoked. *Thus failure of dependency does not have to result in expected noncovaluation*.

The latter asymmetry in coconstrual between names embedded in subjects and pronouns that they do not c-command, on the one hand, and between pronouns embedded in subjects and names they do not c-command, on the other, is partially unexpected in Evans's story (depending on some ancillary assumptions), because dependent identity should be permitted in both cases. I have argued (in Safir 2004*b*) that this asymmetry is related to another one well known to linguists.

33a. Everyone's mother loves him (but no one's lawyer does).

33b. *His mother loves everyone (but his lawyer doesn't).

The bound reading of the pronoun is supported in (33a), and the acceptability of the sloppy reading under ellipsis confirms this success, but the quantifier-bound reading in (33b) is unacceptable

quite apart from the ellipsis in the conjunct that follows it. This is what is known to linguists as the *weak crossover effect* (WCO), which is also illustrated in (34).

- 34a. Who loves his mother?
- 34b. * Who does his mother love?

Wasow (1979) extends Postal's (1971) notion of "crossover" to this case in that the direct object in (34a) appears to the left of the subject at the front of the sentence. Linguists theorize that *who*—in (34b) has moved from the object position after *love*, where objects normally appear, across the subject to sentence-initial position, and this involves "crossing over" the subject (and the pronoun it contains). Linguists hypothesize further, or at least many have, that the unacceptability of (33b) has the same source, if some sort of phonetically unpronounced movement places *everyone's* in a scopal position to the left of the pronoun in (33b). The origin of the term "crossover" depends on that history, but the phenomenon does not. Many characterize WCO effects by one of the following two generalizations.

- 35a. A bound pronoun must be c-commanded by its antecedent.
- 35b. A bound pronoun cannot be embedded in a nominal that c-commands its antecedent.

For both of these definitions, assume "the antecedent" is the variable in the thematic position where the question word originates (before it moves to the front of the sentence) or where the universal quantifier is pronounced (before it moves silently to its scopal position). The first of these generalizations, which is supported by Reinhart (1983*a*, *b*) and Grodzinsky and Reinhart (1993), cannot explain the success of bound anaphora in cases like (31) and (33a) and others discussed in Safir (2004a: 34-7) and at greater length in Safir (2004b), since these are cases where the antecedent does not c-command (it is embedded in the subject). By contrast, the generalization in (35b) works for all of the cases we have looked at so far. Moreover, (35b) looks a lot like Evans's principle, which I have slightly revised and called the Independence Principle. If we assume the extended notion of dependence in (36), the resemblance of the WCO effect to the Independence Principle becomes a consequence of that principle, repeated below.

- 36. Extended Dependence: a nominal α depends on β if α embeds y and y depends on (β).
- 18. The Independence Principle: if X c-commands Y, then X cannot depend on Y.

Extended Dependence seems a very natural notion, semantically, perhaps an inescapable one, and it then extends the force of the Independence Principle in interesting ways, not the least of which being that it derives the existence of crossover effects.

Now let us return to simple cases like (37a-c). While dependence of *her* on *Laura* is permitted in (37a), *her mother* c-commands *Laura* in (37b), so *her* cannot depend on *Laura*. In fact (37b) is the case we showed cannot support sloppy readings under ellipsis (in (32)).

37a. Laura's mother loves her.

37b. Her mother loves Laura.

37c. **She* loves *Laura*.

However, *her* can be coconstrued with *Laura* fairly successfully (especially with a bit of stress on *loves*) (a point also made for the first conjunct of (32)). The success of coconstrual in (37b) contrasts sharply with (37c), which is very difficult to accept without an instantiation context. If both (37a) and (37b) are cases where the pronoun fails to depend on *Laura*, then the contrast between these cases establishes that failure of dependency is not always enough to trigger the expectation of noncovaluation.

This brings us to cases like (38), which Evans does not address.

38. *Laura loves her.

Since *her* in (38) does not c-command its subject antecedent, dependence of *her* on *Laura* must be ruled out by some other principle, such as Chomsky's (1976a) noncoreference rule that became Principle B of his (1981) Binding Theory, simplified in (39).

39. Binding Theory

Principle A: an anaphor must be c-commanded by its antecedent within Domain D.

Principle B: a pronoun must not be c-commanded by its antecedent within Domain D.

Principle C: a name must not be c-commanded by its antecedent.

144

Abandoning Coreference

Let us assume that the necessity of Principle A is not at issue here,¹³ and focus now on Principle C, which Evans's version of the Independence Principle is supposed to encompass, and Principle B, which Evans does not (attempt to) account for. It appears that the cases that Principles B and C are supposed to cover are the ones that induce an expectation of noncovaluation that a strong context can overcome. Moreover, Principles B and C rule out dependent reference in contexts where other dependent forms could occur, such as reflexives, in the case of Principle B, and pronouns, in the case of Principle C.

The complementarity of names versus pronouns and pronouns versus reflexives has long been noted in the linguistic literature; but for a variety of empirical and theory-internal reasons, theories that attempted to capitalize on complementarity as a guiding principle did not gain wide currency until recently. I have chronicled the history of such theories in Safir (2004a), where both the empirical claim about complementarity and a competitive theory of anaphora that derives it are defended at length.

The theory I propose relies on (41), which, though couched here as a principle, is actually developed in the reference cited as an algorithm which applies to syntactic representations to block dependent identity readings.¹⁴

41. Form-to-Interpretation Principle (FTIP)

If X c-commands Y, and Z is not the most dependent form available in position Y with respect to X, then y cannot be directly identity dependent on X.

¹³ Naturally, it would be attractive to derive Principle A from deeper principles. Reuland (2001) makes an interesting, very technical proposal to this effect. I do not endorse his particular proposal, however, because, unlike Reuland, I do not think that A-movement and Domain D coincide. See Safir 2004*a*: 148–56.

¹⁴ The competitive approach to anaphora embodied in the FTIP should not be understood in Gricean terms, as in the approaches of Reinhart (1983*b*) with respect to Principle C effects and Levinson (1987, 1991) with respect to Principle B effects. Reinhart relies on speaker intentions to refer, which I am arguing against as unnecessary and misleading, and Levinson does not distinguish dependent reference from coreference, and hence does not adequately distinguish pragmatic and syntacticosemantic effects. For discussion and argumentation against these proposals, see Safir 2004*a*: 61–6.

⁴⁰a. *He* believes that *he*/*Oscar is incompetent.40b. *He* loves *himself*/**him*.

42. Most Dependent Scale: syntactic anaphor>>pronoun>> name.

The FTIP would apply to (40b) to exclude *him* as identity-dependent on *he*, because *himself* is a more dependent form that could occur in that position (*himself* must satisfy Principle A and can do so in that position). The FTIP thus determines that a pronoun cannot support the dependent reading in that position, because there is a reflexive that could do so. In the case of (40a), a reflexive is not possible in the position of the dependent form, because Principle A would not permit it, but *he* is available and could appear in that position. Since *he* is a more dependent form than *Oscar*, *Oscar* cannot support the dependent reading in (40a).

There are several important points about the FTIP that are relevant to our discussion. First, the FTIP collapses Principles B and C into one principle, and since Evans did not address Principle B, which a more complete theory of the sort he proposed would have required, the FTIP is not "extra", and it is certainly an improvement over having separate principles, as in the Binding Theory.

Second, as is befitting a unifying principle, FTIP captures the fact that both Principle B and Principle C effects are instances where an expectation of noncovaluation is produced in addition to the failure of dependency. To capture this fact, I add a principle inducing obviation, one that refers to the output of FTIP, but not to syntactic structure at all.

43. Pragmatic Obviation: If FTIP determines that Y cannot be identity-dependent on X, then X and Y are not expected to share a value.

The difference in the effects produced by FTIP, on the one hand, and Pragmatic Obviation, on the other, is evident from the following ellipsis paradigm.

- 44a. *Almost everyone* thinks Oscar and Arthur are incompetent. Even Oscar thinks Oscar is incompetent, and maybe even Arthur does.
- 44b. You may not think we are looking at *John* and *Mary*, but if not, why did *he* put on *John's* coat, and then why did Mary?

No bound reading is possible for (44a) such that Arthur thinks Arthur is incompetent; rather, a strict reading is required, under

146

which Arthur thinks Oscar is incompetent. Similarly, *he* and *John* can be covalued in (44b), but (44b) permits only a strict reading, where Mary subsequently puts on the coat that John has put on, rather than a reading where Mary puts on Mary's coat.

The last point may not seem like such an advance. Isn't it the case that Evans's theory employing the Independence Principle already ensures noncovaluation for cases where dependent identity readings are blocked? The answer to this, quite simply, is "no", not since we extended the notion of dependency with (36). Examples like (37b) are instances where the Independence Principle blocks dependency, but there isn't an expectation of non-covaluation. We have already seen this for a version of (3c), restated here as (45).

45. The irrefutable evidence that *he* had lied to all of *his* previous wives convinced *Freddy's* current wife to file for divorce.

In other words, the interpretive effect of the Independence Principle is that dependent identity interpretations are blocked, not that the expectation of noncovaluation is induced. This is just as well, because if all blocked dependency were to have this effect, then our theory would face an internal contradiction for cases like (46).

46. He loves his mother.

In (46), *he* cannot depend on *his* because *he* c-commands *his*, yet we paid no attention to such cases because *his* can depend on *he*. This is not injurious to assumptions about dependent identity—indeed, it is the right result—but if noncovaluation is tied to failure of dependent identity, then the failure of dependency of *he* on *his* should mean that noncovaluation is expected, even as dependency of *his* on *he* is permitted. This is one reason why I limit the force of Pragmatic Obviation to the output of FTIP, not to the output of the Independence Principle.

The upshot of this section is that dependent identity readings can be blocked by either the Independence Principle or the FTIP, but the latter principle is not considered by Evans. The need for a distinction between these principles is established by a difference that Evans misses: namely, that only FTIP-blocked dependent readings result in the expectation of noncovaluation, induced in this system by Pragmatic Obviation.

3.2 CROSS-SENTENTIAL DEPENDENCY

The Independence Principle in itself says nothing about the possibility of cross-sentential dependencies. If such dependencies exist, then neither the FTIP nor the Independence Principle would ever apply to block them, since these principles can only apply where c-command holds and c-command is inherently a sentence-internal matter. If we do not assume that dependencies are legislated by rule, as Evans does, then dependent identity readings are predicted to be possible cross-sententially wherever they are not blocked by linguistic or pragmatic phenomena.

As long as dependent readings are assumed to be limited to cases where a pronoun has a quantificational antecedent, it is reasonable to assume that the distribution of dependent identity readings would be at least in part a function of quantifier scope (see, for example, Koopman and Sportiche (1981: 150) and Safir (1984: 626)).

47. The Scope Principle: If *X* is a variable bound by a quantifier *Q*, then *X* is in the scope of *Q*.

Most likely, the Scope Principle is either an axiom or a natural consequence of any serious semantic theory that includes quantifiers and variables and I shall simply assume it. With respect to the matters at hand, it has been assumed in most syntactic theories that scope is bound by sentence grammar. If so, it follows that dependent identity readings arising from quantifier binding are all limited to sentence grammar.

However, when Evans argued that examples like John loves his mother and Everyone loves his mother could have the same sort of dependent identity reading, he made it possible to consider whether dependent identity readings could be sustained over discourses as long as scope is not involved. Recall now that Evans assumed that cross-sentential interpretations should be established pragmatically, but this is consistent with what I say here, since I am assuming *all* dependent identity readings (apart from most syntactic anaphor readings) are coconstruals established pragmatically. In the case that a pronoun is coconstrued with a quantifier, the interpretive component will assign the only reading such a coconstrual will permit (i.e. a bound reading, not a coreferential one), but the assignment of coconstrual itself is not enforced by the grammar. After all, *his* in *Everyone loves his mother* does not have to be coconstrued with the quantifier if it refers to someone else familiar in the discourse.

Recall, however, Evans's first objection to Lasnik's theory: a pronoun should be treated as dependent on its sentence-internal nonquantified antecedent rather than representing a coincidence of independently referential terms. To ensure that this is the case, I introduce another simple structure-independent principle.¹⁵

48. Preferred Covaluation: covaluation arises from dependent identity unless dependency is blocked.

Thus the coconstrued interpretation of *Marcus loves his mother* is now required to arise in the same way as *Everyone loves his mother*, not by a rule, Fregean or otherwise, that enforces coconstrual, but by an interpretive condition that requires a coconstrual, regardless of how it arises, to have a particular semantic form (parallel to what I said about quantifier-bound readings). This is sufficient to eliminate the need for a rule that *enforces* coreference or coconstrual, insofar as coconstrual is always possible unless some version of it is blocked.

On the other hand, if the Independence Principle blocks a term A from depending on a term B, then A can still be covalued with B without any expectation of noncovaluation, unless the FTIP applies. If a pronoun is used without a linguistic antecedent (as in (4), for example), then there is obviously no relationship to which Preferred Covaluation could apply. If dependent identity is blocked by FTIP, then Pragmatic Obviation insures an expectation of noncovaluation. The way it is stated in (48), Preferred Covaluation is itself asyntactic, and one way of thinking of it is as a preference to rely on previous mention, if possible, to establish an identity for a term.

¹⁵ The view that dependent anaphora is preferred to express covaluation is proposed by Reinhart (1983*a*: 167), who puts it as follows: "When syntactically permitted, bound anaphora . . . is the most explicit way available in the language to express coreference So when coreference is desired, this should be the preferred way to express it." The force of her proposal is quite different from in my account. She assumes that dependent readings are possible only under c-command, a view rejected here. Also, the notion that bound readings are more explicit is mistaken: *Even Oscar thinks Oscar is incompetent* is more explicit for the strictly covalued reading than *Even Oscar thinks be is incompetent*, since the use of a pronoun allows a bound reading as well as a strictly covalued one. See Safir 2004*a*: 32–3. As pointed out in n. 4, Reinhart's (1999) usage of 'covaluation' is incompatible with bound anaphora, so she has to state this preference in a rather different way.

The interaction of Preferred Covaluation and the assumption that dependencies can be intersentential provides a natural account for the contrast between the paragraphs in (49a, b) (drawn from Safir 2004*b*: 53).

- 49a. His back was to us when we came in. He swivelled in his chair to face us. The penetrating eyes of Count Marzipan were trained upon us.
- 49b. Count Marzipan was brooding. His back was to us when we came in. He swivelled in his chair to face us, his penetrating eyes trained upon us.

Since no quantification is involved that requires sentence-internal dependency and none of the pronouns c-command *Count Marzipan*, the pronouns are permitted to be dependent on *Count Marzipan* in both (49a) and (49b). In Safir 2004*b*: 53 I put it this way:

Clearly (49a) and (49b) have a different status. Without any context, examples like (49a) are the stuff of mystery stories, where a pronoun is introduced that we have no referent for and we must wait for a plausible candidate to appear that supplies a value for the pronoun. The effect that disfavors backwards coreference (and hence backwards dependence) in these cases appears to be nothing more than Preferred Covaluation, which forces us to defer the assignment of a dependent reading until we have an appropriate antecedent. Where the antecedent is finally introduced in a position that permits a dependent reading, the tension created by Preferred Covaluation is resolved.

If we do not assume Preferred Covaluation, then we must have some other account of the preference for previous mention to establish an antecedent in simple sentences. The point I am making here is that once Preferred Covaluation is assumed, (49a, b) and the sentence-internal cases fall under the same generalization.

It was one thing to show that it is *possible* to think of crosssentential coconstruals as dependent identity readings that do not rely on the scope of a quantifier (as in (49a, b)), but it is another to show that cross-sentential coconstrual *must* ever be thought of this way. To support the stronger claim, we must revisit the readings permitted by the verb phrase (VP) ellipsis construction.

As mentioned earlier with respect to (31) and (32), the interpretation of elided VPs can produce an ambiguity between strict and sloppy readings, as illustrated in (15b) repeated below as (50a). 50a. Marcus loves his mother and Milton does too.

50b. ... and Milton loves his mother too.

50c. ... and Milton loves *his* mother too.

Earlier, we focused on the sloppy reading of (50a) represented in (50b); but there is, of course, a strict reading whereby Milton loves Marcus's mother, as illustrated in (50c). However, if Marcus and his are assumed to be coconstrued, then the reading for the second conjunct can only be strict or sloppy—it cannot be a third party reading. That is to say, the second conjunct cannot be interpreted to mean that Milton loves a mother who is neither his own nor Marcus's. It is typically said of such cases that the interpretation of the elided portion of the second conjunct must satisfy a parallelism requirement with respect to some closely similar antecedent VP, i.e. loves his mother in this case. Part of that requirement is that the pronoun that is part of the ellipsis must either be part of a copied dependency, or else get its value directly from the pronoun in the first conjunct. This is as close as we get to a linguistically required coconstrual rule, incidentally, and even in this case, there is more than one possible value that can be assigned.

Now that the basic phenomenon is clear, we can address the issue of how the strict reading arises. It is typically said in most accounts of ellipsis that the sloppy reading recapitulates the dependency relation in the antecedent VP, but that the strict reading copies a reading from the first conjunct that is "merely coreferent", not dependent, between Marcus and his (see e.g. Grodzinsky and Reinhart 1993). This seems wrong on two counts. The first is that Preferred Covaluation (a version of which Grodzinsky and Reinhart assume) tells us that if the first conjunct allows dependent identity to express coconstrual, then it is dependent reference that must be employed-there should be no ambiguity in the first conjunct. This requirement goes to the heart of Evans's critique of Lasnik's theory. A retreat from this position would be to say what is copied in the VP ellipsis construction is the reference or value for Marcus in the first conjunct. However, reference is the wrong notion, as section 5.0 will show, and Preferred Covaluation tells us that coconstrual must arise by dependency if possible.

In other words, the strict reading should be thought of as an identity-dependent reading connecting the pronoun in the antecedent directly to the elided but interpreted pronoun. This is possible on my account, for which dependency relations can extend

across sentences. Thus both strict and sloppy readings can arise from dependent identity relations without hypothesizing an ambiguous interpretation for the coconstrual in the antecedent sentence, or introducing rules of coconstrual that rely on any other relation but dependent identity.

4.0 THE IRRELEVANCE OF SPEAKER INTENTIONS AND ENFORCED COREFERENCE

Up to this point, I have not introduced any linguistic rule that refers to structure in order to restrict covaluation, unless the covaluation in question is one of dependent identity. Coconstrual, apart from the use of inherently dependent terms (as in the case of reflexives that respect Principle A) and the requirements of parallelism, is never required as a matter of linguistic form. If coconstrual is posited to hold between two nominals, however, the interpretation of the coconstrual relation is mandated by linguistic principles such as the FTIP, the Independence Principle, Pragmatic Obviation, and Preferred Covaluation. In the case of quantifier-bound pronouns, semantic necessity requires that the coconstrual take on a certain interpretive force (and then the Scope Principle must be satisfied). Where principles appeal to linguistic structures, as the FTIP, the Independence Principle, and the Scope Principle, properties of syntactic form must be respected and influence the interpretation of coconstruals accordingly. However, coconstrual is never fully blocked. In the cases where Pragmatic Obviation applies, coconstrual is possible if it overcomes expectations on noncovaluation, but whether or not covaluation holds is not determined by the grammar. In this respect I am arguing for a theory of coconstrual that Evans would have regarded as pragmatic.

Evans (1980: 236–7) believes that he has an argument against such a pragmatic theory based on the existence of *E-type* readings for pronouns. He presents the following two sentences, which are of a type originally pointed out by Geach (1962).

- 51. John owns some donkeys and feeds them at night.
- 52. Every villager owns some donkeys and feeds them at night.

Evans points out that *them* in (52) couldn't possibly refer to any particular set of donkeys (different owners own different donkeys),

but that it is not a bound variable either (i.e. it is not the case that there is some set of donkeys such that everyone owns them and feeds them at night). The relevant reading is the E-type reading, which, he is arguing, could not arise by any pragmatic assignment of a value for *them* that happened to correspond to *some donkeys*. Evans proposes a "Fregean" substitution rule as the linguistic rule that must be appealed to to achieve an E-type interpretation for (52) such that substituting any particular villager for *every villager* will yield sentences like (51). What is more important for our concerns is what he makes of this state of affairs (I have adjusted the example numbers to match mine):

the Fregean treatment presupposes that there is an interpretation of the pronoun in (52) on which its reference is determined by linguistic rule, and not by "considerations relating to situation, communicative intention, and the like" [Evans is quoting Chomsky 1976b]... 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.

The reference to "some salient group of donkeys" is, for our purposes, no different from an interpretation of *them* that has nothing to do with donkeys, supposing, in the case of (52), that only people with donkeys can get to the refugee camp to feed the refugees, where *them* = the refugees. The contention is that the speaker's intention to communicate the E-type reading must be known by his addressees if the addressee is to correctly interpret the pronoun (insuring that each x who satisfies a singular sentence like (51) feeds all the donkeys he owns).

While it is true that the addressee must have a linguistic rule or algorithm, whatever that linguistic device is, that ensures that the E-type reading will be restricted in the proper way, there is no need for the addressee to know what the speaker intends in order to know what an E-type reading of (52) would be if the speaker happened to intend that meaning. The point I am making is that anyone listening to (52) need only know that the E-type interpretation is possible, among others, as a matter of linguistic form, and that no linguistic criterion decides which of these interpretations the speaker intends. Thus Chomsky's (1976b) appeal to "considerations relating to situation, communicative intention, and the like" apply to choice the between possible readings here, but no linguistic rule need appeal to the speaker's intentions in any way. This is exactly parallel to the line I have taken for quantifier-bound interpretations—if a pronoun is coconstrued with a quantifier, then it must be interpreted as a bound variable and satisfy the Scope Principle, but no linguistic rule *requires* an otherwise independent pronoun to be a bound variable.¹⁶ In short, whatever account of the E-type reading one appeals to,¹⁷ it appears that this distinction between what linguistic form permits and what speakers intend can be maintained.

Perhaps one more example, this one involving Pragmatic Obviation, may be of use in establishing the disconnect between what linguistic form makes possible and what a speaker may intend. It is not hard to concoct a situation wherein speaker intention *could not* be relevant to the set of anaphoric interpretations a given sentence is permitted to have for someone unaware of the speaker's intentions. Imagine that I am at the beach, sitting behind a rock. Albert, whom I know nothing of, strolls along alone on his morning walk, but I don't see him. I suddenly hear him musing, "That man is hiding behind John's rock!" Now I can take this to mean that Albert sees someone behind a rock he takes to be the one that John owns or that somehow has something to do with John. Since Albert is alone, registering his observation, his remark is not addressed to anyone-that is to say, he is not intending to communicate a message to someone who does not know what he intends by what he says. Now suppose that in saying this Albert is identifying the person behind the rock as John, since that rock is the place where he was told that John was likely to be found-in other words, hiding behind John's rock is how one is to identify John

¹⁶ As mentioned with respect to Principle A, there are forms that must have antecedents and that may be inherently dependent, such as English *himself*. Linguists have explored these forms and their readings in some detail. See Safir 2004*a* for extensive discussion and references.

¹⁷ None of this diminishes the importance of Evans's insights about E-type pronouns, which opened a new vein of discovery that has been exploited for the last 25 years in the linguistics literature. See e.g., Elbourne (2001) for a recent discussion with references. It is not necessary for what I have to say to commit myself to a particular analysis of E-type pronouns. The non-scope-based account of dependency offered here permits us to say that E-type pronouns are dependent on an antecedent that does not have scope over them, whether that antecedent is taken to be a definite description of some sort or something else. (as in the coat example (27C)). I know none of this. I don't even know if I am the person that Albert sees behind the rock he associates with John. In short, I know nothing about Albert's intentions, but I do know on the basis of what he says and my linguistic competence that I should not expect *John* and *that man* to be covalued. If I knew all that Albert did, I would have a completely different expectation, but that is not a matter of linguistic form.¹⁸

From this perspective, if someone intends to use anaphora in a sentence to communicate a message, then at a minimum that would-be communicator must respect the restrictions on what a listener can linguistically take a sentence to mean. Short of using linguistically marked bound anaphora (e.g. reflexives), it is incumbent on the communicator to provide enough context, or bring enough of the context to bear, to help the listener light on the meaning the communicator intends. Linguistically, the intended coconstrual is not assured without that context.¹⁹

One final clarification is in order, however, if it has not been made clear already. When I speak of *coconstrual interpretations* that are or are not possible, I am not presuming that formal restrictions on coconstrual interpretations are providing a complete

¹⁸ This is, in effect, just an exemplification of Chomsky's (1971: 19) remark (concerning views of Russell): "Though consideration of intended effects avoids some problems, it seems to me that no matter how fully elaborated, it will at best provide an analysis of successful communication, but not of meaning or of the use of language, which need not involve communication or even the attempt to communicate."

¹⁹ Compare the approach in the text with the language of intended coreference as employed by Reinhart (1983*a*: 167, 1983*b*: 76), who proposes (i) and (ii).

- i. *Speaker's strategy*: When a syntactic structure you are using allows boundanaphora interpretation, then use it if you intend your expressions to corefer, unless you have some reasons to avoid bound-anaphora.
- ii. *Hearer's strategy*: If the speaker avoids the bound anaphora options provided by the structure he is using, then, unless he has reasons to avoid bound-anaphora, he didn't intend his expressions to corefer.

The hearer does not need to know what the intentions of the speaker are to know how a given coconstrual is permitted to be interpreted. I have reinterpreted the "unless he has reasons to avoid bound-anaphora" locution as a predisposition not to expect anaphora where the FTIP blocks a dependent interpretation. Reinhart's view that all bound readings that are blocked result in some sort of obviation is one I reject. Grodzinsky and Reinhart (1993) and Reinhart (1999) present more nuanced versions of this idea, but the reliance on the view that bound readings are always licensed by c-command, which I also reject here, underlies these proposals as well.

interpretation of the sentence that could be used to determine, in a given context, whether the utterance containing a possible coconstrual is true or not. A complete interpretation of a sentence, including extralinguistic factors, such as assignment of values to all of the pronouns, will have to respect the restrictions that formal grammar imposes, but formal grammar is not responsible for assigning these values. This does not change the logic of my critique of Evans's account of E-type interpretations, however. Formal grammar does not require speaker intentions to compute whether such a reading is possible or not. Assignments of value to the pronouns will be necessary to determine which of the possible interpretations permitted by formal grammar corresponds to the one the speaker meant in a given context, but one does not need to know what a speaker means to apply the part of "the rule" that gives a particular interpretation.

5.0 THE EXPANDED WORLD OF IDENTITY READINGS

It may strike some as odd that I have avoided expressing the coconstruals I speak of as instances of *coreference*, sticking mostly to descriptions of coconstrual that I characterize as covaluation or dependent identity. The reason for this apparent circumlocution is that most coconstruals do not necessarily involve coreference, and some of the cases of bound readings or obviation I have discussed cannot be said to involve coreference or noncoreference at all.

First of all, I set aside some cases that might come directly to mind: namely, those where the antecedent is itself a mythical entity. Thus *Pegasus turned his head* may not be said to be coreference if *Pegasus* does not refer; but I am not interested in such cases, which do not involve any mismatch between the value for *Pegasus* and the value assigned to *his* on the coconstrued reading. The cases I have in mind all involve a kind of mismatch between terms that still holds under coconstrual.

For example, the readings in (53) all involve what I call $proxy^{2\circ}$ readings, whereby the dependent is some sort of representative token of the identity of its antecedent. A sample of the range of

 $^{^{\}scriptscriptstyle 2o}$ I do not use the term "proxy" to mean "pronoun of laziness", as Evans occasionally does.

proxy relations includes likeness in (53a, b), where *he* refers and *she* refers to the wax statues of Fidel and Marlene (53a) or people wearing the costumes representing Fidel and Marlene in (53b); author/work (53c), where *he* refers to Grisham's writings, and player/vehicle (53d, e) (all discussed in more detail and with references in Safir 2004*a*: 112–14):

- 53a. As they strolled through the wax museum, *Fidel* could not help thinking that *he* would have looked better in a uniform and **Marlene** could not help thinking that **she** would have looked better without one.
- 53b. The masquerade ball was a bit disconcerting. It seemed to *Marlene* that everywhere *she* looked, either *her* nose was too long and *her* chin too weak.
- 53c. Grisham claims that he is even more suspenseful in Swahili.
- 53d. After *her* last shot, *Alice's* ball was close to the gate, or it was until the Red Queen knocked *her* into the bushes.
- 53e. *Patton* realized that *he* would be vulnerable to a flanking movement.

Most speakers permit (53e) to be true even if Patton is directing his army (*he*) from a thousand miles away. As Jackendoff (1992), who was the first to note the wax museum cases, points out, proxy cases show that coconstrual is not necessarily about coextension or any normal sense of coreference.

Now we can test to see if proxy readings are a variety of coconstrual that shares the same properties as those that Lasnik's Noncoreference Principle (Chomsky's Principle C) applies to.

54. *He thinks Yeats reads better in English than Swahili.

Whether or not (54) describes Yeats's own opinion of his reading abilities, or it describes a proxy situation, such that Yeats is expressing an opinion about how his work is experienced in English and Swahili, we have an expectation that *he* and *Yeats* are not covalued—the FTIP prefers a pronoun, blocks a dependent interpretation, and then Pragmatic Obviation requires an expectation on noncovaluation. To say that *he* and *Yeats* are not coreferent on the proxy interpretation would tell us nothing, since on that interpretation they are not coreferent to begin with. Indeed, according to Lasnik's Noncoreference Principle, coconstrual between *he* and *Yeats* should not be odd at all. Moreover, it appears that the relevant covaluation can be ameliorated in the same way that Evans argued that Lasnik's cases could.

55a. Everyone thinks Yeats reads better in Swahili. 55b. Even *Yeats* thinks *Yeats* reads better in Swahili.

In this instance, the subordinated *Yeats* in (55b) seems to depend on the previous mention of *Yeats* in discourse, thus permitting the proxy interpretation. However, it does not seem likely that any notion of coreference worth the name could be rehabilitated to account for these usages.

One might try to argue that proxy interpretations are merely elliptical expressions or stand-ins such that the proxy Yeats stands for Yeats's poetry and the proxy his stands for his/poetry, in which case Chomsky's Principle C would distinguish the cases in the usual way. However, if the syntax is simply sensitive to the expanded interpretation of the stand-in, why are animate singular pronouns used? Yeats says it reads better in Swahili does not give us an anaphoric reading where it refers to Yeats's poetry, nor would Yeats says they read better in Swahili referring to his poems. Using real ellipsis we can show that the use of inanimate pronouns or pronouns that don't match in number cannot support bound readings.

56a. Yeats says he reads better in Spanish and so does Lord Dunsany.56b. Yeats says it reads better in Spanish and so does Lord Dunsany.56c. Yeats says they read better in Spanish and so does Lord Dunsany.

A sloppy reading is possible for (56a), such that Lord Dunsany says Lord Dunsany's poems read better in Spanish, but sloppy readings are not possible for (56b, c). The view that proxy pronouns are substitutions for descriptions containing dependent pronouns would not lead us to choose animate number-matching pronouns over inanimate non-number-matching ones. The view that these proxies are dependent identities offers a more tractable account: matching identities require matching pronouns.

Contrary to the position I take here, Jackendoff (1992; 1997: 55, 73–4) proposes that rules of grammar must have access to whether or not a proxy reading is intended, an argument not unlike Evans's E-type pronoun argument, and one that can be answered in the same way. If proxy readings are necessarily dependent readings, such that the proxy depends on its animate antecedent, as argued in

Safir 2004*a*: 112–14, then they should succeed wherever dependent identity readings are not blocked (by FTIP or the Independence Principle), but where those principles reject dependency, as in (54), the reading fails. This accounts for the asymmetries of interpretation in (57a, b) ((56a) from Jackendoff (1997)).

- 57a. Ringo fell on himself.
- 57b. In Michaelangelo's time, *David* indicated how handsome *he* was thought to be.

Imagine that Ringo is in the wax museum, stumbles, and falls down on the statue of him that was just about to be hoisted into place. As Jackendoff points out, this reading is strained, but it contrasts sharply with a reading that is not at all possible; (57a) cannot describe a situation whereby someone bumps into the statue of Ringo such that it falls on Ringo the person. If the inanimate proxy must depend on an animate antecedent, then the reading where the statue falls on Ringo is precluded by the Independence Principle, since the proxy would c-command its antecedent. A similar set of judgments hold of (57b), which could be a statement made by one David who is enamored of himself, but it cannot be a statement about how the statue of David indicated how handsome the biblical David was. Cases like (57a) and (57b) do not violate FTIP, since the reflexive and the pronoun are the most dependent forms available, respectively.

Thus it appears that there is no need to appeal to some level of conceptual relations between proxies and their antecedents or to speaker intentions to know if one term is permitted to be identitydependent on another. A dependent interpretation is available if it is not blocked linguistically, and the patterns of proxy interpretation respect these limits. Moreover, in cases where proxy interpretation is blocked because dependency is blocked, the interpretive result is not noncoreference or even expected noncoreference, but rather one of expected noncovaluation.

Another sort of coconstrual relation that the term 'coreference' seems inadequate to characterize is what I call the *guise* relation, illustrated by the contrast in (58) (a version of a class of examples first discussed by Lakoff (1968)).

58a. A: If I were you *I* would hate *myself*. B: I do. 58b. A: If I were you, *I* would hate *me*. B: I do.

Examples like (58a) support a bound interpretation, such that B's response means that B is a self-hater. While not every speaker is comfortable with the acceptability of A's remark in (58b), it is clear that B's response can only mean that B hates A, not that B is a selfhater. Thus, even though both I and me refer to the speaker, I is most easily taken to refer to A in the shoes (having the perspective) of speaker B, while me is presumably A in his own shoes. The reason that A's statement in (58b) is less than fully acceptable is that it violates FTIP (myself is a more dependent form that could appear in the same syntactic position), and the failure of dependency does not permit a bound reading for the ellipsis in (58b). However, Pragmatic Obviation, which indicates that I and me should be expected not to be covalued, can be overcome by the fact that the first person requires that every first-person mention by the same speaker must refer to that speaker. In other words, Pragmatic Obviation between I and me is partially neutralized by the indexicality of person.²¹

The key point about these guise examples, however, is that a dependent identity reading, the self-hater reading, cannot be unblocked by the indexicality of the first person, and yet there is an interpretation of (58b) for which coconstrual can still be said to

²¹ If we use third-person examples and violate FTIP, it is much harder to recover any relevant meaning.

- i. If Bob were Bill, he would believe that he was smart.
- ii. If Bob were Bill, he would believe that *Bob was smart.
- iii. If Bob were Bill, *he* would hate *him*.

If Bob lacks the sort of confidence that Bill has in him, then (i) could be used to say that Bob, if he could see himself through Bill's eyes, would see better how smart he, Bob, is. Alternatively, (i) could have a bound reading: suppose that Bill is very self-confident; then if Bob were just like Bill, Bob would be self-confident about his intelligence too. However (ii) disfavors both of these readings, since the dependent reading is blocked, and indexicality does not help to overcome the expectation of noncovaluation for the unbound reading (for most speakers). To my ear, similar blocked coconstruals hold for (iii), though of course the bound coconstrual is permitted with *himself* in place of *him* (speakers differ as to whether there is a strict coconstrual for (iii) with *himself*—see the discussion of c-antecedency in Safir 2004a: 107-11. This suggests that the consciousness of Bob in Bill is enough to block the relevant covaluation when FTIP induces Pragmatic Obviation—a rather subtle distinction, to be sure, but something that the notion of noncoreference does not begin to address.

succeed. The coconstrual that succeeds is a coreferent one in some sense, since the speaker is coextensive with the speaker, but one faces tortured questions about what coextension means when the consciousness of one individual is housed in the perspective of another. These distinctions are not confined to reflexives.

59. If I were any one of you, I would ask the president to give me a pardon.

The bound reading of *me* in (59) is one that might be appropriate in a situation where the president's aides are all guilty of a crime and the speaker (the president's chief of staff) is advising them that they will each need a pardon. The unbound reading in (59) might be salient if the speaker were making a veiled threat; if the president is not prevailed upon to give the speaker a pardon, then the speaker might have a lot to say about the guilt of everyone else in the room. The bound readings in (58a) and (59) require that the identity of the bound form is completely dependent on the interpretation of its antecedent, while the readings that are not bound do not seem to rely crucially on the identity of their antecedent at all.²² It appears that the identical guise reading requires a dependent identity relation, and if a dependent identity reading is not blocked, then the identical guise reading is possible.²³

There are other interesting cases where coconstrual is not complete even under bound interpretation, but these two are enough to show asymmetries between two terms that must be regarded as covalued.²⁴

 22 Evans (1980: 240) assumes that first- and second-person pronouns are like proper names in that they cannot be referentially dependent. Examples like (59) show that Evans was wrong about first-person (and presumably second-person) pronouns. By contrast, Schlenker (2000) assumes that all first-person pronouns are bound by null operators. I argue against this proposal in Safir (2004*c*).

²³ This is not what Lakoff (1996: 98) makes of this. He proposes a conceptual distinction between the "subject", which represents one's subjectivity, and the "self", which roughly corresponds to one's body/place, and claims that "the use of anaphoric pronouns is based on this split." However, this can't be about the shape of the anaphoric morpheme *myself*, since the same sort of reading is possible for first person pronouns in (59). Moreover, it is only metaphorically about body/place, since most people share the intuition that (59) could be telepathed between aliens lacking corporeal form. What the right metaphor for guises is, is interesting, but whatever it turns out to be, it appears that complete identity of guises is possible only where dependent readings are, and this possibility is all that the syntax has to contribute.

 24 Evans (1977: 98–9) mentions pronouns bound by the subjects of attitude verbs as inadequately treated by the term "coreference". As Lewis (1979) has pointed out, there are *de se* readings and non-*de se* readings, and the former show some asymmetries

Thus it turns out that there are coconstrual relations between terms that must be at least covalued, even if there is no dependent identity relation between them (as in the unbound reading for first-person guises), and that there are coconstrual relations which must be dependent identity relations, but where extensional coreference is not possible (proxies bound by animate antecedents).

A variety of weighty semantic questions open up about what the identities covalued by linguistic form consist in such that we can say that two identities are the same by one notion of coconstrual and not by another. Indeed, for the unbound guise cases, it is not even clear what it means to say that two distinct guises are somehow covalued. My contribution to this question is largely empirical (developed in more detail in Safir 2004a), but I take these questions to be both open and very interesting.

6.0 CONCLUSION: ABANDONING COREFERENCE

In order to evaluate linguistically what a sentence is *permitted* to mean (not what it actually means), we do not have to know what a speaker intends to say. Grammar permits us to determine a range of meanings that a given coconstrual can have and compute which meanings it cannot have—the rest is not a matter for the grammar

between antecedent and dependent. For example, (i) describes a moment in that unfortunate king's life when he did not think his wife had met his father.

i. Oedipus thought his mother had never met his father.

However, the description *his mother* is not a use of the pronoun *his* that reports what Oedipus knew about himself, while the use of *his* in *his father* could refer either to the man Oedipus thought was his father (the *de se* reading) or the man that actually was his father unbeknownst to him (the non-*de se* reading). The *de se* reading for both instances of *his* is pragmatically unlikely in the ancient world for biological parents. The *de se*/non-*de se* distinction is not one between readings that are bound and those that are not, as Chierchia (1989) and Huang and Liu (2001) have pointed out. For example, the same ambiguities arise for the bound reading in (ii).

ii. *Anyone in Oedipus's position* at that point in the story would presumably think that *his* mother had never met *his* father.

The reason I raise the matter in this context is that the distinction is linguistically relevant, in that there are languages that actually mark *de se* readings morphologically, such that the form of the pronoun used could indicate that the *de se* reading is required. For discussion of morphologically induced *de se*, see Safir (2004c) and references cited there.

at all. In saying this, I am certainly not advocating that it is of no consequence for *anybody* to examine notions of what people intend to accomplish by uttering what they do—doubtless a complete picture of communicative situations requires such a project. I am explicitly arguing that the full interpretation of a sentence is something greater than the result of formal grammar. In other words, I am insisting, as Lasnik and Chomsky do, on a line between formal grammar and the uses to which the products of formal grammar are put.

From this perspective there is no place for the words "intended coreference" or "intended reference" in linguistic theory. Once Evans had facilitated the distinction between dependent identity and covaluation, and once the varieties of identity that can be coconstrued in these ways are distinguished, not even the simple term "coreference" is viable any longer. Instead we are left with various coconstrual relations between nominals supporting identities. One such relation between nominals, dependent identity, can be blocked by linguistic form, while another, covaluation, can be rendered unexpected; but neither of these linguistically significant coconstrual relations crucially requires extensional equivalence for the referents that those nominals may be employed to pick out. It is swimming upstream to legislate against a term so commonplace as "coreference", so it will likely go on being used in an informal way to describe coconstrual-but no thoughtful use of the term "coreference" is any longer appropriate in discussions of linguistic form.

Evans and the Sense of "I"

JOSÉ LUIS BERMÚDEZ

This paper focuses on two enduring features of Gareth Evans's work. The first is his rethinking of standard ways of understanding the Fregean notion of sense, and the second his sustained attempt to undercut the familiar opposition between Russellian and Fregean approaches to understanding thought and language. Evans's The Varieties of Reference provides the only worked-out development of how a Fregean should understand singular thoughts. Singular thoughts are ways of thinking about objects that require the existence of the object being thought about. Paradigm singular thoughts are perceptual beliefs of the sort that might be communicated through demonstrative expressions such as "this" or "that" and beliefs about oneself or the present moment that might be expressed through token-reflexive expressions such as "I" or "now". Suppose we characterize Russellian approaches to understanding language and the thoughts it expresses in very broad terms as holding that all that is required for a correct and adequate specification of a thought (the meaning of a sentence or the content of a propositional attitude) is a specification of the object(s) being thought about and the properties attributed to it/them. It looks very much as if singular thoughts will lend themselves to a Russellian approach, and there have been a range of influential arguments to this effect (Kaplan 1989; Perry 1977). In The Varieties of Reference Evans offers a counterweight to this line of argument. Starting out with general reflections on Frege's notion of sense, Evans argues in Part I, on the basis of a distinctive way of

My thanks to Mark Sainsbury, Chris Gauker, Fiona Macpherson, Jane Heal, and Stephen Butterfill for helpful comments on earlier drafts of this paper.

understanding Frege's notion of a mode of presentation, that Fregean senses are in general object-dependent in precisely the way that singular thoughts are object-dependent. He then goes on in Part II (and in the related paper "Understanding demonstratives" (1981b)) to work out in considerable detail how this understanding of sense can be applied to thoughts that would most naturally be expressed using demonstrative and token-reflexive expressions. The discussion goes to the heart of many of the most central issues in the philosophy of language and of thought.

An important set of issues comes into focus when we ask whether indexical expressions have a Fregean sense.¹ John Perry's two influential papers "Frege on demonstratives" (Perry 1977) and "The problem of the essential indexical" (Perry 1979) argued powerfully that indexicals cannot be accommodated within a Fregean theory of sense, because there are requirements upon the Fregean notion of sense that the context dependence of indexical expressions precludes them from satisfying. Instead, we need to separate out different strands of what we might intuitively think of as the "meaning" of sentences expressed using demonstratives. In particular, he suggests in "Frege on demonstratives" that we need to distinguish the sense of a sentence employing indexical expressions from the thought expressed by that sentence. Evans responded to Perry in 'Understanding demonstratives'. In that paper, and in Part II of The Varieties of Reference, he offers a positive account of how we might understand the senses of sentences involving a range of indexical expressions. His account develops the idea that the Fregean senses of referring expressions are object-dependent and involve ways of identifying and keeping track of the object or objects being thought about (taking "object" in a loose enough sense for the present moment to count as an object).

My primary concern here is with just one part of the general debate opened up by Evans's work: namely, the status of the indexical expression "I". I shall be exploring the peculiar difficulties that "I" poses for a Fregean theory and show how Evans's account of the sense of the first-person pronoun can be modified to meet those difficulties. I begin, in section I, with some general comments on

¹ For the purposes of this paper I shall talk about demonstrative and tokenreflexive expressions as distinguishable categories within the general category of indexical expressions. Both demonstrative and token-reflexive expressions are context-dependent, but whereas demonstrative expressions "latch onto" objects through some sort of act of ostension (such as pointing), using token-reflexive expressions to pick out an object requires no such gesture.

José Luis Bermúdez

the Fregean approach to thought and language, identifying three different explanatory tasks that the notion of a thought is called upon to perform. Section II explores the distinctive difficulties that indexical expressions in general, and "I" in particular, pose for Fregean approaches to thought and language. We will consider in section III Perry's proposal to deal with these difficulties by fractionating the notion of a thought so that there is no single thing that performs the three functions identified in section I. Rejecting this proposal takes us, in section IV, to Evans's own account of the sense of 'I'. Evans's account can be seen as a working-out of Frege's well-known dictum that "everyone is presented to himself in a special and primitive way in which he is presented to no one else" (Frege 1918: 25). Evans's account has several dimensions, one of which stresses the importance in self-conscious thought of judgments based upon forms of information derived from sources that can only provide information about oneself. Evans is quite explicit that his account of the sense of "I" has the consequence that "I"-thoughts are not shareable. We see in section V that, despite Evans's best efforts, this cannot be reconciled with the objectivity of "I"-thoughts. It is, I will argue, a requirement upon an account of the sense of "I" that it respect the symmetry constraint, so that what I say in a given context using "I" should be the same as what you might say in that same context using "you". In section VI I propose a development of Evans's account that meets the symmetry constraint.

I FREGE'S CONCEPTION OF THOUGHT: A SKETCH

For Frege thoughts are simultaneously

- (1) the bearers of truth and falsity (the things² that are true or false);
- (2) the objects of propositional attitudes (the things that are believed, hoped, etc.);

² The description of thoughts as things should not be taken too literally. It need not be taken to suggest more than, first, that we can refer to thoughts (by talking, for example, of the thought that p); second, that we have some grasp of the identity conditions of thoughts (of what makes thoughts the same or different); and third that thoughts are not in any sense reducible either to individual ideas or acts of thinking or to the sentences that express them. It is true that Frege suggested in various places that thoughts exist in a "third realm", but we can set these comments aside. They are not essential to the overall picture.

Evans and the Sense of "I" 167

(3) the senses of assertoric sentences (the things expressed by sentences that purport to describe the way things are).

The interconnection of these three different dimensions of the notion of a thought sets the background for the problems posed by indexical expressions.³

Frege's distinctive idea is that the essence of a thought is given by the fact that it is a candidate for truth or falsity. Fundamental to the analysis of a thought is its truth condition: namely, what would have to be the case in the world for that thought to be true. Starting from the truth condition, we can analyze the parts of a thought in terms of how they contribute to determining that truth condition.

For Frege, judgment is the fundamental propositional attitude. To form a judgment is to commit oneself in thought to a certain state of affairs holding in the world. The judgment is true just if that state of affairs does hold, false otherwise. Judgment itself is a psychological state—a relation to the thought, rather than a part of the thought. The thought itself is the postulate (for want of a better word) that the relevant state of affairs holds in the world. To judge the thought to be true is to endorse the postulate—to take it to be true. And what we endorse when we form a judgment is precisely what we understand when we understand an assertoric sentence (where an assertoric sentence is one that characterizes the world as being a certain way).

Frege's notion of sense is correlative with the notion of understanding, so that the sense of a linguistic expression, whether that expression is a proper name, a logical constant, a predicate, or a complete sentence, is what a competent language-user understands when he understands it. The connection with the other two aspects of the notion of thought comes because Frege takes the understanding of a sentence to be basic, and the notion of truth to be fundamental to understanding. We understand a sentence insofar as we understand its truth condition, and we understand the parts of a sentence insofar as we understand how they contribute to determining the truth-value of a sentence.⁴

³ It should be clear that my account of Frege owes much to Michael Dummett, and in particular to Dummett 1973.

⁴ There are various complicating factors here, not least of which is that something needs to be said about fictional statements that are not directly truth-apt. Evans himself has very interesting things to say about this in ch. 10 of *The Varieties* of *Reference*. His views are illuminatingly discussed in Sainsbury 1999.

José Luis Bermúdez

Our understanding of the constituent words of a sentence is parasitic on our understanding of the sentences in which they can feature. We understand a word insofar as we understand the contribution that word makes to determining the truth condition of such sentences. For a proper name, that understanding consists in knowing what it would be to single out an object as its semantic value. For a predicate, it consists in a means of determining whether or not an object falls within the predicate's extension. We understand logical operators (or at least the truth-functional logical operators) insofar as we understand how they contribute to determining the truth conditions of complex sentences on the basis of the truth-values of their constituent sentences.

The three aspects of Frege's conception of a thought dovetail together. For Frege, thoughts are the bearers of truth and falsity, the objects of the propositional attitudes, and the meanings of assertoric sentences. Indexical expressions are problematic, because they seem to prise these aspects apart. Section III illustrates how this fractionation of the notion of a thought threatens not only to reduce the Fregean notion of sense to incoherence, but also to render problematic the very notion of psychological explanation. First, though, we will turn in the next section to considering how Evans proposed to put flesh on the bones of this general schematic account.

II EVANS'S INTERPRETATION OF FREGE

The notion of sense is ineliminably bound up with the notion of truth. The sense of a sentence is its truth condition, which is a function of the senses of its subsentential components. This is often put by saying that sense determines reference. But how should "determination" be understood here?

According to popular interpretation, the relation of determination (as applied to referring expressions) is the relation of satisfaction. The sense of a referring expression is a definite description or cluster of definite descriptions. The referring expression refers to a given object because that object is the unique satisfier of the associated description or cluster of descriptions. Proper names, therefore, are synonymous with definite descriptions, and in order to understand a proper name, one needs to understand the sense of the synonymous definite description. This way of understanding how sense determines reference has been subjected to a battery of criticism, most of which has to do with what are taken to be its counter-intuitive modal implications. Kripke has been the prime instigator of this line of attack (Kripke 1980). According to Kripke, referring expressions such as proper names are rigid designators, which is to say that they refer to the same object in all possible worlds (in which that object exists), whereas definite descriptions are non-rigid designators that can shift reference across possible worlds.⁵

The repudiation of the descriptive view is central to Evans's development of the notion of sense. Evans's conception of sense actually requires treating proper names and non-descriptive referring expressions as rigid designators.⁶ Evans takes his lead from what he calls Russell's principle, according to which a subject cannot make a judgment about something unless he knows which object his judgment is about. Evans interprets "knowing which" as requiring that the thinker possess *discriminating knowledge* of the object being thought about, where that in turn requires being able to distinguish that object from all other things (1982: 89). We have, he maintains, an intuitive grasp of three different forms that such discriminating knowledge of an object by perceiving it. Or by being able to recognize it.⁷ Or by knowing distinguishing facts about it.

⁵ This is not the place to discuss Kripke's arguments. It is worth pointing out, though, that there is a far simpler argument from the behavior of proper names in complex sentences to show that proper names cannot be synonymous with definite descriptions—an argument that does not appeal to problematic intuitions about modality. Proper names cannot be synonymous with definite descriptions because they do not display scope ambiguities with respect to negation, whereas definite descriptions do (when they are construed in Russell's manner). There are two ways of understanding the negation of the sentence "The teacher of Alexander was the greatest living Stagyrite", but only one way of understanding the negation of the sentence "Aristotle was the greatest living Stagyrite". But, of course, the issue is not really whether proper names are synonymous with definite descriptions, but rather whether they function as non-rigid designators.

⁶ See the introduction to this volume at pp. 7–8 for Evans's understanding for referring expressions and his reasons for thinking that definite descriptions are not referring expressions. Note that Evans allows a category of so-called descriptive names that function as rigid designators although their reference is fixed descriptively. See Evans 1982: §2.3

⁷ It is not simply perceptual recognition that is at stake here. Evans is operating with a broad sense of recognition that extends, e.g., to recognizing the sort of information that would be relevant to determining the truth-value of sentences about the object.

Each way of possessing discriminating knowledge corresponds to a distinct type of linguistic referring expression.

The discriminating knowledge that derives from standing in a perceptual relation to an object will most obviously be expressed linguistically by means of demonstrative expressions. The understanding of most proper names will deploy recognitional discriminating knowledge, while knowledge of discriminating facts about an object will most naturally be expressed linguistically through definite descriptions. It is Frege's basic idea that we should think about the sense of referring expressions in terms of knowing what it would be to identify objects as their semantic values. We can now see how Evans develops this basic idea. Different types of referring expression will be correlated with different ways of identifying objects: demonstratives with perceptual identification, (most) proper names with recognitional identification, and definite descriptions (and some proper names) with what I characterized earlier as identification through satisfaction.

Evans devotes relatively little attention to identification through satisfaction (apart from the category of descriptive names), and we will follow him in this. Most of *The Varieties of Reference* is devoted to recognitional and perceptual identification, both of which rest upon the thinker's standing in certain information links to the object being thought about. These information links yield the thinker's discriminating knowledge (or what Evans calls the subject's Idea, where an Idea of an object is the conception of the object that allows the thinker to differentiate that object from all other objects). In the case of demonstrative referring expressions these information links are obviously perceptual, but Evans envisages a far wider range of information links in the case of recognition-based thoughts, including information links based upon memory and testimony. In each case the sense of the relevant expression is understood in terms of the ability to exploit the appropriate information links.

Descriptive identification, in terms of which many authors have proposed that the sense of referring expressions can be completely understood, plays at best a peripheral role for Evans, applicable almost exclusively to those expressions that have the surface structure of definite descriptions. It is this basic distinction between descriptive identification, on the one hand, and recognition-based and demonstrative identification, on the other, that allows him to argue that a great many singular referring expressions are Russellian. The following passage is instructive. Evans and the Sense of "I" 171

In the case of description-based identification, a subject may have a fully *coherent* Idea-of-an-object, *a*, even if there is nothing that would be satisfied by that Idea.... But this is not so in the case of either of the other modes of identification. In those cases, if there is no object that would be identified by the purported mode of identification, then there is no coherent Idea-of-an-object, even in this sense. In these cases it is therefore true, in the strictest sense, that where there is no object, *there is no thought*. (1982: 136)

It remains unclear, however, how Evans's general approach to the sense of referring expressions is to apply to the first-person pronoun, which does not seem obviously to involve either recognition-based identification or demonstrative identification. I sketch out the basic contours of Evans's account of the sense of the first-person pronoun in section IV. First, though, we need to consider why indexical expressions in general, and the first-person pronoun in particular, pose prima facie difficulties for a Fregean notion of sense. That will be the task of section III.

III INDEXICALS AND FREGEAN THOUGHTS

The object picked out by an indexical expression is a function of the context in which the expression is uttered. In the case of demonstratives such as "this" or "that", the object referred to is most frequently determined by the ostensive gesture accompanying the utterance (or by whatever else makes a particular object peculiarly salient in that context). Simple rules determine a particular person as the referent of token-reflexive expressions such as "I" and "you"—roughly, "I" refers to the utterer of the sentence and "you" to the person (or persons) addressed. These simple rules are the basis from which we can understand more complex cases involving, for example, deferred reference (as when you write a letter on my behalf using the first-person pronoun).

Given that indexical expressions have varying reference, any account of the sense of those expressions confronts a stark choice. Does a given indexical expression have a single sense holding across different contexts of utterance, or does it have a different sense for each different context of utterance? For the first-person pronoun, is there a single sense of "I", or is each person's use of "I" to refer to himself associated with a different sense of "I"? It is clear that there are pressures in each direction. On the one hand, it seems plausible to demand that what I understand when I hear you saying "I am hungry" is the same as what you understand when you hear me say "I am hungry". There is a sense in which we are saying the same thing. But on the other, this is incompatible with the principle that the sense–reference relation cannot be one–many. If sense determines reference and reference is context-dependent, then there cannot be a single sense of "I".

In the face of these conflicting pressures John Perry has suggested fractionating Frege's unitary notion of sense into two distinct notions: the notion of sense, on the one hand, and the notion of a thought, on the other (Perry 1977).8 The sense of a sentence containing an indexical expression is the context-insensitive component of what we might intuitively think of as the "meaning" of that sentence. So, the sense of a simple sentence of the form "I am F" is composed of the incomplete sense of the predicate "F" conjoined with an element corresponding to the linguistic rule that a token of "I" refers to the utterer of that token. In a simple sentence of the form "It is F now", the incomplete sense of "F" is completed by an element corresponding to the linguistic rule that a token of "now" refers to the time of utterance. And so on for the other indexical expressions. The *thought* expressed by the utterance of an indexical expression, on the other hand, is somewhat closer to a singular thought in the sense discussed earlier. It is composed of the actual object picked out by the relevant indexical expression, together with the sense of the relevant predicate expression.

Let us return to the three dimensions of Frege's notion of a thought identified in section I. Thoughts are:

- (1) the bearers of truth and falsity (the things that are true or false);
- (2) the objects of propositional attitudes (the things that are believed, hoped, etc.);

⁸ Perry uses somewhat different terminology in "The essential indexical", where "sense" is replaced by "belief state" and "thought" by "belief". Here is how he characterizes the distinction there: "The shoppers, for example, are all in a certain belief state, a state which, given normal desires and other belief states they can be expected to be in, will lead each of them to examine his cart. But, although they are all in the same belief state (not the same *total* belief state, of course), they do not all have the same belief (believe the same thing, have the relation of belief to the same object)....Belief states individuated in this way enter into our common sense theory about human behavior and more sophisticated theories emerging from it. We expect all good-hearted people in that state which leads them to say 'I am making a mess' to examine their grocery carts, no matter what belief they have in virtue of being in that state" (Perry 1979: 181). (3) the senses of assertoric sentences (the things expressed by sentences that purport to describe the way things are).

In the case of indexicals, Perry suggests, the third dimension comes apart from the second. The thoughts that are the objects of the propositional attitudes expressible by sentences involving demonstratives are context-sensitive and object-dependent, whereas the senses of those sentences are context-insensitive and do not include the objects referred to by the indexical expressions they contain. It is senses that carry the burden of psychological explanation, while it is thoughts that are picked out in propositional-attitude ascriptions.

Perry thinks that this fractionation of Frege's notion of sense is independently motivated. He has two arguments, each starting from the distinctive explanatory role of the states of mind expressible in sentences involving indexicals. The first line of argument comes across clearly in the following passage.

We use senses to individuate psychological states, in explaining and predicting action. It is the sense entertained, not the thought apprehended, that is tied to human action. When you and I entertain the sense of "a bear is about to attack me", we behave similarly. We both roll up in a ball and try to be as still as possible. Different thoughts apprehended, same sense entertained, same behavior. When you and I both apprehend the thought that I am about to be attacked by a bear, we behave differently. I roll up in a ball, you run to get help. Same thought apprehended, different sense entertained, different behavior. Nevertheless, it is surely the thought apprehended that is the indirect reference of a sentence containing a demonstrative reference in the scope of "believes". (Perry 1977: 494–5)

For sentences involving indexical expressions, Perry claims, content and functional role, come apart. The very person picked out by the first-person pronoun is part of the content of the belief expressed by a sentence involving "I"—and the beliefs expressed by two different persons each uttering equiform tokens of a given "I"-sentence are correspondingly different. But two people will respond in the same way to the sentence "A bear is about to attack me", even though that sentence gives rise in each of them to a different belief. So, we need a further notion to capture what it is that leads these two people to behave in similar ways.

The second argument also stresses the demands of psychological explanation. Perry starts from the assumption (discussed in section I) that Frege is committed to the thesis that all referring expressions

are synonymous with definite descriptions. This leads him naturally to the thought that, if a given indexical expression has a Fregean sense, it must be possible to substitute some coreferring definite description *not involving any indexical expressions* for that expression in any sentence in which it might feature.⁹ Perry shows with considerable plausibility that any candidate definite description will lack the immediate implications for action possessed by the putatively synonymous indexical expressions cannot have diverging implications for action in a single context, it follows that indexical expressions cannot be synonymous with definite descriptions.

Clearly, this second line of argument will have a limited impact on those Fregean theorists, such as Evans, who deny that all referring senses have to be synonymous with definite descriptions. But Perry's points do make very clear an important constraint upon any non-descriptive notion of sense applicable to indexicals: namely, that it should capture the *cognitive dynamics* of such expressions. Leaving aside the claims about definite descriptions, let us turn to Perry's suggestion that we need to distinguish between the sense of a sentence involving an indexical and the thought expressed by that sentence in order to accommodate the requirements of psychological explanation.

It is counter-intuitive to sacrifice the basic tenet of propositionalattitude psychology that people act the way they do because of what they believe. According to Perry, what people believe (the content of their belief) is not directly relevant to explaining how they behave. My belief that I am about to be attacked by a bear is a belief about me (and correspondingly different from your belief that you are about to be attacked by a bear), and this fact can only be captured at the level of the thought expressed by the indexical sentence "I am about to be attacked by a bear". But the fact that the sentence "I am about to be attacked by a bear" is about me, as opposed to you, is not on Perry's view a part of the explanation of why a certain type of behavior is correlated with uttering that sentence.

Suppose we leave aside the analysis of sentences involving indexicals and instead ask about indexical beliefs *simpliciter*. The

⁹ The qualification rules out, e.g, taking "I" to be synonymous with "the utterer of *this* token sentence".

Evans and the Sense of "I" 175

content of an indexical belief, according to Perry, is an objectdependent proposition. But what work does this object-dependent proposition do within the cognitive economy, given that the things that explain behavior are not object-dependent and are more general? The object-dependent proposition is what is either true or false. The bearers of truth and falsity are also what feature in inferences. This is most obvious when the inferences are truthfunctional. If the validity of an inference is a function of the truth-values of its constituents, then those constituents must be truth-apt. But the constituents of truth-functional inferences must also be capable of featuring in non-truth-functional inferences. With this in mind Perry's position becomes even more puzzling. The sorts of things to which we appeal in giving psychological explanations of behavior are not the sorts of things over which inferences are defined. How, then, can we make sense of the idea of a person reasoning their way towards acting in a certain way. It looks very much as if the conclusion of a practical inference must be an object-dependent proposition. Let us suppose that this practical inference issues in action. It is natural to think, then, that an adequate explanation of the relevant action will correctly identify the conclusion of the practical inference. And yet this is precisely what Perry is committed to denying.¹⁰

¹⁰ It is important to clear up a potential confusion at this point. Perry draws a close parallel between his notion of sense and Kaplan's notion of character (Kaplan 1979, 1989). Kaplan, like Perry, fractionates the notion of thought into two components: character and content. The content of an utterance, roughly speaking, is the proposition it expresses in a given context. This proposition can be understood as a function that assigns a truth-value to each index, where an index specifies a world, a time, an agent, a location in that world, etc. We can understand the contents of subsentential linguistic expressions accordingly. The character of a linguistic expression, in contrast, is the component that corresponds to linguistic meaning. An expression's character is a function that assigns a given content to each context. So, for example, the character of "I" is the function that assigns to each context the constant function from possible worlds to the agent of the context. In other words, the character of "I" is effectively that the denotation of any token of "I" is always the utterer of that token. One of Kaplan's principal contributions to the study of indexicals is his development of a logic of demonstratives in which the notion of character plays an integral part. However, Kaplan's logic of demonstratives is perfectly compatible with my claims about the split between inferences and psychological explanation. Although Kaplan's character (Perry's sense) does feature in his logic of demonstratives, it does so in a very subordinate role. Kaplan holds that the character of a linguistic expression is given by linguistic rules such as those deployed in the recursive definitions that serve to define truth and denotation in a context. There are no inferences at the level of character, although we need character to identify the

Given the counter-intuitive position in which Perry ends up, it is worth looking again at his argument that object-dependent propositions are too specific for psychological explanation. The argument rests on two implicit principles. The first is that all tokens of a given behavior-type require the same explanation. The second is that giving the same explanation for two different behaviors that are tokens of the same type involves attributing the same psychological states to the authors of the two behaviors. It is easy to see how those two principles rule out giving explanations of different tokens of a given behavior-type in terms of object-dependent propositions—provided, of course, that we really do have different tokens of a different behavior-type in the cases that Perry discusses.

Perry is assuming that if our bodies move in comparable ways, then we are behaving in the same way, so that there is just one behavior that is instantiated both when you roll yourself up in a ball and when I roll myself up in a ball. It is true that there is a way of thinking about behavior in which this is indeed the case. This is the sense of "behavior" in which a behavior is just a series of bodily movements. But one might think that psychological explanations explain actions rather than bodily movements, where actions are at least in part individuated by their goal. Looking at the aims of psychological explanation in this way, it is no longer so obvious that my rolling up in a ball is the same action as your rolling up in a ball. After all, my rolling myself up in a ball is intended to save me from the bear, while your rolling yourself up in a ball is intended to save you from the bear. One might be inclined to say that in this case we have two different actions that are effected by similar sets of bodily movements.¹¹ If one looks at the matter like this, then Perry's two principles do not yield the same result. What I am doing is saving myself from the bear. This action-type can have other tokens, such as, for example, my shooting the bear. These tokens would, in accordance with Perry's two principles, have the same explanation: namely, my belief that I am about to be attacked by a bear. Your rolling yourself up into a ball is a token of a completely different action-type: namely, the action-type of your saving

objects over which inferences are defined and evaluated. Truth and validity arise only with reference to specific contexts, whereas the whole point of the notion of context is that it abstracts away from the specifics of context. (I am grateful to Chris Gauker for encouraging me to address this issue).

¹¹ A similar point is made by Evans (1982: 203-4).

yourself from the bear, and, like the other tokens of that type, it is to be explained in terms of your belief that you are about to be attacked by a bear.

Of course, our actions are related. But this does not mean that there is a single thing (the sense of the sentence "I am about to be attacked by a bear") that explains both our behaviors. We can do justice to the similarity by noting that there is a significant similarity between the explanations that might be offered of my action and yours respectively. Consider the proposition expressed by the sentence "ILB is about to be attacked by a bear" (leaving open for the moment whether that proposition is to be understood in Perry's way or in some other way). We see that proposition as made up of an element corresponding to the proper name "JLB" and a propositional function corresponding to the incomplete expression "----is about to be attacked by a bear". Both my thought that I am about to be attacked by a bear and your thought that you are about to be attacked by a bear involve that very same propositional function being completed by constituents that in some sense correspond to the first-person pronoun we might use to express our beliefs. Surely the fact that both thoughts can be interpreted in terms of a single propositional function captures the respect in which they are similar.

This brings us back to a new formulation of the problem with which this section began. Suppose we accept, *pace* Perry, that there must (in the case of indexicals in general and "I" in particular) be a single thing that does the explanatory work that Perry parcels out between thought and sense. Is this to be understood in terms of Frege's notion of sense? The following section considers Evans's attempt to answer the challenge.

IV EVANS ON THE SENSE OF "I"

Evans discusses "I" in two different places. In "Understanding demonstratives" he offers an account of the sense of "I" in the context of a general discussion of demonstratives (Evans 1981b). In *The Varieties of Reference* Evans devotes an entire chapter to self-identification. There he investigates what he calls each person's "I"-idea—the discriminating knowledge of oneself that allows one to think "I"-thoughts. I shall assume in this paper that the two

discussions are continuous, so that what Evans says about "I"-ideas can be taken to represent his views about the sense of the first-person pronoun. An "I"-thought, according to Evans, is a thought that might typically be expressed through a sentence containing the first-person pronoun. Since Evans adopts an orthodox Fregeanism, according to which the sense of a sentence can be equated with the thought it expresses (as opposed, e.g., to positions such as that of Perry), it seems to follow that the sense of the first-person pronoun should be equated with the first-person component of the "I"-thought.

Frege famously wrote that "everyone is presented to himself in a special and primitive way in which he is presented to no one else" (Frege 1918: 25). One strand in Evans's account of the sense of "I" interprets this idea in a way that harmonizes with the general thesis (discussed in section II) that the sense of a referring expression is to be understood in terms of the utterer's discriminating knowledge of the referent of that expression. Everyone who uses the first-person pronoun with understanding does so in virtue of ways they have of thinking about themselves that are both primitive and not available to anyone else.

Using the first-person pronoun to refer to oneself requires being in touch with a particular object (namely, oneself), but not in a way that involves picking oneself out as the referent of the pronoun. Rather, it involves being in receipt of certain types of information information that is distinctive precisely in virtue of being the sort of information that does *not* require identification of a particular object as the source of that information. The types of information underwriting what Evans somewhat confusingly calls selfidentification are such that there can be no question but that one is oneself the source of the relevant information. Evans characterizes these types of information by using Shoemaker's notion of *immunity to error through misidentification*. This notion is defined in terms of the impossibility of a certain type of error:

to say that a statement "*a* is ϕ " is subject to error through misidentification relative to the term '*a*' means that the following is possible: the speaker knows some particular thing to be ϕ , but makes the mistake of asserting "*a* is ϕ " because, and only because, he mistakenly thinks that the thing he knows to be ϕ is what '*a*' refers to. (Shoemaker 1968: 557)

Judgments are immune to error through misidentification when this type of error is impossible. Immunity to error is an epistemological

notion relativized to the grounds upon which a judgment or statement is made, so that, e.g., the same sentence can be employed to make two different statements, only one of which is immune to error, if the grounds on which the statement is made on the two occasions of utterance are appropriately different. Not all judgments that would be expressed by sentences involving the firstperson pronoun are immune to error through misidentification (relative to the first-person pronoun). The relevant judgments are just those that are grounded in and based upon the appropriate sources of information—sources of information that can provide information only about oneself.

Autobiographical memory is one such source. When one remembers something "from the inside", there can be no question but that the person who had the remembered experience is oneself (leaving aside the possibility of what Shoemaker has called quasimemories—see Shoemaker 1970 and Evans 1982: §7.5). Introspection is a further source. If I come to believe that I am having a certain thought, I might well be mistaken, but I cannot come to believe through introspection that someone is having a certain thought and be mistaken about who that person is. Ordinary visual perception is a source of self-specifying information about one's own spatial location and trajectory. One's field of view is structured in a way that offers information about one's spatial relations to perceived objects and about the route one is taking through the perceived environment (Bermúdez 2002, 2003). This information can be inaccurate, but cannot be information about anyone else's spatial location or trajectory. Somatic proprioception also provides selfspecifying information that is immune to error through misidentification. If, for example, I form the judgment that my arms are folded on the basis of feedback from joint position sense, then there is no sense (as the world currently is) in which I can be mistaken about whose arms it is that are folded.¹²

We can now see how Evans proposes to develop Frege's suggestion that "everyone is presented to himself in a special and primitive way in which he is presented to no one else". Understanding Frege's "way of being presented to oneself" as involving ways of thinking of oneself (in line with his general way of understanding

¹² I am prescinding here from debates posed by the possibility of being hooked up to someone else's body. For further discussion see Ayers 1991 and Cassam 1995.

what Frege says about modes of presentation), Evans elucidates what is "special and primitive" in terms of the capacity to think thoughts, and to make statements, that are immune to error through misidentification relative to the first-person pronoun. The sense of the first-person pronoun is the ability to keep in touch with the person that one in fact is, the person referred to by one's use of the first-person pronoun. This ability is a function of one's possessing sources of information about oneself that are immune to error through misidentification.

There are two further dimensions to Evans's account of the first person. One has to do with the "output" side of first-person judgments. Being appropriately in touch with the person one in fact is is a function of one's ability to act on judgments derived from the appropriate information sources. Evans's account speaks directly to the issues about psychological explanation and cognitive dynamics raised by Perry. If I see a bear coming towards me, I acquire information that is suitably immune to error through misidentification to the effect that there is a bear coming towards me. So I act in the appropriate manner, immediately and without needing to identify myself as the person towards whom the bear is coming. There is no room for the thought that might be expressed as "The bear is coming towards someone, but is that me?" and no delay in moving straight into action. It might turn out that the bear has designs on the person behind me, rather than on me, but this would not be an error of misidentification in the required sense. It would be a mistake about the ramifications of the fact that the bear is coming towards me.

With respect to both the input and the output sides of the sense of the first-person pronoun, Evans can be seen both as fleshing out some of Frege's sketchy comments on "I" and as developing an account of the functional role of "I". The final strand in his account of "I"-ideas departs both from Frege and from considerations of functional role, as standardly understood.

Indispensable though these familiar ingredients (an information component and an action component) are in any account of the Ideas we have of ourselves...they cannot constitute an exhaustive account of our "I"-ideas. So long as we focus on judgements which a person might make about himself on the basis of the relevant ways of gaining knowledge, the inadequacy may not strike us. A subject's knowledge of what it is for the thought "I am in pain" to be true may appear to be exhausted by his capacity to decide, simply upon the basis of how he feels, whether or not it is true—and similarly in the case of all the other ways of gaining knowledge about ourselves. However, our view of ourselves is not Idealistic: we are perfectly capable of grasping propositions about ourselves which we are quite incapable of deciding, or even offering grounds for. I can grasp the thought that I was breast-fed, for example, or that I was unhappy on my first birthday, or that I tossed and turned in my sleep last night, or that I shall be dragged unconscious through the streets of Chicago, or that I shall die. (Evans 1982: 208–9)

We can, Evans emphasizes, effectively think about ourselves from a third-person point of view, entertaining possibilities to which we have no informational links (as in the vivid examples he gives). The truth conditions of these thoughts are far more complex than the truth conditions of thoughts about oneself that are based upon introspection, proprioception, and other sources of information that is immune to error through misidentification.

Consider a sentence of the form "I am F", where "F" is a predicate whose applicability to oneself can be determined through one of the informational links we have been discussing-hence where the sentence as a whole is immune to error through misidentification relative to the first-person pronoun. The truth condition for "I am F" is intimately tied to the subject's ability to use the relevant information link to determine whether the predicate applies. Because "I am F" is immune to error through misidentification, the subject does not have to identify himself in order to grasp the truth condition of the thought that the sentence expresses—the thought will be true just if the subject can detect *F*-relevant information in the favored way. The issue, from the subject's point of view, is simply whether or not there is F-ness. But when it comes to "I am F" sentences that are not immune to error through misidentification, the subject's grasp of the truth condition must involve, not simply sensitivity to the presence or absence of F-ness, but an ability to determine what it would be for it to be he who is F. And this in turn means that the subject must be able to think of himself in a much richer way than is required in order to grasp the truth condition of sentences that are immune to error through misidentification.

Evans returns at this point to Russell's principle and to the idea that thinking about an object involves having discriminating knowledge of that object. For Evans this requirement holds no less when the object in question is oneself than when it is an object in

the distal environment. So, the final strand in Evans's account of "I"-ideas (and hence of the sense of the first-person pronoun) is that the thinker possess discriminating knowledge of himself, where this knowledge is understood in terms of the subject's ability to locate himself within the objective spatio-temporal world.¹³ Here is how Evans puts it.

It seems to me clear that as we conceive of persons, they are distinguished from one another by fundamental grounds of difference of the same kind as those which distinguish other physical things, and that a fundamental identification of a person involves a consideration of him as the person occupying such-and-such a spatio-temporal location. Consequently, to know what it is for $[\delta = I]$ to be true, for arbitrary δ , is to know what is involved in locating oneself in a spatio-temporal map of the world. (Evans 1982: 211)

Being able to locate herself within a spatio-temporal map of the world allows a thinker to determine the truth condition of any thought (whether that thought involves predicates that are susceptible to error through misidentification, or immune to such error). With this third dimension in his account of "I"-ideas, Evans moves away from Frege and closer to a broadly Kantian conception of self-conscious thought, on which the ability to think about oneself is intimately tied to the ability to think about the objective order of things.¹⁴

V PRIVACY VERSUS OBJECTIVITY

Evans explicitly endorses Frege's suggestion that the sense of "I" is private. Against Perry, who maintains that the idea of a private sense is incompatible with Frege's conception of the objectivity of thought, Evans argues forcefully that there is no incompatibility between privacy and objectivity.

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.

¹³ See Evans 1982: §7.1 and the appendix to ch. 7.

¹⁴ This aspect of Evans's view has been taken up in recent books by Campbell (1994) and Cassam (1997). See also Bermúdez 1998.

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. (PW [Frege 1969] 251)

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. (Evans 1982: 313)

Evans proposes two respects in which thoughts involving the sense of "I" (as he is proposing to understand it) should count as objective. First, they are not reducible to the contents of an individual consciousness. Second, they exist and have a truth-value independently of anyone entertaining them.

Why might thoughts of a given class not be reducible to the contents of an individual consciousness, even though they are not shareable? There would be such irreducibility if there were thoughts of that class that have never been thought by anyone. This certainly makes perfectly good sense for, e.g., mathematical thoughts (on all non-constructivist views of mathematics). It cannot, however, be applied to non-shareable "I"-thoughts (or indeed to any nonshareable token-reflexive thoughts), simply because the identity of any given token-reflexive thought is a function of the episode of thinking. Recall that the reference of "I" in a sentence involving the first-person pronoun is fixed by the token-reflexive rule that (in standard cases) picks out the speaker of the relevant token sentence as the person to whom the pronoun refers. Since an "I"-thought is a thought expressed by a sentence involving the first-person pronoun, one would expect the reference of the first-person component of the thought to be fixed in an analogously token-reflexive way. The whole point of token-reflexivity, and indexicality in general, is that the identity of the thought is determined by the context in which it is thought-in which case, there is no thought without an episode of thinking. If this episode has to be an episode featuring the subject of the thought (as it must be if the thought is non-shareable), then I see no sense in which the thought can exist and have a truth-value independently of being thought by the particular person who is the subject of the thought.

Senses are not just the objects of propositional attitudes and the bearers of truth and falsity. They are also the units of communication what is understood when someone utters a sentence with understanding and when someone else hears that sentence with understanding. A thought that is non-shareable is a thought that is non-communicable. Frege himself seems to be in two minds about how communication is secured in sentences involving the first-person pronoun. On the one hand, the famous discussion in "The thought" that we have already considered appears to entail the non-shareability of the sense of "I". On the other hand, however, in the posthumously published "Logic", Frege suggests that the thought that I entertain is the same as the thought that I communicate in a sentence involving the firstperson pronoun.

It is not necessary that the person who feels cold should himself give utterance to the thought that he feels cold. Another person can do this by using a name to designate the one who feels cold. (Frege 1914: 1969: 235)

The emphasis here is not only on the communicability of "I"-thoughts, but even on the possibility of expressing them with sentences not featuring the first-person pronoun.

There is an ambiguity in the passage from "Logic". What is the name that someone else might use to designate me when I feel cold? Is it a proper name or a pronoun? This is particularly important in the present context. If the objectivity of "I"-thoughts requires shareability, then it must be possible for someone else to think a thought that I would express using "I". So we can ask: how would that person express this thought? What words would they use to express the same thought that I would express using "I"? We are looking for a token utterance that would be synonymous with my token utterance. Clearly, this would not be an utterance employing the first-person pronoun. No two utterances involving "I" can be synonymous if they are uttered by different people. As we saw in section III, there are good reasons for thinking that utterances involving "I" will generally not be synonymous with utterances involving proper names and definite descriptions, even when those names/descriptions pick out the same person as the use of "I". Of course, there will be some cases in which proper names and/or definite descriptions will work just fine. And there are no doubt cases where someone might use the third-person pronoun successfully to express the same thought that I would express using "I". But there

seems no prospect of finding a rule to the effect that, whenever there is a thought expressed by a sentence involving "I", that thought will equally be expressible by a sentence involving a proper name, or a definite description, or the third-person pronoun. The most plausible candidate for such a rule would be a token sentence involving the second-person pronoun, accompanied by a suitable way of fixing reference (either anaphorically or ostensively). As Mark Sainsbury has emphasized, a token of "I" can be synonymous with a token of "you", even though the type-expression "I" is not synomymous with the type-expression "you" (Sainsbury 1998). The thought that I express with a token of "I" in a given context just is the thought that you express with a given token of "you" in that same context (or a suitably related context), so that when you understand my utterance "I am F" you do so by grasping a thought that you would yourself express by saying "You are F". Let us say, then, that the objectivity and shareability of "I"-thoughts imposes a constraint. This is the constraint of preserving the token-equivalence of "I" and "you" within a given context. I will call this the symmetry constraint.

Evans's account of the sense of the first-person pronoun does not meet the constraint. Since he holds that the sense of the first-person pronoun is private, my utterance of "I" cannot be synonymous with vour utterance of "vou". If my understanding of the firstperson pronoun is tied to my sensitivity to, for example, proprioceptive information from my own body, then there is no sense in which my utterance of "I" can be synonymous with your appropriately related utterance of "you" (for the understanding of which proprioceptive information from my body is not in any sense an issue). The point holds more generally. What Evans calls the information component of an "I"-idea is a set of information links that can hold only between a person and himself. A fortiori, therefore, thoughts containing such an "I"-idea are not shareable-and so (on our assumed equivalence of each person's "I"-idea and his grasp of the sense of the first-person pronoun) we have the consequence that the sense of the first-person pronoun is not shareable.

Evans's argument that objectivity does not require shareability is an unsuccessful attempt to show that this result is not as unpalatable as it might immediately appear. But, even if it had been successful, objectivity is not the only issue here—and an objective thought that cannot be communicated or denied hardly fits the

Fregean picture of thought and linguistic understanding. Evans is well aware of this, and is explicit about his rejection of the Fregean model of communication in *The Varieties of Reference*.

We do not have Frege's full model of the role of sense in communication; for we do not have the thesis that communication between speaker and hearer requires them to think of the referent *in the same way* (in any plausible or natural sense of that phrase). The nearest we come to the full Fregean model is with expressions like "here" and "now"; the furthest we move away from the full model is with expressions like "I" and "you". (Evans 1982: 316)

There is a tension between this and some of the other things that Evans says about communication. A few pages before the quoted passage Evans comments that "it is a fundamental, though insufficiently recognized, point that communication is *essentially* a mode of the transmission of knowledge" (Evans 1982: 310). It is true that Evans is careful to elucidate this transmitted knowledge in a de re manner, so that what is transmitted in a sentence of the form "xis F" is knowledge of x to the effect that it is F. But many would agree with Frege that the transmission of knowledge in communication is best seen in terms of a fully determinate de dicto proposition. If I know of x that it is F, then there must be some way in which I am thinking of x. Suppose that it is different from the way in which you think of x. Then it is perfectly possible for the proposition that I apprehend when you tell me that x is F to be a proposition that you would deny (and would be perfectly rational in denying). It is hard to see how this can count as an instance of the transmission of knowledge. Applying this back to the case of the first person yields further support for the symmetry constraint.

VI RETHINKING EVANS'S ACCOUNT OF THE SENSE OF "I"

Evans's account has three different strands which he calls the information component, the action component, and the location component. The information component has to do with the thinker's sensitivity to sources of information about himself that are immune to error through misidentification. The action component has to do with the thinker's disposition to take a certain category of thoughts about himself to have immediate implications for action—those thoughts whose typical linguistic expression would be the firstperson pronoun. The location component has to do with the thinker's ability to locate himself within the spatio-temporal order. Of these three components it is the first that most directly generates the difficulties identified in the previous section. My sensitivity to information about myself that is immune to error through misidentification is not something that you can share. Hence, building such sensitivity into the sense of the first-person pronoun will have the result that "I"-thoughts are incommunicable.

There are two motivations for including an information component in an account of the sense of the first-person pronoun. The first is an analogy between the first-person pronoun and other referring expressions. Evans's general picture of the sense of referring expressions involves the holding of an information link between thinker and object. In the case of "I", the most obvious information links are precisely the ones that Evans discussesintrospection, somatic proprioception, autobiographical memory, and so on. But these information links do not perform the same function as the information links invoked to explain other referring expressions. On Evans's general picture, information links allow the subject to identify a particular object as the referent of the referring expression. But the whole point of the information links that he discusses with respect to the first-person pronoun is that they do not make possible any sort of identification of a given person as oneself. They are immune to error through misidentification precisely because the judgments to which they give rise do not involve any such identification. So, the work that the information component is doing in explaining the sense of the first-person pronoun cannot have to do with making possible identification of the thinker as the referent of the first-person pronoun.

Evans's second motivation for introducing an information component into an account of the sense of the first-person pronoun is the idea that we can understand the truth conditions of certain thoughts about ourselves simply by understanding how we would determine their truth-value on the basis of the deliverances of sources of information that are immune to error through misidentification. So, for example, there need be nothing more to my grasp of the truth condition of the thought that I am in pain than my ability to judge that I am in pain when I feel pain—and there need be nothing more to my grasp of the truth condition of the thought that I am thinking about my car than my ability to judge, on the basis of introspection, that I am thinking about my car.

But why should this have anything to do with the sense of the first-person pronoun? Surely what is at stake here is not a way of thinking about myself, but a way of thinking about what it is to be in pain-or to be thinking about a certain object. Consider the corresponding linguistic expressions "----is in pain" and "-----is thinking about—". Evans holds that we understand these linguistic expressions by understanding what it would be for them to be "completed" by arbitrary names of objects to yield full sentences expressing a thoughts. Understanding thoughts is, of course, a matter of understanding what it would be for them to be true. In the case of the predicate "-----is in pain", therefore, one would expect this understanding to include the difference between its first- and third-person application criteria—the difference between applying the predicate to ourselves on the basis of how we feel and applying it to others on the basis of what they say and how they behave. Similarly for the dyadic predicate "-----is thinking about-----". Here too we would expect an account of what it is to understand this predicate to distinguish between the introspective grounds on which one might apply it to oneself and the rather different grounds on which one might apply it to others. It looks, therefore, as if the sensitivity to peculiarly self-specifying information that Evans builds into his account of the sense of the first-person pronoun is already accommodated in any account of what it is to understand the relevant predicates. It is true that a thinker can only understand thoughts that might be expressed with sentences of the form "I am F", where "F" is a predicate that can be applied to oneself on the basis of information that is immune to error through misidentification, if they are appropriately sensitive to the relevant information and know that they are. But this sensitivity is part of what it is to understand the predicate "-----is F", rather than forming part of the sense of the first-person pronoun.

Neither motivation for introducing the information component into an account of the sense of the first-person pronoun is compelling. There is a significant disanalogy between the first-person pronoun and other referring expressions in this respect. Evans is surely right (given his assumptions about Russell's principle and the *Evans and the Sense of "I"* 189

dependence of successful reference upon discriminating knowledge) to propose a general model for referring expressions in which the ability to identify the object referred to depends upon the existence of some sort of information link between thinker and object. And he is right also to stress that we do have dedicated sources of information about ourselves. But he is wrong to think that these information sources serve as an information link underpinning mastery of the first-person pronoun of the sort that we find underpinning mastery of other referring expressions.

Suppose, then, that we drop Evans's information component from an account of the first-person pronoun. One immediate consequence is that the sense of the first-person pronoun ceases to be so obviously private and incommunicable. But can we move from this to an account that explains both what it is to utter the first-person pronoun with understanding and what it is to understand someone else's utterance of the first-person pronoun? And, in particular, can we move to an account of the sense of the firstperson pronoun that meets the symmetry constraints discussed in the previous section?

Let us start with what is involved in understanding someone else's utterance of the first-person pronoun. One needs to be able to identify the utterer of the token sentence-one needs to know who uttered the sentence in question. But in what does this knowledge consist? It is not necessary to be able to track down the physical source of the token, since it is frequently enough to know what one would have to do to track down the physical source of the token. If I overhear someone whose voice I do not recognize say "I" in the next room, then I know what I need to do to track down the physical source. I need simply go next door. Yet the ability at stake is clearly the ability to locate a given individual-to know what one would have to do to arrive at a position where one can apply the token-reflexive rule. Equally clearly, the location that needs to be determined is *egocentric*. What matters is that I should be able to locate the utterer of "I" relative to my own position. In fact, the canonical way of understanding a token of "I" involves being able to put oneself in a position relative to the speaker such that one can think about him in second-person terms (it is irrelevant that one might not be the addressee of the sentence in question) or in those third-person terms in the expression of which "he" functions primarily as a demonstrative pronoun (more or less synonymous with "that man").¹⁵

This account of what it is to understand someone else's utterance of "I" clearly meets the symmetry constraint. It also preserves one element in Evans's account of the sense of the first-person pronoun. Evans thinks that an integral element in one's "I"-idea is the ability to think of oneself as located on a spatio-temporal map of the world—to be able to identify oneself with an element in the objective order. My understanding of another person's utterance of "I" is constituted by my ability to identify a particular person at a particular spatio-temporal location as the utterer of the token. If I am not able to locate (or at least be in a position to go about locating) the utterer of the token within the objective order, then I cannot properly be understood as having understood the relevant utterance. Everything said so far, therefore, is fully compatible with Evans's account of the sense of the first-person pronoun (now that the information component has been dropped).

But this, after all, is the easy case. Does it give us any clues regarding an account of what it is to use "I" with understanding? Is there a comparable requirement on one's own self-referring use of "I" that one be able to locate oneself within the objective order? It is tempting to think not, since I can truthfully utter the sentence "I have no idea where I am". More generally, an objection to any attempt to place substantive requirements upon the ability self-consciously to refer to oneself in thought or speech is that it seems possible for a thinker who does not meet the requirement to wonder what is happening. Surely, it is argued (most famously in Anscombe 1975), an anaesthetized amnesiac temporarily deprived of the use of his senses is perfectly able to reflect upon what is happening to him.

It is clear why we do not want to deny that the anaesthetized amnesiac would be able to think about himself self-consciously.¹⁶

¹⁵ It is true that not all cases of apparently successful communication involving "I" will take this form. Telephone conversations with people one does not know proceed perfectly well when one is in complete ignorance of the location of the other person. But there are of course degrees of success in communication and degrees of understanding. A conversation can proceed perfectly well in circumstances in which one might hesitate to attribute full understanding to the parties concerned. The notion of understanding is an ideal to which we may approximate more frequently than we attain it.

¹⁶ Nor is Evans committed to denying that the anaesthetized amnesiac would be able to think about themselves. It is true that Evans does countenance the possibility

Such a person could be asking questions to which there are perfectly determinate answers ("Where am I?", "Are my arms above my head?", "Why is this happening to me?"). And if he can think them, then he can obviously think about himself. Clearly, then, it cannot be a condition of his being able to think about himself that he know the answer to these questions, or any like them—and, *a fortiori*, nor can knowing the answer to any such question be part of the sense of the first-person pronoun. But does this rule out some form of location component in the sense of the first-person pronoun? In one obvious sense it does. The anaesthetized amnesiac cannot identify himself as a particular person at a particular spatio-temporal location. But let us look again at what Evans says.

It seems to me clear that as we conceive of persons, they are distinguished from one another by fundamental grounds of difference of the same kind as those which distinguish other physical things, and that a fundamental identification of a person involves a consideration of him as the person occupying such-and-such a spatio-temporal location. Consequently, to know what it is for $[\delta = I]$ to be true, for arbitrary δ , is to know what is involved in locating oneself in a spatio-temporal map of the world. (Evans 1982: 211)

The requirement here is not that one actually be able to identify oneself as the person occupying such-and-such a spatio-temporal location, but rather that one knows what is involved in locating oneself in public space. If one understands that one is such a person, then one can understand the truth condition for any sentence of the form "I = δ " where " δ " is a name or description picking out a particular person. It seems clear that on *this* way of understanding the location component, it is easily satisfied by the anaesthetized amnesiac.

It might seem that we have accommodated Anscombe's problem case only at the cost of thinning down the idea of a location component in the sense of the first-person pronoun so far that it becomes impossible to meet the symmetry requirement. I understand

that an "I"-thought might "fail to net any object at all" (1982: 253). But the examples he gives are ones in which either the information links to one's own body are systematically distorted (as would happen if one were receiving kinaesthetic information from someone else's body) or the thinker discovered himself to be a brainin-a-vat. Evans has the resources to deal with Anscombe's case. I discuss one such resource below.

someone else's utterance of the first-person pronoun just if I can locate the utterer of "I" relative to my own position—and this, I suggested, is effectively the ability to put myself in such a position relative to the utterer of "I" that I can address him in the second person. But this seems very different from the bare knowledge of what is involved in locating oneself in public space. After all, that bare knowledge can take the form simply of knowing that one is *here*, without one's being able to identify here as anywhere except the place where one is. The utterer of "I" does not seem to need any substantive capacity to locate himself. So how can there be the equivalence between what the utterer understands and what the hearer understands demanded by the symmetry thesis?

At this point we can turn to one of the most suggestive aspects of Evans's thinking about the first person: namely, his analysis of egocentric spatial thinking and the relation between "here"-thoughts and thoughts about space in general. Evans repeatedly emphasizes the interdependence of "I"-thoughts and "here"-thoughts. Consider the following passage.

It is not the case that we *first* have a conception of which material object in the world we are (or what it would be to establish that), and *then* go on to form a conception of what it is for us to be located at a particular place. It is true that [p = here] is the same thought as [I am at p]; but this does not mean that I identify p as *where I am*. This would raise the question 'How do I identify myself, and make sense of my being located somewhere?', but—if we had to keep the capacity to grasp 'here'-thoughts out of the picture—would make it impossible to answer it. (Evans 1982: 153)

Evans's point is that the ability to think of oneself as being *here* is far richer than the trivial ability to think of oneself as being located at the place where one is. A subject cannot have "here"-thoughts in isolation. "Here"-thoughts are part of a complex network of egocentric spatial thoughts defining an *egocentric space*, in which the spatial relations of objects are coded in a frame of reference centered on the thinker.¹⁷ But nor, Evans argues, can a subject have a conception of egocentric space in isolation. This is where Evans makes one of his Kantian moves, arguing that the ability to think about space in egocentric terms cannot be separated from the subject's capacity to think objectively about space (to possess what

¹⁷ As Evans points out (1982: 157), an egocentric space is not a special kind of space, but a distinctive way of representing space.

Evans and the Sense of "I" 193

psychologists and cognitive scientists call a *cognitive map*).¹⁸ What is *un*-Kantian about Evans's understanding of spatial knowledge, however, is that he analyzes the thinker's capacity to superimpose their representation of egocentric space on their representation of objective space (and vice versa) in essentially practical terms—in terms of the thinker's ability to find his way about (Evans 1982: 165).

On Evans's account, therefore, the location component in the sense of the first-person pronoun involves an essentially practical ability to integrate egocentric and objective ways of thinking about space. This underwrites one's sense of oneself as a physical object and makes it possible to grasp the truth conditions of thoughts about oneself. One obvious consequence of this is that, provided Anscombe's anaesthetized amnesiac has retained this essentially practical ability to find his way around (even if only in a dispositional sense), then Evans's account allows him to have genuine "I"-thoughts (on the assumption that we have dropped the information component from an account of the sense of the first-person pronoun). Less obviously, but more importantly, this points the way to seeing how the proposed modified version of Evans's account meets the symmetry requirement.

My use of the first-person pronoun with understanding rests upon an ability to grasp the truth conditions of the thoughts thereby expressed, where those truth conditions are states of affairs involving a certain person (namely, myself). My grasp of which person that is depends upon my being able to locate myself within the spatial environment. Consider now your understanding of my use of the first-person pronoun. You understand my use of the first-person pronoun just if you grasp that very same truth condition-that is, the state of affairs involving me. Your grasp of which person is at stake depends upon your ability to locate that person (namely, me) within the spatial environment. As we saw when discussing Perry in section III, it is often necessary to individuate actions in terms of their ends. The reasons that lead us to individuate actions in this way extend to abilities, so that the ability you exercise in understanding my utterance of "I" is in fact the same ability I exercise in uttering "I" with understanding. And so the symmetry constraint is met.

¹⁸ For further discussion of the relations between egocentric and objective ways of thinking about space, see Campbell 1993 and 1994.

VII CONCLUSION

We see, therefore, how a modified version of Evans's account of the sense of the first-person pronoun resolves the fundamental problems that indexical expressions and indexical thoughts pose for a Fregean theory of sense. The account that Evans offers resists the temptation to fractionate the notion of sense and provides us with a suitably unitary notion of the sense of "I". It does so, however, at a price. Evans abandons the shareability of "I"-thoughts-and hence the possibility of understanding someone else's "I"-thoughts (on the assumption that one cannot understand a thought without grasping it). The price is not worth paying. Fortunately, we need not pay it, because Evans is mistaken in thinking that what he calls an "information component" is required to explain the sense of the first-person pronoun. Once we drop that element of his account, and instead emphasize the points he makes about self-location, we see that the sense of the first-person pronoun is perfectly shareable. Building on Evans's insights into first-person thought and the firstperson pronoun, we can develop an account of the sense of "I" that allows the thoughts expressed using "I" to count as fully Fregean thoughts, shareable and with determinate truth conditions. This is surely one of Evans's most enduring contributions to the philosophy of thought and language.

Information Processing, Phenomenal Consciousness, and Molyneux's Question

JOHN CAMPBELL

Ordinary common sense suggests that we have just one set of shape concepts, which we apply indifferently on the bases of sight and touch. Yet we understand the shape concepts, we know what shape properties are, only because we have experience of shapes. And phenomenal experience of shape in vision and phenomenal experience of shape in touch seem to be quite different. So how can the shape concepts we grasp and use on the basis of vision be the same as the shape concepts we grasp and use on the basis of touch?

I think this is the intuitive puzzle that underlies the question sent by the Dublin lawyer Molyneux to John Locke. This concerns a man born blind, who learns by the use of touch to discriminate cubes from spheres. "Suppose him now to gain the use of his sight. And suppose him to be presented with a cube and a sphere, of nighly the same bigness. Quaere, will he be able to tell, by the use of his vision alone, which is the sphere, and which the cube?" (Locke 1975: II. ix. 8).

In his seminal paper 'Molyneux's question' (1985b), Gareth Evans agreed in posing the underlying issue as a problem about our concepts of shape: are the concepts grasped and used on the basis of vision the same as the concepts grasped and used on the basis of touch? And he gave an argument which aimed to show and explain how it can be that it is the same shape concepts that we exercise on the basis of vision as on the basis of touch. Evans's argument uses the notion of an egocentric way of representing space, or 'egocentric space'. For present purposes, we can follow Evans in characterizing

John Campbell

egocentric space as a space defined by the axes up, down, left, right, in front and behind, and centred on the subject (cf. Evans 1982: 153-4; 1985*b*: 384). Here are the steps of his argument:

- (1) Shape concepts have their meaning in virtue of their relations to egocentric space.
- (2) Egocentric space has its content in virtue of its relations to behaviour.
- (3) Egocentric space in vision and in touch has its content in virtue of its relations to the very same behavioural repertoire.
- (4) Consequently, egocentric spatial content is of the same type in vision as in touch.
- (5) Since egocentric space is the same in vision as in touch, shape concepts have the same content whether they are acquired and used on the basis of vision or on the basis of touch.

Evans presents his conclusion, (5), as addressing the fundamental issue underlying the original Molyneux question. The argument is that if it is the same type of content, egocentric content, that we use indifferently in vision and in touch, and if we apply shape concepts on the basis of egocentric content, whether in vision or touch, then it will be true that the newly sighted man will be able to say immediately which is the sphere and which the cube. The reason is that he will simply be applying the same concept on the same egocentric basis as before.

Evans was one of the first to introduce the idea that there is a distinction between the type of representational content used in our thought and talk, which he called 'conceptual content', and the kind of content that is involved in biological information processing, which he called 'nonconceptual content'; and he tried to provide principled ways of distinguishing them, and a view of their relation to one another. (See Evans 1982, index entries under 'conceptual and nonconceptual content', especially p. 157. For an overview of the current state of play, see Gunther 2003.) This distinction immediately bears on the argument sketched above, in steps (1)-(5). For Evans takes it that the problem about shape concepts-about what is going on at the level of conceptual content-has to be resolved by looking at the relation of shape concepts to space represented egocentrically-which he takes to be a nonconceptual level of representation. I will review the basis of this distinction between conceptual and nonconceptual content in §2 below.

For the moment, we can remark that the key move here, the move that makes all this bear on the intuitive puzzle raised by Molyneux's question, is that Evans assumes that the phenomenal content of perceptual experience is to be given in terms of nonconceptual-content (Evans 1985b: 386-8). The puzzle was to understand how to acknowledge that we have a unitary set of shape concepts, exercised indifferently in all the sensory modalities. For experience of the shapes is what provides us with our knowledge of what the shape properties are, and the phenomenal character of vision seems so different from the phenomenal character of touch (cf. Grice 1962). In effect, Evans provides a way of addressing this question. Since the egocentric content of visual experience is the same as the egocentric content of tactile experience, the two senses can in principle be making the very same contribution to our understanding of shape concepts. In the relevant respect, then, the phenomenal characters of the two senses are the same. In effect, what happens here is that Evans substitutes a problem about the architecture of our information processing and its relation to conceptual thought for a problem about the relation of phenomenal experience to conceptual thought.

It seems to me, though, that the appeal to nonconceptual content does not give a convincing characterization of the phenomenal content of experience, and I will pursue this point below. Briefly, the problem is that, on the face of it, we would expect that an account of the basis of conceptual content should vield the result that sameness of content will be transparent to the subject. This is indeed implicit in the above argument, when it is assumed that sameness of the shape concepts applied on the basis of sight and of touch should mean that the subject realizes that it is the same shapes being perceived by sight as by touch. And we would ordinarily assume that sameness of phenomenal content should be apparent to the subject. If two aspects of your experience have the same phenomenal content, it should seem to you that they do. But there is no such transparency of the content involved in biological information processing. The subject need not register, in any way or at any level, sameness of the content involved in two biological information processing stages.

There is a further, related problem in Evans's discussion. He seems to assume the following principle: if an information-processing routine can be applied to the informational content of one sensory

John Campbell

modality, it can also be applied to the informational content of any other sensory modality. So if shape information can be derived from the egocentric spatial information in one modality, the subject can equally well derive it from the egocentric spatial information in another modality. This principle needs only to be spelt out to be seen to be problematic; and it does not help that the sensory information is assumed to be part of phenomenal content. I suspect, however, that what underlies the mistake here is a supposition that a phenomenal content is somehow 'central', and that the computational processes applied to any such content must consequently also be 'central' and equally applicable to any other phenomenal content.

I will take up this point in $\S4$ below, where I look at steps (1) and (5), the argument that shape concepts have meaning in virtue of their relations to egocentric space. It is, I think, wrong to suppose that shape information is derived from egocentric information in the way that Evans supposes; it is a mistake to suppose that shape concepts have their meanings in virtue of their relations to egocentric space. The main points I have made in the last two paragraphs arise even if we set that mistake aside. In $\S5$ I will argue that the relations of shape concepts, which are concepts of categorical properties.

In S_{1-3} I will discuss steps (2)–(4), in which Evans aims to establish that egocentric content is of the same type in vision as in touch. My aim here is to set out the basic problem about the type of content that Evans is supposing that we find in perceptual experience. On the one hand, you might take it to be conceptual content, in which case whether we have the same contents in different sensory modalities will be transparent to the subject; but we lose any picture of the relation of experiential content to the content ascribed in information-processing accounts of perception. Alternatively, you might take experiential content to be the kind of law-governed content found in information-processing accounts of perception; but then we have no reason to suppose that sameness of content will be transparent to the subject.

However, Evans offers an argument for the transparent sameness of egocentric content in the different sensory modalities which might seem to transcend this dilemma, so I begin by looking at it, in §1.

I. EGOCENTRIC SPACE

Evans gives a brief summing up of his argument that egocentric spatial content is the same whatever the sensory modality. The argument is that '[t]here is only one egocentric space, because there is only one behavioural space' (Evans 1982: 160; cf. Evans 1985b: 389–90). The idea here is that each sensory modality has its spatial content in virtue of its relations to behaviour. Moreover, it is not as though each sensory modality has its own particular repertoire of behaviours associated with it. That is, it is not as though there is one set of behaviours which are particularly appropriate in response to visual input, another set of behaviours which are responses to auditory input, and so on. Rather, there is a single set of behaviours which are suitable as responses to spatial perceptual input in whatever sensory modality. Evans thinks that this establishes that egocentric spatial content is the same in all sensory modalities.

To evaluate this line of argument, let us begin with the idea that egocentric space has its content in virtue of its relations to behaviour. I think that there is a confusion which can make this idea seem more straightforward than it is. Suppose we consider a submarine commander who, let us assume, uses latitude and longitude co-ordinates in instructing the ship's computer and navigation systems as to just how the craft should move. It is, we suppose, quite a sophisticated submarine; which at the highest level of control uses only latitude and longitude co-ordinates. Moreover, we can suppose that most of the information the commander has about where he is and where he wants to go comes in the form of latitude and longitude co-ordinates. In that case, latitude and longitude play a special role in navigating the ship. But this does not mean that this co-ordinate system has its meaning in virtue of its role in the control of the submarine. Rather, the co-ordinate system has its meaning entirely prior to its use in navigation, and this prior meaning is exploited when the system is used in control of the submarine.

It seems quite plausible that egocentric space plays a special role in our ordinary high-level control of our own movements. And it seems quite plausible that very often, the information that the subject has from perception about the location of this or that target comes in egocentric form. So egocentric space may play just the kind of role in the control of spatial action that I just envisaged for

John Campbell

latitude and longitude in the case of the submarine. But in this case too, it does not follow that egocentric spatial terms have their contents in virtue of their role in the direction of action. It remains possible that the egocentric terms have their meanings in virtue of some quite separate range of facts.

Notice, incidentally, that the latitude and longitude system is being used at the highest level of decision making. It may well be that the execution of a command to move to a particular destination will mean that a lot of computing machinery has to operate. It may be that this will involve translation from the high-level command into lower-level frames of reference used in more immediate control of the submarine's steering system. Ultimately, indeed, the instructions issued may be entirely non-spatial—simply to fire one or another engine, for example. The commander may know nothing of exactly what is going on at these lower levels. And the meaning of the latitude and longitude system is still not given in terms of its relations to these lower levels.

It may be that the submarine commander can be said to possess the latitude–longitude system of representation only in virtue of the fact that it plays a role in the explanation of his behaviour. Similarly, someone might be said to possess an egocentric system of representation only in virtue of the role that the system plays in the explanation of his behaviour. But in neither case would it follow that the content of either system of representation, the latitude– longitude system or the egocentric system, had to be explained in terms of its connections with behaviour.

At any rate, whatever the motivation for the idea, what are the implications of supposing that egocentric spatial content is actually constituted by its implications for behaviour? The natural way to implement that idea would be to suppose that identifying the egocentric location of an object is identifying something like an affordance of the object, in the sense of Gibson (1979). That is, knowing the egocentric location of the thing is a matter of knowing that it affords grasping if you move thus-and-so, that it affords avoidance if you move thus-and-so, and so on. And the same affordance has to be presented in the same way to the subject, whichever sensory modality is used in finding out about it, since the subject has only one behavioural repertoire.

What is doing the work here is the idea that the same affordances are being identified in the very same ways: that is, they are being identified from the perspective of the agent who may be acting to use them. So if both vision and touch, for example, represent an object as being just to one's right, they are both, on this account, representing the object as 'reachable thus-and-so'. And the 'thus-and-so' has to be spelt out in the same way both times, this argument continues, because in both cases the reaching is being specified in just the same way, from the viewpoint of the agent who may execute it.

If we accept this interpretation of the suggestion that egocentric spatial content is constituted by its implications for behaviour, we can see the force of Evans's argument that 'there is only one egocentric space, because there is only one behavioural space'. The idea here is that egocentric content identifies the locations of objects merely as affordances. But there is only one set of affordances provided by the egocentric location of an object, no matter through which modality the location is identified. This is the force of the point that there are not different behaviours associated with each sensory modality. The idea is that whatever the sensory modality, identification of the location of an object is always identification of it as providing the very same set of affordances. It is for this reason that egocentric space has to be the same, whatever the modality.

The trouble with this gloss on the content of egocentric identifications of location is that we would ordinarily take spatial location to be the categorical basis of these affordances. That is, we think that it is the relative locations of the thing and the agent that explain why it is possible for the agent to act on the thing. We do not suppose that egocentric location is actually constituted by the possibility of the agent acting on the thing. This comes out when we consider the relation between the basic egocentric frame and other systems of reference to places. It is probably true that, initially, perception merely identifies the locations of objects in a basic egocentric frame, in which locations are specified in terms merely of their relations to the subject, and not in terms of their relations to one another. It is this basic system of identifications that you might take to be a set of identifications of Gibsonian affordances. But we can also operate with egocentric terms, using an egocentric frame that is centred not on oneself, but on an arbitrary object one can see. For example, I might say that 'the window is to the right of the door', and this may be true even though the window and the door are both on my left. In this case, there may be

John Campbell

no assumption that the door itself has an intrinsic right or left; rather, I take my own right or left, and project them on to the door, using it as a reference object. Linguists describe this as use of a deictic frame of reference (cf. Garnham 1989). It can also happen, though, that I identify locations in terms of the egocentric frame generated by the axes of an object other than myself, such as another person, or a car, for example. So I might say that my bicycle is in front of and slightly to the left of a car that you and I can both see. Linguists describe this as use of an intrinsic frame of reference (Garnham 1989). The first point to notice about deictic and intrinsic place identifications is that they can be straightforwardly derived from basic egocentric information about where the various objects are with respect to me, together perhaps with some information about the shapes of the objects, from which their intrinsic axes can be derived. Indeed, this seems to be why we find it so easy, in ordinary vision, to find deictic and intrinsic locations of seen objects.

The point now about deictic and intrinsic identifications of places is that they are rich in causal significance. It might be that the relation of the door to the window has implications for the structural safety of the building. Or the relation of my bicycle to the car might affect how safe the bicycle is from being flattened by the car. But these causal implications of location can't be derived from the affordances I described earlier, concerning how I myself should go about reaching or avoiding seen objects. The natural reading of the situation is that the basic egocentric identifications of location are the grounds of affordances, rather than being constituted by these affordances, and that from these grounds of affordances we can determine further spatial relations, which ground further causal implications.

If, however, we think of egocentric locations as the grounds of affordances, then, on the face of it, we lose the argument that the spatial contents of different sensory modalities must be identifying the same affordances in the same ways. We have to acknowledge that it is possible that a single egocentric location could be identified in two quite different ways, yet still be grounding the very same possibilities of behaviour. That is, a subject could be identifying locations in one way by vision, and in a different way by hearing, and yet these two different ways of identifying locations could ground the very same behavioural reactions; they could ground the ascription to the object of the very same sets of affordances. The Molyneux's Question

difference between the identifications of places in the two sensory modalities would emerge in the fact that it could still be informative for the subject to find that it is the same range of places that is being identified by vision as by hearing. Of course, the subject might be expected to recognize that the place identifications in the two modalities typically ground the very same behavioural responses, and that those responses are typically equally successful whether based on vision or on hearing. And this means that the subject would have available an inference to the best explanation, which would argue that it must be the very same range of places that is being identified by vision as by hearing, since the very same behaviours that yield success in response to visual input also yield success in response to auditory input. But this implies that we are dealing with different ways of identifying places in the different sensory modalities, so that it is a substantive inference to the best explanation that determines that it is the same range of places that is being identified in these different ways. If it really were transparently obvious that it is the same range of places being identified by vision as by hearing, there would be no scope for there to be such an inference to the best explanation. The identity of the places identified through vision and hearing would simply be guaranteed by the mode of place identification used.

2. CONCEPTUAL VERSUS NONCONCEPTUAL CONTENT

Suppose we were to accept that there is a sameness of egocentric spatial content across sensory modalities. Under what circumstances would this sameness of content mean that the sameness of the places identified in the different modalities was transparent to the subject? I think we can focus this problem by drawing a distinction between two types of content: on the one hand, the information-processing content that scientists use in characterizing the operations of brain systems involved in perception, for example, and on the other hand, the kind of conceptual content that we ascribe to each other in our everyday common-sense talk about beliefs and desires and so on. The kind of content in which Evans was interested was supposed to be the content of consciousness, not merely brain states, and to be nonconceptual. But I think we can triangulate the kind of phenomenal content he was after by comparing it to these other two types.

There do seem to be quite sharp differences between these two other types of content. Consider the question why cognitive science ascribes content to brain systems at all. It is, after all, often argued that this is in itself a mistake, that states of the brain cannot literally represent aspects of the external environment. If we want to characterize brain states, why not simply describe the anatomy and physiology and leave it at that? One traditional reason for not leaving it there is that cognitive science seems to have found lawsrough-and-ready, ceteris paribus laws no doubt, but laws none the less-which are stated at the level of content. That is, we can, for example, describe human hearing as performing a kind of processing to establish the location of a sound, and the description may be stated at the level of content. There are laws dealing with the kinds of illusions and breakdowns to which such a system is prone. And these laws are relatively indifferent to the details of the physiology of the system in which they are realized. They would apply equally to a different species, with a quite different physiology to ours, which had none the less, and perhaps for similar reasons, developed auditory systems working on similar principles.

It is important to note that, as has frequently been observed since Fodor (1983), there is a certain modular organization to the brain systems studied by scientists (cf. Coltheart 1999). For present purposes, we can take modularity to be a matter of informational encapsulation-that is, that information processed within one system is not generally available to all other input systems-and domain specificity-that is, the various input systems are processing different sets of initial raw data. And for present purposes, we can acknowledge that the conformity of input systems to this rough working definition may be a matter of degree. That is, there may be some overlap in the raw data being processed by different systems, and on occasion there may be some capacity for one system to make use of information processed by another system. The present point is that the laws governing the processing of contents of a particular kind will, in general, be module-specific. That is: suppose, for example, you are told that there is somewhere in the subject's brain a representation of a particular stimulus as at a particular location. The significance of this representation will depend on the Molyneux's Question

laws governing the inputs which can produce such a representation and the outputs which such a representation can generate, perhaps in conjunction with other representations. So you only know the significance of the subject's brain having a representation of the location of a stimulus when you know what the relevant laws are. But the relevant laws will be, as I shall say, 'module-specific'. That is, you will know the significance of the subject's brain having that representation of the location of a stimulus only when you know in which module the representation figures. There is no general presumption that representations will be processed in the very same way, in whatever modules they figure. So to grasp the significance of the subject's brain having that representation of the location of a stimulus, we need to know in which module the representation figures, and we need to know the laws governing the processing of representations within that module.

Of course, the outputs from one module will often be the inputs to one or more other modules, so we cannot say that the only relevant laws are those governing the processing of contents within modules; we also need to acknowledge the existence of laws about the relations between the outputs of one system and the inputs to another. In effect, we have already noticed the existence of these kinds of connections, when we considered the ways in which vision can calibrate touch. And we will also have to acknowledge the importance of laws about the relations between external stimuli, input systems, and the environmental effects of actions based on processing.

In contrast to the contents ascribed to information-processing modules, there are the contents ascribed in common-sense psychology, when we attribute particular thoughts and speech acts to one another. These are generally taken to be subject to a battery of a priori constraints; certainly they are taken by Evans to be so. In particular, there is what Evans called the 'Intuitive Criterion of Difference' governing the ascription of conceptual contents:

the thought associated with one sentence S as its sense must be different from the thought associated with another sentence S as *its* sense, if it is possible for someone to understand both sentences at a given time while coherently taking different attitudes towards them, i.e. accepting (rejecting) one while rejecting (accepting), or being agnostic about, the other. (Evans 1982: 19) So if two sentences express the same thought, it must be immediately recognizable by the subject, in the sense that the subject cannot coherently take conflicting attitudes towards them.

This immediately marks a point of contrast between conceptual contents and information-processing contents. When two informationprocessing contents are contents in different modular systems, there is no guarantee that their sameness of content must be registered in any way or at any level. They may simply be in different modules, with their significance regulated by quite different sets of laws.

Moreover, even within a single module, since we are dealing only with the empirically discovered laws governing the processing of information within that module, there will be no a priori guarantee that within the module there could not be two tokenings of the very same single content—say to the effect that a particular stimulus is at a given location—such that within the module there was acceptance of one token content and rejection or agnosticism about the other.

We can see the difference between information-processing contents and conceptual contents very plainly if we consider the phenomenon of asymmetric cross-modal transfer of learning. Streri describes the phenomenon as follows: 'At the age of 5 months, babies show haptic recognition of the shape of objects which they have already seen, but it has not been possible to observe the reverse transfer' (Streri 1993: 130; cf. Streri and Pecheux 1986). These findings are puzzling if we take ourselves to be dealing here with conceptual contents in vision and in touch, which must now be assumed to be identical, to explain the ability of the infants to recognize haptically the shapes they have seen, and now assumed not to be identical, to explain the inability to recognize visually the shapes they have explored haptically. The findings are, however, relatively unproblematic, conceptually at any rate, if we take ourselves to be dealing with information-processing contents governed by laws which may vary over time as the child matures. We have simply discovered something about the empirical laws governing the contents in question.

Evans is emphatic that the egocentric spatial content of the senses is not conceptual content. So there is no reason to suppose that egocentric content will be subject to the 'Intuitive Criterion of Difference': it could be that two spatial contents were the same, yet Molyneux's Question

the subject could rationally assess them in conflicting ways. And if we think of egocentric content as law-governed, module-specific content, then there is positive reason to dispute the idea that sameness of content in different modules must be transparent to the subject. Even if content, in this sense, is 'subjectively available' to the subject, there is no reason, so far, to suppose that sameness or difference of egocentric content in different sensory systems must be transparent to the subject.

3. THE CONTENTS OF EXPERIENCE

I think the fact is that the notion of a nonconceptual content of experience comes under great pressure at this point. On the one hand, it is supposed to be a kind of foundation for conceptual content, and when we reflect on that, it can easily seem that sameness or difference of nonconceptual content must be transparent to the subject. Consider, for example, a demonstrative like 'there', referring to a perceived place. This is conceptual identification of a place, and when we refer on the basis of vision to a place, and then refer on the basis of touch to the very same place, the sameness of conceptual content has to be grounded in a sameness of nonconceptual content. And since the conceptual demonstratives are subject to the Intuitive Criterion of Difference, this makes it look as though sameness of the place identified must be apparent to the subject. And that transparent sameness of place can only be grounded in a transparent sameness of place at the level of nonconceptual content.

The problem is that the nonconceptual content of experience is usually thought to be content of the very same kind as is ascribed in information-processing accounts of perception. The idea is that this kind of information-processing content at some point becomes 'subjectively available', and that a discovery of this point is what will make the link between ordinary subjective experience and scientific accounts of perception. This is all but explicit in the very idea of a 'neural correlate of consciousness': the idea is that conscious experience has a certain representational content, and brain processing has a certain representational content, and what is sought, by those looking for a neural correlate of consciousness, is the point at which the content of the brain processing is the very

John Campbell

same as the content of experience. But this just requires that the content of consciousness should be content of the very same type as information-processing content. And, as I have been stressing, content of this type, co-ordinate with module-specific information-processing laws, is not in general transparent: sameness of this type of content need not be apparent to the subject.

Evans himself gives a vivid characterization of the relation between the content of conscious experience, information-processing content, and conceptual content, as follows:

we arrive at conscious perceptual experience when sensory input is not only connected to behavioural dispositions in the way I have been describing 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 on the informational properties of the input. When there is such a further link, we can say that the person, rather than just some part of his brain, receives and possesses the information. (Evans 1982: 158; cf. Evans 1985b: 387)

Although the passage is not fully explicit, the natural reading is that experiential content is the very same content as brain-processing content, only it is brain-processing content that is input into a 'thinking, concept-applying and reasoning system'. We can, that is, make sense of counterfactuals such as 'if this content had not been input into a thinking, concept-applying and reasoning system, then it would have been mere brain-processing content and not conscious at all'.

Just to emphasize the main point here: Evans is operating with a single generic notion of nonconceptual, informational content, which he uses in characterizing both the deliverances of conscious perception, and the information processing carried out by the brain. His characterization of the spatial content of auditory input, for example, is a characterization of this generic notion of content: 'auditory input—or rather that complex property of auditory input which encodes the direction of sound—acquires a non-conceptual spatial *content* for an organism by being linked with behavioural output in, presumably, an advantageous way' (Evans 1982: 156). And he glosses the account as follows:

So far I have been considering the non-conceptual content of perceptual informational states. Such states are not ipso facto perceptual *experiences*—that is, states of a conscious subject 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 subject of experience. (Evans 1982: 157-8)

Since there is, of course, a distinction between the case in which nonconceptual content is not the content of an experience, and the case in which the nonconceptual content is the content of an experience, this raises the question of how that distinction is to be explained. And it is here that Evans appeals to his idea that the content is the content of conscious perceptual experience when it is not only connected to behavioural dispositions in the 'advantageous' way indicated, but is also 'the input to a *thinking*, *concept-applying and reasoning system*' (p. 158). But, by his own lights, Evans has given the constitutive account of this generic type of content in advance of any appeal to consciousness or its being input to a thinking, concept-applying, and reasoning system.

This is indeed driven home in his discussion of Molyneux's question, where, as we have seen, his whole point is that we can appeal to the way in which nonconceptual spatial content is constituted by its links to behaviour, in advance of any appeal being made to the way in which this content is linked to the thinking, concept-applying, and reasoning system, in order to establish that the spatial concepts are shared across the modalities.

Incidentally, Evans uses the same general strategy in arguing that conceptual thought, unlike the content of conscious experience, must conform to what he calls the Generality Constraint. His idea here is that conceptual thought is subject to the requirement that anyone capable of grasping the thought that a is F must also be capable of grasping the thoughts that b is F, that c is F, and so on, for all the other suitable singular ways of thinking they understand; and they must be capable of thinking that a is G, that a is H, and so on, for every other suitable predicative concept they grasp. But the nonconceptual content of experience is subject to no such constraint:

It is one of the fundamental differences between human thought and the information-processing that takes place in our brains that the Generality Constraint applies to the former but not to the latter. When we attribute to the brain computations whereby it localizes the sounds we hear, we *ipso facto* attribute to it representations of the speed of sound and of the

John Campbell

distance between the ears, without any commitment to the idea that it should be able to represent the speed of light or the distance between anything else. (Evans 1982: 104)

Without any further argument, Evans then takes it for the remainder of the book that the content of perception, whether conscious or not, is not subject to the Generality Constraint. This procedure makes sense only on the supposition that we are here dealing with a single generic notion of nonconceptual content, which can be used equally in connection with conscious and non-conscious states. Since Evans aims to derive the distinctive characteristics of conceptual thought from the fact that it meets the Generality Constraint, he could hardly acknowledge that the content of conscious perception meets the Generality Constraint without maintaining that the content of conscious perception too is conceptual. But the passage just quoted is the only argument he feels obliged to give for the claim that the content of conscious perception is not subject to the Generality Constraint; and this procedure makes sense only on the supposition of a single generic notion of nonconceptual content.

On this interpretation, then, the content of experience is the same as the content of information-processing brain systems. The problem is now that, as we have seen, sameness of the contents produced by information processing in different sensory modalities will not in general be transparent to the subject, even if those contents are contents of experience. For the module-specific laws which govern those contents will not in general guarantee transparency.

You might argue that the guarantee of transparency comes just because the nonconceptual content is input into a 'thinking, concept-applying and reasoning system'. If the same concept is applied in response to non-conceptual input in two different sensory modalities, then the transparency of conceptual content implies that it will be apparent to the subject that it is the same external property or particular that is being perceived through the two sensory modalities. But that reaction stands Evans's approach on its head. The whole point of Evans's approach was to ground the transparent unity of our shape concepts in the transparent unity of egocentric space. Suppose for a moment that the transparent unity of egocentric space is somehow grounded in the transparent sameness of the concepts we apply on the basis of sight and touch. This means that we have lost the explanation Evans set out in the five-step argument with which I began, whose point was to explain how it can be that we are applying the very same shape concepts on the basis of sight as of touch. The whole strategy was to establish transparent sameness of egocentric content across the different sensory modalities, and to argue that this was the basis for the transparent unity of the shape concepts we use. This is evident in Evans (1982), where the transparent unity of egocentric space is argued for without shape concepts being mentioned at all. Or, to put it another way, if we begin with the assumption that egocentric space has no transparent unity in advance of the content being input to a 'thinking, concept-applying and reasoning system', we will not be able to establish the unity of the shape concepts we actually have merely by saying that they are responses to egocentric content.

I think the truth is that it is a mistake to approach this topic by identifying the content of conscious experience with either conceptual content or information-processing content. We have to acknowledge that there is such a thing as the phenomenal content of experience, and that it is related to conceptual content and to information-processing content. But phenomenal content does not have to be identified with either conceptual or information-processing content. On the face of it, there are three quite different sets of phenomena here—conceptual content, information-processing content, and phenomenal content—and we ought simply to articulate their relations to one another without feeling compelled to provide reductions.

Which relations ought we to be considering? We have to acknowledge that the conscious experience of a subject is causally explained, in part at any rate, by the content of the information processing carried out in that subject's brain. Part of the reason why the subject has a conscious experience with this particular phenomenal content is that the brain processing had a particular informational content. There must further be causal-explanatory relations between the subject's conceptual judgements and the contents of the underlying brain processing. The whole methodology of cognitive science as applied to human subjects depends on the idea that their verbal reports, which presumably are in general conceptual, can be explained in part by the contents of the underlying brain processing in various more or less modular sub-systems. So there are certainly relations between information-processing content and phenomenal content, and between information-processing contents and conceptual contents.

What about the relations between phenomenal content and conceptual content? As I began this essay by saying, it seems to me that this is the really difficult issue raised by Molyneux's question. On the face of it, the phenomenal contents of our experiences in different sensory modalities are quite different—isn't that why it is generally apparent to us whether we are seeing or touching an object? And on the face of it, we have the shape concepts we do because of our experiences of shape. We do not think of shapes merely as hypothetical possessors of various functional roles; rather, we take it that in experience we encounter the categorical properties themselves. How, then, could it be possible for us to acquire and use the very same shape concepts on the basis of vision as on touch, given the apparent differences in the phenomenal contents of sight and touch?

Since this is plainly a difficult question, it is natural to do what Evans does, and replace it with a more tractable problem about the architecture of perceptual information processing in various modalities, and its relation to conceptual thought. This exercise is actually helpful in addressing the harder problem, just because of the relations I have already remarked between information processing and phenomenal experience, and between information processing and conceptual thought. But the exercise will not in itself constitute a solution to the Molyneux problem.

On this understanding, then, let us look finally at Evans's picture of the information-processing architecture, and its relation to our concepts of shape. It seems to me that this picture is instructively mistaken. There are morals here for anyone attempting an assault on the Molyneux problem.

4. SHAPE CONCEPTS AND EGOCENTRIC SPACE

As I said at the outset, Evans seems to operate on the assumption that if an information-processing procedure can be applied to the conscious content of any one sensory modality, then it can be applied to the conscious content of any sensory modality. There is no basis for this assumption. If we are assuming that the phenomenal content of sensory experience is information-processing content, then the computational processes applied to that content may be modulespecific; there is no reason in general to suppose that an operation which can be carried out within one modular system can equally be carried out within all modular systems. Discussing this issue is a little bit complicated here, because the example we have to deal with is Evans's idea that the subject somehow derives information about the shapes of the objects perceived from egocentric information about the locations of their parts. The idea is that if the subject performs this computation within one sensory modality, then it must be possible for the subject to execute the computation within any other sensory modality that provides such egocentric spatial information. I think the real problem is the idea that this computational procedure must be general-purpose, rather than modalityspecific, because it is being applied to contents of consciousness. What complicates matters is that it does not seem right anyway to say that we derive shape information from egocentric spatial information. Here is what Evans says about this:

When we think of a blind man synthesising 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. (Evans 1985b: 389)

Presumably Evans does not suppose that we find the shapes of objects by articulate verbal reasoning based on knowledge of the egocentric locations of their parts. This crucial passage seems to be suggesting, rather, a subpersonal computation for finding shapes. It is not here made fully explicit how the computation of shape from egocentric location is supposed to go. The suggestion seems to be that the computation begins with the parts of the object—presumably

John Campbell

specified as already possessing their own particular shapes-and that the shape of the object as a whole is derived from this information together with information about the egocentric locations of the parts. So, for example, consider the similarities and differences between the shapes of a teacup and a bucket. The teacup has as parts a bowl and a handle at its side. The bucket has as parts a bowl and a handle over the top. Evans's proposal would then be that the relation between the handle and the bowl of the teacup is derived from information about their respective egocentric locations. Similarly, the relation between the handle and the bowl of the bucket is derived from knowledge of their respective egocentric locations. We can contrast this with a theory in which the relations between the bowl and the handle in these two cases are given in an object-centred frame of reference, using primitives such as 'over the mouth of the bowl' or 'down one side of the bowl'. There seems to be no particular reason to suppose that the derivation of this kind of information has to go in the way Evans envisages (cf. e.g. Bruce et al. 1996: ch. 9, 'Object Recognition', for a review of the possibilities here).

Suppose, for a moment, though, that Evans's picture is correct, and that shape information is derived from egocentric information in touch. And suppose that we have a subject capable of deriving shape information from egocentric tactual information, as Evans must suppose Molyneux's newly sighted man to be. Suppose too that this subject does also have egocentric visual content, and that this is transparently the same as his tactual egocentric content. It still does not follow that this subject would be capable of identifying shapes on the basis of vision. For the ability to extract shape information from egocentric information could still be modalityspecific; that is, the capacity to perform this kind of derivation might be something that the subject has in relation to touch, but not in relation to vision.

To see this, it may help to consider a case in which spatial information really is derived from egocentric spatial information, so that something like Evans's picture is correct. Suppose we go back to the distinction I drew in §1 above between a basic egocentric frame, on the one hand, and the use of deictic and intrinsic reference frames, on the other. So, for example, when I look around me and see where everything is, for the purposes of reaching and grasping myself, I am making use of basic egocentric visual information. In effect, I see where objects are in relation to me, but I am not concerned with their spatial relations to one another. If, however, you ask me where the tennis racket is, I might say, 'It is to the right of the ball', and here I am projecting my own left and right on to the ball and using that deictic frame of reference to locate the racket. And when I say 'The racket is on Bill's left', I am using Bill's intrinsic axes to generate a frame of reference, and locating the racket in that frame of reference. Now these computations of deictic and intrinsic locations are performed 'on demand' by the visual system. They have to be derived from the basic egocentric information one has in vision; they are not performed automatically: you have to look to see whether the racket is on Bill's left, for instance. (See Logan 1995 for detailed development of this point.) But it is not, either, as though the determination of deictic and intrinsic locations is a matter of explicit calculation by the subject. The subject does not, for example, have to engage in verbal reasoning to find deictic and intrinsic locations: it really is a visual matter. The subject has only to look to find out the deictic and intrinsic locations of things.

So this case seems to meet a part, at any rate, of Evans's picture: deictic and intrinsic locations are being derived from basic egocentric locations. And this derivation is a perceptual matter. Suppose now that we have a subject who can find deictic and intrinsic locations on the basis of vision alone. And suppose that this subject also has tactual information about the egocentric locations of the various objects around him. Would it follow from this that the subject is able to use this basic egocentric information in touch to find the deictic and intrinsic locations of things? Evans's answer is that the subject is bound to be able to do so. He is arguing that a subject who can extract shape information from egocentric visual information must also be able to extract shape information from egocentric tactual information, if it is transparently the same egocentric information that is presented in both sensory modalities. Just so, a subject who can find deictic and intrinsic locations on the basis of vision alone ought to be able to perform the very same operations to find deictic and intrinsic locations on the basis of touch.

Once we have set out the reasoning here explicitly, it is evident that there is a problem. The problem is that the computational procedure that is being used to derive the deictic or intrinsic information about location may be modality-specific. That is, the computational procedure may be available for the deliverances of vision but not for the deliverances of touch. It seems entirely possible that there could be a subject who could find deictic and intrinsic locations on the basis of specifically visual attention alone, but who could not compute deictic and intrinsic locations on the basis of touch alone, even though touch provided basic egocentric information.

I think it is easy to see the picture that Evans is using here. He is taking it that after visual or tactual information processing becomes conscious, once we are at the level at which the informationprocessing contents are 'subjectively available', any further operations performed on the now conscious contents cannot be modalityspecific but must be general-purpose, central-system operations. And this seems simply to be a mistake. It is true that verbal reasoning applied by the subject to information he has in perceptual awareness seems to be general-purpose. Any verbal reasoning I can perform on my visual information is also reasoning I could apply to transparently similar tactual information. But the subpersonal processing applied to visual egocentric information, whether conscious or not, may still be modality-specific and not available for use on information provided by touch.

5. SHAPE AS CATEGORICAL

There is a line of thought in the literature which runs somewhat as follows. Shape properties have causal significance. The shape of an object has endlessly many implications for how it will behave in interactions with other objects. To understand a shape concept, you have to grasp something of the causal significance of the shape property. Indeed, even to perceive a shape property, you have to grasp something of its causal significance. This is part of the point of Bennett's distinction between shape-blindness and colour-blindness (Bennett 1971). That someone is colour-blind can easily escape detection. If someone were shape-blind, however, it would affect every aspect of interaction with the surroundings; it could not escape notice.

The idea, then, is that the causal significance of a shape property is the same whether it is identified on the basis of vision or on the basis of touch; the idea is that round things roll, whatever the modality through which they are perceived, and that to perceive something as round, in whatever modality, you must perceive it as having a tendency to roll. So you might argue that this causal significance to the property can be constant across the sensory modalities, even though the appearance of the object varies. This seems to be something like Judith Jarvis Thompson's idea in her discussion of Molyneux, where her point is that even the newly sighted subject, if he really is seeing the shapes of the objects before him, must grasp that the properties perceived will have the same causal significance, whether they are perceived by sight or touch (Thompson 1974). In contrast, there could be no such thing as perceiving a colour through some modality other than vision, because all there is to a colour is what is given in perception, and the colour perception has no specific causal significance which could be held constant and associated with an appearance in some non-visual modality. In the absence of causal significance, the sensory appearance of the colour is thought to be modality-specific.

Evans developed a version of this idea in 'Things without the mind', when he spoke of shape properties as embedded in a primitive mechanics of our surroundings: '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' (Evans 1980*a*: 269). In contrast, 'no single *sensory* property can be defined in relation to different senses' (Evans 1980*a*: 270).

One way of pursuing this line of thought would be to ask whether the shuffle through egocentric space is really essential to Evans's approach. His idea was to argue that shape concepts are tied to egocentric representations, and that 'there is only one egocentric space, because there is only one behavioural space'. But couldn't we argue directly: 'there is only one system of shape concepts, because there is only one behavioural space'? That is, you might argue that the implications of roundness for how you interact with the object are exactly the same in vision as in touch, so you must perceive the object as yielding exactly the same affordances, whether you see it or touch it; and that is all there is to seeing it as the same shape again. In effect, this is a form of the idea that to grasp shape concepts is to grasp the causal significance of shape properties; the proposal is that this grasp of causal significance is provided by a grasp of the affordances of objects.

One problem with this is that the ability to move and act on objects in ways appropriate to their shapes seems to be quite different from the ability to apply shape concepts explicitly to them. There are patients who can reach and grasp successfully, while being incapable of successfully comparing the shapes of two seen objects. And there are patients who are incapable of successful reaching and grasping, who can none the less correctly compare and contrast the shapes of seen objects (Milner and Goodale 1995). And ordinarily we think that our grasp of shape properties is not exhausted by our grasp of causal significance. An explicit grasp of shape concepts is not merely a matter of making articulate the causal connections we implicitly grasped in our unreflective manipulations of objects. We do not think of an object's possession of a shape property as a matter merely of the object having a collection of dispositions to behave in various ways, or as a matter of the object merely being disposed to be affected by us in various ways. This comes out very clearly when you think of what happens when there is a change in the shape of an object. Suppose, for instance, that you take a piece of paper and fold it into the shape of an aeroplane. Many of the dispositional properties of the piece of paper have now changed: it has various tendencies it did not have before. If you really thought that there was no more to the paper having a shape than its having such tendencies to behaviour, you would have to suppose that the dispositional characteristics of the paper had somehow been affected directly, and somehow affected en masse. But we have no picture of how you could affect the dispositional characteristics of an object except by affecting the grounds of those dispositions; and we would ordinarily take it that changing the shape of the paper is changing the categorical grounds of those various dispositions. And what makes it so compelling that we have encountered shape as a categorical property is that our phenomenal experience seems uncontrovertibly to be experience of shape as categorical, not merely experience of a collection of unsubstantiated threats and promises.

I think it is fair to say that the current philosophical literature on Molyneux manages only to point, in one way or another, to the sameness of the collections of dispositions that are associated with the shape properties we ascribe on the bases of sight and touch. If, though, we assume that the shape concepts we ordinarily apply on the bases of vision and touch are concepts of categorical properties, Molyneux's Question

rather than merely collections of dispositions, then we need to know more than that the collections of dispositions we are ascribing on the bases of sight and touch are the same. We need to know that our phenomenal experience in sight and in touch confronts just the same categorical shape properties, in just the same ways. And so far, that question is still wide open.

"Another I": Representing Conscious States, Perception, and Others

CHRISTOPHER PEACOCKE

What is it for a thinker to possess the concept of perceptual experience? What is it to be able to think of seeings, hearings, and touchings, and to be able to think of experiences that are subjectively like seeings, hearings, and touchings?

This question is of philosophical interest for multiple reasons. To understand, explain, and predict the thought and action of others, you must know what they perceive. This requires you to possess the concept of perception, or at least to represent in some form that the other person perceives. Each of us every day rests his life on his correct application of the concept of perception. When you cross the road, or drive, your future depends on your ability to know that someone else sees you.

The concept of perception is also crucial to more first-personal projects of thought. To assess critically the way you reach your own judgments, to revise and improve your methods of reaching beliefs, requires you to be capable of thinking of the perceptual experiences that led you to make or withhold various judgments. For this too it is necessary that you be capable of thinking of your own perceptions.

The question of what it is to possess the concept of perception is also of interest to the philosophy of mind more generally. Perception

This material was presented in lectures at New York University in 2003, and I drew upon it in the first of my Immanuel Kant Lectures at Stanford University later in the same year. My thanks to José Luis Bermúdez, Michael Bratman, John Campbell, Victor Caston, Mark Crimmins, Alison Gopnik, Stephen Schiffer, Elliot Sober, and an anonymous referee for helpful comments.

is one of the mind's states that relate it most directly to the non-mental world. Can a good treatment of possession of the concept of perception provide a model for possession of concepts of other mental states with distinctively close relations to the world? Do features of a good treatment generalize? And do they permit us to make sense of the striking empirical phenomena displayed by children's acquisition of the concept of perception? These are some of the questions I will be attempting to address.

A perceiving thinker who has the capacity to appreciate that others also perceive is on the way to thinking of others as subjects like himself—to thinking of another person as "another I", in Zeno's phrase. "Another I" was reportedly Zeno's answer to the question "What is a friend?". If we strip the notion of thinking of someone as "another I" of the elements of identification and sympathy that Zeno no doubt intended, Zeno's phrase captures perfectly what is involved in thinking of another as a subject like oneself. It is a real challenge to say what is involved in such thinking. I will try to indicate in the course of this chapter points at which the approach aims to contribute to meeting that challenge.

I start by considering the first-person case, that in which a thinker judges that he himself sees. The case falls within a general area to which Gareth Evans made original, important, and influential contributions, notably on the self-ascription of belief (1982: 225–6). Evans showed how treatments that are in various respects outward-looking do not merely accommodate the distinctive epistemic features of first-person thought. An outward-looking treatment is actually required if we are to do justice to those distinctive epistemic features. While Evans's own remarks on the different topic of the self-ascription of belief, and though I shall be offering an alternative account, I hope that what I present here respects the generally outward-looking reorientation that he recommended.

After considering the first-person case, and some of the epistemic and metaphysical ramifications of the treatment I offer (\S I-2), I go on to compare it with Evans's account (\S 3). From that I move to discuss the relation between first-person and third-person ascription and to the explanation of some developmental phenomena (\S 4). I conclude with a discussion of the extension of the model presented to the self-ascription and other-ascription of action and intentionality (\S_5).

1. THE CORE RULE

Aristotle held that it is by sight that you perceive that you see.¹ The heart of Aristotle's idea seems to me right, provided that we understand it as follows: it is by sight that you know that you see. Suppose you see that

That desk is covered with papers.

This visual knowledge about the world gives you a good reason to make the self-ascriptive judgment

I see that that desk is covered with papers.

This is a transition you are entitled to make, from a conscious state you enjoy to a judgment. If a thinker comes to judge, by this means, that he sees that that desk is covered with papers, his judgment can thereby be knowledge.

This is the starting point of a general model of self-ascriptive knowledge of one's own perceptual states. Because the thinker sees that

p,

he moves, rationally, to the judgment

I see that *p*,

and thereby gains knowledge that he so sees. If a thinker comes to judge that he sees that p in this way, and does so for the reason that he sees that p, then he is following what I call the Core Rule. More specifically, it is the Core Rule for vision, for the case of seeing-that. One can equally formulate the Core Rule for other sense modalities. Here sense modalities are regarded as individuated by their phenomenology, rather than by the identity of the sense organs whose states cause perceptions in the modality.

Following the Core Rule for seeing does not require the thinker to have the concept of seeing-that in advance. It just requires a

```
<sup>1</sup> On the Soul 3.2, 425<sup>b</sup> 12–17.
```

222

"Another I"

differential sensitivity to the cases in which one sees that something is the case, as opposed to perceiving it in some other modality, or knowing it not through the senses at all. A thinker may also be in error about whether a state is a seeing-that. But in any case in which he seems to be seeing-that something is the case, he is entitled, *ceteris paribus*, to make the transition to a self-ascription of a seeing.²

It would be a misunderstanding of the Core Rule to think that following it involves making a transition from a belief or judgment that one is seeing. Rather, following the Core Rule involves making a transition from a seeing-that itself. Since the conclusion of the Core Rule is that one sees that p, that misunderstanding of the Core Rule would construe it as making a transition from one content to the same content again. It would also be a transition from a state that presupposes that the thinker already has the concept of seeing.

One can also formulate a Core Rule for seeing an object, as opposed to seeing-that. Suppose our subject x sees object o, under mode of presentation m. Then he is entitled to judge

I see *m*.

Here of course the mode of presentation m is employed in our subject x's thought, rather than mentioned. The resulting judgment is about the object presented under m, namely the object perceived, rather than being about m. m might be expressed linguistically by the phrase "that door over there", accompanied by a pointing gesture. A full characterization of m would specify the egocentrically identified apparent location of the perceived object in relation to the subject, and the way in which the object is perceived. In both this most recent case of object-seeing, and the preceding case of seeing-that, it is the subject's seeing, of the respective kinds, that makes rational the subject's judgment about his seeing.

What I have given so far can be described as the positive part of the Core Rule. There is also a negative part, having to do with the conditions under which a thinker is entitled to judge that

² Here, as in the case of entitlement to perceptual beliefs, I would argue that the factive entitling state is more fundamental in explaining the nature of the entitlement relation than is the non-factive state of having an experience as of seeing that p. I offer some reasons in support of this position in Peacocke 2004: ch. 4, sect. I(d).

Christopher Peacocke

he does not see that p. If a subject is not in a position to judge, knowledgeably, that p, simply by virtue of what he sees to be the case, then he certainly does not see that p, and no further information is needed to establish that he does not see that p. If a thinker does not see that p, then he is entitled to judge

~(I see that p).

I call this the negative component of the Core Rule.

Here there is a difference between the concept of seeing-that and an observational concept. No such negative clause as we have just given for the case of seeing-that holds for an observational concept. From the fact that some speck or tiny dot, for instance, is not experienced by the subject as square, even when it is being observed, it does not follow that it is not square. Its shape may just be too small to see, or be perceived in any other of the subject's sense modalities. This difference is one of the marks which distinguishes possession of an objective concept of things in the world from possession of a psychological concept like seeing-that.

I suggest that following the Core Rule for any given sense modality is part of (one clause of) the possession condition for the concept of perceptual experience in that modality. To possess the concept of visual experience, the thinker must be following the Core Rule for vision; and so forth.

The Core Rule is not, and could not be, an exhaustive account of what it is to be able to judge the content "I see that p". That content contains the first person, and the present tense, which also have a life outside judgments of "I see that p". The Core Rule is just one piece of a jigsaw. Other pieces of the jigsaw are required to have a full picture of mastery of "I see that p". The other pieces would be accounts of mastery of the other conceptual constituents of "I see that p". It is a more general task in the philosophy of mind to describe these other pieces correctly, and to show how they interlock to form a full picture of mastery of "I see that p".

I further suggest that what I shall call "the Extended Core Rule" for vision is a component of the possession condition for the concept of visual experience, considered as applicable both in perceptual and in the illusory, or more strictly non-perceptual, case. The Extended Core Rule, in the case of vision, states that if the thinker is in a state that is subjectively as if he sees that p (at least in

respect of his visual experience), or subjectively as if he sees an object given under mode of presentation m (in respect of his visual experience), then he is entitled to judge

I have a visual experience as of p's being the case

or

I have a visual experience as of m

respectively. A subject's judgment of such a content, made for the reason that he is in the entitling state, can in ordinary circumstances be knowledge.

The Extended Core Rule will, perhaps surprisingly, not cover all cases in which someone is entitled to self-ascribe an experience with a given content. Consider an experience as of looking at the "impossible" object constructed by Penrose. This is a triangular 3D model, similar to prototypes drawn by Escher, which when viewed from a certain angle gives an experience in which corner A seems to be closer to the viewer than corner B, corner B seems to be close than corner C, and yet corner C seems to be closer than corner A.³ (It is not really so, of course.) Now a thinker cannot soundly reach a self-ascription of this experience by regarding it as subjectively of the same kind as an experience in which he sees that this content holds. Since the content is inconsistent, there are no such genuine seeings that it holds, nor could there be. Hence there are no experiences that are subjectively similar to such genuine seeings.

One way to attribute the correct content to the experience, e, of seeing the model is as follows. (I do not claim it is the only solution to the problem; there may well be others.) e is subjectively similar to genuine seeing e' that A is closer than B; it is subjectively similar to a genuine seeing e'' in which it is seen that B is closer than C; and it is subjectively similar to a genuine seeing e''' in which it is seen that C is closer than A. The content of e is thus determined by its subjective similarity relations to several genuine seeings, and not all of these seeings can be identical with one another. We call this "the multiple similarity" solution to the problem. We will henceforth take the Extended Core Rule to

³ For photographs of Penrose's "impossible object", see Gregory 1974: 369; 1970: 54–7.

employ a notion of subjective similarity for an experience that allows such similarity to be determined by multiple similarities to different genuine seeings.

There are many attractive consequences of incorporating the Core Rule into the possession condition for the concept of experience.

(i) It explains and justifies the sense in which one's own perceptions are not given to one in any mode other than is made available simply by the ability to have the perception itself. A *fortiori*, the perception is not given in some further perceptual mode. Despite some divergences to be noted later, this is a point on which I am in agreement with Gareth Evans when, in *The Varieties of Reference*, he writes: "[The subject's] internal state cannot in any sense become an *object* to him. (He is *in* it.)" (1982: 227).

Evans's remark is a little Delphic, but it has a natural elucidation. Whenever we perceive some spatial, material object or event, we perceive it in some sense modality. When something is perceived in some sense modality, it becomes an object to the subject. The modality in which one perceives some particular chair—be it by sight, or touch— is not in any way a priori determined by the object or the event itself. In the case of a particular perception, however, there is a way in which the perception is given in thought that does not involve any sense modality not fixed by the event itself. The mode in which the perceptual experience is given to the thinker who enjoys the experience is a priori determined by the perception itself. No further sense modality is involved. I refer to this feature of thought about perception as its "unadorned character".

What is the explanation for the difference between the unadorned character of a subject's thought about his own perceptions and the adorned character of his perceptual thought about spatial, material objects and events? Perceptual experience is itself a conscious state that can thereby itself function as a reason for the thinker to make judgments. It can enter the possession condition for concepts in a way that spatial, material objects, events, or states of affairs in themselves, not considered as given in any particular sense modality, cannot.

(ii) Incorporating the Core Rule into the possession condition is the first step towards capturing the respect in which the concept of perception is first-personal. If the Core Rule is part of the possession condition for the concept of perception, then there is a clause dealing specifically with first-person application in the possession condition.

It is important to formulate sharply the sense in which the concept of perception is first-personal, if the Core Rule is correct. Quite generally, it is not sufficient for a concept F to be first-personal that there is a special way of coming to know that one is F oneself. There is a special way, in ordinary circumstances, of coming to know that one is touching one's own toes, but the general concept x is touching x's toes is not one that involves the first person in any deep way. One's knowledge of what it is for an arbitrary thing to be touching its toes does not in itself have specific connections with the first person. The deeper sense in which the first person is involved in the general concept of seeing something to be so is that one's knowledge of what it is for an arbitrary thing to have that property makes reference in one way or another to what is involved in first-person ascription of that property.

I say "makes reference in one way or another", because there is more than one way in which there can be such a connection to first-person ascriptions. One way is that so famously criticized with what justice, we will touch upon later—by Wittgenstein: the idea that your conception of what is involved in another person's having a certain sensation is that they are having the same type of experience as you when you are in pain—that is, when you can truly self-ascribe "I am in pain". But this is not the only way in which there can be a special connection between the understanding of the general property and the first person, and I shall describe another way a few paragraphs hence.

For enthusiasts of the study of first-person thought, I note also that the occurrence of the first person in "I see that p" when it is reached in this way is representationally independent, in the sense I used in *Being Known* (1999: 266–74). That is, when the thinker is following the Core Rule, his reason for judging as he does is not that he is in some state with the representational content "I see that p", which he then takes at face value. His reason is simply his being in the state of seeing that p.

(iii) The clause containing the Core Rule can explain why selfascriptions of perception made in this way are rational, and can yield knowledge. Any context in which a thinker follows the Core Rule for, say, the visual case, will be a context with respect to which the self-ascription "I see that p" will also be true. The entitlement to make a self-ascription of a seeing in the given circumstances respects the general principle that, corresponding to every entitlement, there is an objective norm of correctness.⁴ Self-ascriptions of seeings made by following the Core Rule are correct.

This is so for a priori reasons. In the spirit of a rationalist account, the account also holds that transitions respecting the Core Rule lead to true belief because of the nature of the states and concepts involved in the transition.⁵

(iv) All experiences with representational content, whether genuine perceptions or illusions, are, in respect of the sense modalities in which they occur, subjectively as if they are perceptions. (This is why there is such a thing as taking perceptual experience at face value in the first place.) Not only is this a feature of the subjective experiences themselves; it is also a feature that is immediately obvious to us when we think of perceptual experience as perceptual experience. If our account of possession of the concept of perceptual experience incorporates the Core Rule and the Extended Core Rule, we can explain this fact. The Extended Core Rule implies that anything that is thought of as a perceptual experience is thought of as the same, subjectively, as an experience in which one genuinely perceives something to be the case. Incorporating the Extended Core Rule and the Core Rule into the account of possession of the concept of perceptual experience explains our appreciation of the primacy of the genuinely perceptual case in the phenomenology of perceptual experience. This primacy of the fully veridical case must be present in any other domain to which the Core Rule and the Extended Core Rule generalize.⁶

(v) Incorporating the Core Rule into the possession condition for the concept of seeing plausibly implies that one cannot fully possess the concept of seeing unless one knows what it is like to see. A plausible account of knowing what it is like to be in a given kind

⁴ There is a neo-rationalist treatment of these issues in Peacocke 2004: ch. 1.

 $^{^5}$ For more on such rationalist approaches, see Peacocke 2004, especially the second principle of rationalism, in ch. 2.

 $^{^{6}}$ I would argue that this feature is present for the generalization given in \$5 below.

of conscious state is that one possesses a capacity to recognize that one is in that state, on the basis of being in that state. But this is precisely what one does in following the Core Rule.

(vi) Perceiving that p is certainly an externally individuated state, for many reasons. Whether someone is perceiving that p depends on their relations to external states of affairs. Perceiving that p is a form of knowing that p, and whether one knows something depends in part on what could easily have been the case (on what happens in nearby possible worlds, as one says). What could easily have been the case is something that depends on multiple conditions concerning matters far outside the perceiver's head. If, as I am suggesting, the concept x perceives that p is individuated by its connections with the externally individuated relation of perceiving that p, then it follows that the concept is also externally individuated. So this is another case in which not only the intentional content of a state is externally individuated, but so is the psychological relation to the intentional content.

More specifically, on the present treatment the concept of perception is what I have called "instance-individuated", in the sense I discussed in "Explaining perceptual entitlement".⁷ Although a possession condition for the concept of perception that incorporates the Core Rule emphatically does not treat it as an observational concept, it does share one feature with observational concepts. It entails that in order to possess the concept, the thinker must be willing to apply the concept in response to instances of the concept, have this property. As one might put it, we have here an internal externalism. This internal externalism is consistent with the unadorned character of a subject's thought about his own perceptions.

(vii) As Mark Crimmins noted to me, a thinker can employ the Core Rule for seeing without having much idea at all of how sight works, either of its neurophysiological and computational bases, or of light as the environmental medium of transmission of the information of visual information. This attractive feature will be present in some of the later applications of the Core Rule.

⁷ In Peacocke (forthcoming) and in Peacocke 2004: chs. 2–4.

(viii) The Core Rule vindicates the Aristotelian-like doctrine that it is by sight that you know that you see. It does this without any regress in the content of seeing, and without any attribution of reflexivity in the content of the seeing.⁸

The Core Rule also has some implications for the theory of epistemic entitlement and norms. If the Core Rule is part of the possession condition for the concept of perceptual experience, then the most fundamental way of coming to know that, for instance, one is seeing something is by first making a perceptually based judgment about the non-mental world beyond oneself. For this reason, one can classify this position on self-ascription of experience as an "Outside-In" theory.

It follows that it cannot be correct to say that our basic means of knowing by perception about the external world is first by knowing that we see something, or see something to be the case. On the present approach, that is precisely the reverse of the correct order of epistemic entitlement.⁹

The second implication of the position for epistemic entitlement concerns the transmission of warrant. It is sometimes said that warrant cannot be transmitted from an observational judgment that p made on the basis of visual perception to the conclusion "I see that p". On the present position, warrant is transmitted in that transition. In fact, the paradigm of entitlement is when a judgment is made in accordance with the possession condition for some concept in the content of the judgment, and the truth-preservingness of this transition follows from the nature of the concepts and contents involved in the transition. (This was the thesis of chapters I and 2 of *The Realm of Reason* (Peacocke 2004).) If that property is not sufficient for entitlement, it is questionable whether anything ever is. But it is, on the present approach, that property that is present when one judges "I see that p" on the basis of seeing that p.

⁸ For extensive discussion of issues of regress and reflexivity, and historical references, see Caston (2002). I believe the Core Rule meets many of the desiderata that Caston formulates, and ought to be considered either as a possible interpretation of Aristotle or as a thesis doing justice to his best insights on this matter.

⁹ The point bears on the diagnosis of what is wrong with Moore's "Proof of an external world". If the arguments of the text above are sound, what is wrong with Moore's "Proof" cannot be that in making the perceptual judgment "Here is a hand", he is already relying on the proposition that he is perceiving. For further discussion, see Peacocke 2004: ch. 4, sect. I(c).

"Another I"

This is of course not to imply that we have here any kind of answer to skepticism about perception. The skeptic is questioning whether we are really ever entitled to take perception at face value. According to him, we never really get as far as the first line of the transition in the Core Rule. If we cannot know that p perceptually, we never see that p, since seeing that p implies knowing that p. But if we do get as far as that, there is nothing erroneous or unwarranted in making the transition to "I see that p".

The Core Rule, in its two parts, together with the Extended Core Rule, can be compared with a competing rule, one we might call an "Inside-In" Rule. This "Inside-In" Rule states, as a primitive rule, that when a thinker has a visual experience as of its being the case that p, he can judge "I have a visual experience as of p's being the case"; and similarly for other modalities. Why should we not use the Inside-In Rule in giving a possession condition for the concepts of experience and perception?

It can hardly be objected that the Inside-In Rule is incorrect. On the contrary, what it proposes as primitive is a consequence of the Core Rule plus the Extended Rule. If someone has a visual experience as of p's being the case, then he is in a state that is subjectively similar to the state he is in when he can apply the Core Rule. Hence, by the Extended Core Rule, he can judge "I have a visual experience as of p's being the case". The question is not, then, of whether the Inside-In Rule is correct, but of whether it is fundamental, or consequential. If the Inside-In Rule is fundamental, what is the rule for self-ascribing seeings? If seeings are fundamentally conceived of as visual experiences that additionally stand in the right kind of relation to environmental states, the question arises of what kind of grounds a thinker can have for thinking that he stands in the right kind of relation. What, for example, on this approach would give the thinker rational grounds for self-ascribing a seeing in quite ordinary circumstances in which he really does see? The obvious answer to this question is that we know that the visual experience stands in the right relations to be a seeing when indeed we see, so we can self-ascribe a seeing. This is quite right; but it evidently relies on the Core Rule, rather than on the Inside-In Rule together with additional materials. I conjecture that if we try to take as fundamental in the order of explanation of understanding a neutral notion of visual experience, whose content may or may not be correct, and try to build up to mastery of a

Christopher Peacocke

notion of genuine perception by additional conditions, without using the Core Rule, we will never reach our intended destination in a way that makes sense of ordinary self-ascription. We have to take the genuinely perceptual case as fundamental, both in the explanation of understanding and in the account of the nature of perceptual states themselves.

2. MODAL STATUS, GRASP OF MODAL STATUS, AND A CHALLENGE TO WITTGENSTEIN

We should distinguish the following two kinds of transition. First, there is a transition from

seeing that *p*

to the judgment

I see that *p*.

When someone makes a transition because it is of the displayed form, the fact that he is in the state of seeing that p is part of the explanation of his moving to the judgment that he sees that p. In a second kind of transition, the thinker moves from a judged content

p

where p is one made available by his seeing—perhaps because it contains a visual perceptual-demonstrative such as "that desk"—to the judgment

I see that *p*.

Transitions of this second kind are not metaphysically necessary. It is metaphysically possible that that desk (actually given in perception) is covered with papers and you do not see that it is covered with papers. In some other possible circumstances, that desk is covered with papers, and you are facing away from it, or you are not in the room at all, or your eyes are closed. Not only is this possible: on your ordinary understanding of the notions involved, you also have some appreciation that it is possible. This appreciation is reflected in—amongst other things—your assessment of the truth-values of counterfactuals. We accept as true the counterfactual "If you had not entered the study, you would not

232

have seen that that desk is covered with papers, but it would still have been covered with papers". The most we can say about transitions of the second kind is that in any context in which the premise is seen to be true, the conclusion will also be true (that is, the transition meets a condition that is a variant of Kaplan's notion of logical validity in the logic of demonstratives). But this is well known to fall short of metaphysical necessity.

The fact that transitions of the second kind are not metaphysically necessary raises a question about the concept of seeing. In what features of the concept is this possibility founded? Can the possession condition for the concept of seeing explain the possibility? What does the explanation show about other treatments of the concept of seeing?

If we consider just the thought-contents involved, the possibility of modal divergence between "That desk is covered with papers" and "I see that that desk is covered with papers" is quite unsurprising. "That desk is covered with papers" has a categorical truth condition, which is fulfilled with respect to some arbitrary possible state of affairs s provided that it holds, with respect to s, that that same desk has the categorical property of being covered with papers. For "I see that that desk is covered with papers", as thought by you, to hold with respect to s, it has to hold with respect to s that you stand in the same psychological relation to the content that desk is covered with papers as you do when in the actual world you see that that desk is covered with papers.¹⁰ It seems clear that nothing rules out that s meets the first condition without meeting the second. The fact that this can be the case is part of our conception of objectivity of the world we perceive, and of our conception of the mind-dependence of perception.

The reason why there is a special problem in accounting for these modal truths about the concept of perception is that the concept is a member of a family for which the possession condition is given by reference to a psychological state that makes application of the concept rational, a family for which there is thereby also what I call a "cantilevering" problem. The concept of seeing, other concepts of perception, concepts of sensation, and observational concepts of material objects and events are each concepts that, in very different

¹⁰ This is an application of the Modal Extension Principle developed in the treatment of necessity in Peacocke 1999: ch. 4, "Necessity".

Christopher Peacocke

ways, all have possession conditions that mention a psychological state that makes application of the concept reasonable. What makes a concept an observational one is that a certain perceptual experience gives reason for applying the concept to a perceptually given object. What makes a concept a concept of sensation is that certain sensations themselves give reason for self-ascribing the concept of sensation; and so forth. Now in all of these cases, the psychological states that give reason for applying the concepts give reason only for making a judgment that contains a particular favored kind of mode of presentation of the object to which the concept applies. The particular favored kind of mode of presentation, or way of thinking, in question is that of the first person for concepts of perception and sensation. The particular favored kind of mode of presentation in the case of observational concepts is that of perceptual modes of presentation of objects and events. But the concept in question applies to objects not given in the favored way. People other than oneself can see, perceive in other modalities, and have sensations. Objects other than those perceptually presented to the thinker can have observational properties. And in other possible states of affairs, there are determinate truths about the extension of all of these concepts, even though of course a rational response to a psychological state (or to its absence) in the application of one of these concepts, can in the nature of the case only be a rational response to an actual psychological state (or an actual absence). So in all of these cases, it is a task for a philosophical theory of concepts to explain how the concept applies beyond those cases that are given in the favored wavs. This is what I mean by the cantilevering problem.

I suggest that the cantilevering we need in the case of the concept of seeing to the modal cases is supplied by a piece of tacit knowledge that involves grasp of a sameness relation. Someone who possesses the concept of seeing tacitly knows that for it to be true with respect to a given possible situation that he sees that p, he must be in the same psychological state as he is in when he reaches the judgment "I see that p" by following the Core Rule in the actual world. He will not be in the same psychological state in a possible state of affairs in which, though that desk is covered with papers, he is not in the same room, and does not see the desk at all.

The presence of an element of tacit knowledge involving a sameness relation is not at all unique to the possession condition

for the concept of seeing, or to other concepts of perceptual states. It should be a familiar point that there is an additional element of tacit knowledge involving a sameness relation in the case of ordinary observational concepts applicable to physical objects and events. We know that something can be square without our seeing it as square, not only because we may not be suitably situated, but because objects too small or too large to be seen at all may also be square. We grasp these possibilities because we have the following piece of tacit knowledge involving an identity relation: that for something to be square is for it to have the same shape as we perceive things to have when we perceive them to be square.

The resulting position presents a challenge to a claim that seems to be present in the writings of the later Wittgenstein. On the account I am in the course of presenting, grasp of a sameness relation plays an indispensable, and intelligible, part in an account of grasping the concept of seeing. This is a property that, on my view, the concept of seeing shares at least with the concept of the past and observational concepts, and arguably with many other concepts.¹¹ Wittgenstein, by contrast, repeatedly said that in a range of cases, including conscious psychological states, grasp of identity of state is to be explained in terms of the truth of two predications of the state in question. Thus, "For that part of the grammar is quite clear to me: that is, that one will say that the stove has the same experience as I, if one says: it is in pain and I am in pain."12 What he seems to be saying is that identity of experience is to be explained in terms of identity of predications, rather than conversely. What is striking in the present case of the concept of seeing is the need to invoke grasp of an identity relation in an account of possession of the concept even for firstperson ascriptions, when we consider embeddings in modals and counterfactuals. When we consider the occurrence of first-person predications of seeing in counterfactuals and in other modals, our understanding of first-person predications of seeings is not free of any tacit use of the notion of sameness of conscious state. It is not an identity-free level of predication that could be used to explain thought about identity of conscious states. To suggest that our understanding of identity of state across actual and possible situations can be explained, on Wittgensteinian lines, in terms

¹¹ See Peacocke 1999: ch. 3, The Past, especially sect. 3.5.

¹² Wittgenstein 1958: §350; see also ibid. p. 190 on the concept of the past.

of identity of true predications in the actual and merely possible cases would simply leave unexplained our understanding of what it is for the first-person predication to be true in the merely possible state of affairs.

There is, I conjecture, a huge amount of constructive philosophical work to be done on the nature of our understanding of identity as it features in solutions to the cantilevering problem. I myself suspect that grasp of identity plays a role even in the most basic cases of the capacity for thought about particular material objects (a case in which the analogue of Wittgenstein's position is particularly problematic).

The additional element of tacit knowledge involved in grasping "I see that p" should affect our conception of how the relation of seeing is fixed as the reference of the concept of seeing. The relation of seeing (seeing-that), between a thinker and a content, is not the unique relation R that makes always truth-preserving in the actual world the transition from the thinker's

seeing that *p*

to the content (with our thinker as reference of "I")

I R that p.

Consider the unintended relation \mathbb{R}^* such that I \mathbb{R}^* that p iff p, where in fact in the actual world I see that p, and for which I $\mathbb{R}^* p$ holds in some other possible world just in case p holds in that world. This gives incorrect modal evaluations, as we just discussed. Such unintended relations are ruled out by the requirement that the relation of seeing is one that is the same psychological relation as holds between the subject and the content when the thinker is following the Core Rule. The deviant candidate \mathbb{R}^* does not meet this condition: it is really a property of contents, fixed in the actual world indeed by a psychological relation to a thinker, but it is not that psychological relation that determines the application of the property in other cases.

3. COMPARISON WITH EVANS'S ACCOUNT, AND WITH AN APPARENT ALTERNATIVE

Evans gives a different account of how a thinker can attribute a content to his perceptual experience. In *The Varieties of Reference*,

he writes: "He [the subject] goes through exactly the same procedure as he would go through if he were trying to make a judgment about how it is at this place now, but excluding any knowledge he has of an extraneous kind. (That is, he seeks to determine what he would judge if he did not have such extraneous information.)" (1982: 227). Evans's idea is that if the subject, using this procedure, determines that he would judge that p under these conditions, then he can ascribe the content p to his perceptual experience.

To explain what he means by "extraneous information", Evans mentions an example of Dummett's (1976: 95). If you see a pile of newspapers at the Smiths' front door, you may judge, "I see the Smiths forgot to cancel their newspapers". But, on Evans's approach, the content *the Smiths forgot to cancel their newspapers* is not to be counted as part of the content to be ascribed to your experience, because it is "extraneous". Also, if you know that your visual experience is an illusion, that knowledge is also extraneous information that is to be excluded in assessing what you would judge when you apply Evans's criterion (1982: 228 n. 39).

It seems to me that the condition that Evans formulates is not necessary for an experience to have a given content. Something can be in the content of a given experience without the subject being willing to make the corresponding judgment Evans mentions. Several different kinds of example show this.

Consider recognitional concepts of individuals. A person can have the capacity to recognize the former Iraqi dictator Saddam Hussein. When he sees Saddam, his visual experience has a content specified in part by using that recognitional concept: it seems to him that Saddam, so thought of, is in front of him. Other things being equal, he will take such visual experiences at face value, and judge that Saddam is in front of him. But his willingness to do so rests, and rationally rests, on his belief that there is only one person, at least in this part of the world, who looks that way. This seemingly extraneous belief is essential for our subject to be willing to move from the experience to the judgment that Saddam is in front of him. When our subject comes to learn that Saddam actually employs three look-alikes, he will not move from the experience to the judgment that Saddam is in front of him. But his visual experience will continue to have that content all the same. So it seems that Evans's condition is not necessary.

Christopher Peacocke

Perhaps Evans could add the requirement that the judgment he mentions can rely on information if that information is necessary if the subject is to be willing to employ the concept in judgments at all. Whether or not that saves the Saddam example, it would not help with others.

Suppose you hear the sounds "Peter leaped". It is in the representational content of your experience that someone said that Peter leaped (its sense, not merely the sound and phonemes). But, we can suppose, it is only because you take yourself to be amongst English speakers that you also judge that the speaker said that Peter leaped. If you took yourself to be amongst German speakers, you would judge not that someone has said that Peter had made a certain kind of jump, but rather that someone had said that Peter is in love ("Pieter liebt"). So Evans's procedure fails to attribute to the experience the content that someone said that Peter leaped.

This is not a problem for the Core Rule, for in such examples a person can certainly hear that someone has said that Peter leaped, and move from that to a self-ascription of such a hearing. He can do this independently of whether he needs additional information before endorsing the content of the experience in a judgment.

There is a range of other examples of a similar sort. If you can see something as a car, a computer, or a phone, it is only because of your background knowledge of the function of these perceptually recognizable objects that you judge that the seen things are cars, or phones, and so forth. If this background knowledge were not there, some of us would not make the judgment that it is a car, or phone, or computer that is in front of us.

A hard line with these examples would be to take the unintuitive view that you do not really see the object as a computer, say, but only as something of a certain size and shape. But not all examples can be handled by such a hard line. The example of "Peter leaped" cannot. It would be a huge misrepresentation of our auditory experience to say that we do not really hear words as having certain senses.

There is a third kind of case demonstrating the non-necessity of Evans's conditions. In cases of this third kind, the content of the experience is so outlandish that we would never judge it to hold, given our background knowledge. A competent magician can make it look as if three pigeons have just come out of his jacket sleeve. We do not judge that they were there. If it is said that we must exclude knowledge of how physical objects behave, or what sorts "Another I"

are around us, we will thereby exclude all sorts of features of our visual experience. We see an occluded object as having a certain shape, as continuing in a certain way behind the occluding object; and our willingness to take these experiences at face value relies on our background information. Another example is provided by such experiences as that of the rising, but apparently curving, rising zigzag jet of water in the display that is, or used to be (c. 1983–4), in the Exploratorium in San Francisco. It had the shape shown in fig. 7.1. The effect was produced by a rising jet of water that was in fact continuously moving back and forth across the arc of a circle, but under carefully timed, and unnoticed, stroboscopic lighting that produced the visual effect of an unsupported continuous jet of water in a zigzag shape. The experience of this striking display was undoubtedly of a curving zigzagging jet of water. We are going to get the right answer from Evans's procedure only if we ask such

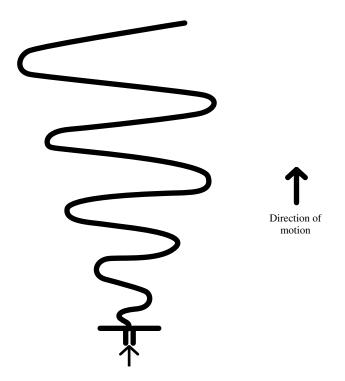


Figure 7.1. The seemingly zigzag jet of water.

questions as "What would I judge if I did not think that the laws of motion did not hold?". It is impossible to believe that such barely assessable questions have to be answered before we can pronounce on the question of the content of our perceptual experience in looking at such a display.¹³

I conclude from this range of cases that it is one thing for a judgment to have a certain content in the circumstances described by Evans, and it is another for the experience to have the same content, even though there is sometimes overlap between the two. There would be complete coincidence if "extraneous information" meant "any content that is not in the content of the perceptual experience": but that would be a very different procedure and criterion from that which Evans suggested. That different procedure would not be genuinely circumstance-dependent, in the way Evans's procedure is. For example, if one just requires someone to judge only what is in the content of the perceptual experience, one would not need Evans's instruction that if one knows one's experience is illusory, one should prescind from that information. Just requiring sensitivity to the content of the experience would be enough, whether it is a genuine perception or not. Neither would the alternative approach preserve the primacy of the genuinely perceptual case in the self-ascription of experiential content, an attractive and important feature of Evans's approach.

Evans's approach is not the only treatment to preserve the primacy of the genuinely perceptual case. The possession condition that incorporates the Core Rule and the Extended Core Rule also attributes explanatory primacy to the genuinely perceptual case. It does so in two respects. The genuinely perceptual states of seeing-that, feeling-that, hearing-that, and the rest are the initial states from which transitions are made in the Core Rule when the thinker makes a self-ascription of an experience. The treatment of thought about illusions as states that are subjectively similar to genuine perceptions also gives an explanatory primacy to the genuinely perceptual case.

¹³ Could Evans solve this problem by appealing to multiple similarity relations, just I myself did in an earlier section? One needs to capture that the jet is going around curves, for this is given in the experience itself. That it is so perceived is not captured by similarities to other genuine seeings overlapping in contents.

Why did Evans adopt his account of the self-ascription of experience, with its circumstance-dependent character? Despite the differences he notes between the two, he was likely partly tempted by the partial parallel with his treatment of the selfascription of belief (1982: 225-6). In the case of belief, he suggests, a thinker can employ the procedure of asking himself a question about the world, whether p holds, and then self-ascribing the belief that p just in case his answer is affirmative. In both the case of belief and the case of perception, the self-ascription of a psychological state is said to be dependent upon what, in certain circumstances, one thinks about the world. But beyond this parallel between the two cases on his own view of them. I conjecture that there was a further reason for Evans's (arguably uncritical) adoption of his circumstance-dependent account of the self-ascription of experience. The reason lies in his account of experience itself. In a passage that has achieved some notoriety. Evans wrote in The Varieties of *Reference* that

we arrive at conscious perceptual experience when sensory input... 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 on the informational properties of the input. When there is such a further link, we can say that the person, rather than just some part of his brain, receives and possesses the information. (1982: 158)

On this conception of experience, what makes something the content of an *experience* must be its relations to thought and concept application. On such an approach, what is the most obvious resource to invoke in specifying how a state must be related to "the conceptapplying system" if it is to be an experience with a given representational content? The most obvious resource is to say the content of the experience is determined by the content of the judgment that the thinker would make when enjoying the experience and when prescinding from other extraneous considerations. As we have seen, this does not work. But it has to be some resource of this sort, involving relations to states with conceptual content, if, like Evans, one does not think that there is a notion of the representational content of experience available independently (in respect of philosophical explanation) of relations to states with conceptual content.

Christopher Peacocke

In my judgment, the idea that the very same state that can occur in a creature that does not possess concepts and which is not an experience can nevertheless be an experience in a genuinely concept-using creature has no legitimate attractions. Far from supporting the thesis that the same perceptual experiences can occur in more primitive animals, the idea is incompatible with that thesis. For what it is worth, I suspect that Evans would have dropped the idea had he lived longer and been able to revise The Varieties of Reference. The most primitive conceptual contents of thought are individuated by their relations to a system of perceptual states whose content is autonomous in relation to the level of conceptual content. Those perceptual states must indeed make a contribution to their subject's conception of the layout of the world around him. But all this can happen at a level of non-conceptual content that is far more primitive than the level of conceptual content individuated in terms of rationality and judgments.14

4. THE POSSESSION CONDITION, EMPIRICAL PHENOMENA OF DEVELOPMENT, AND OTHER-ASCRIPTIONS

What should be the relation between the possession condition for a concept and empirical psychological phenomena involving possession of that concept? The relation between the two is complex and multifaceted. Here I want to emphasize one of the tasks of a theory of possession conditions that is particularly pertinent to issues surrounding possession of the concept of perception. (Some of the other tasks I have discussed in Peacocke 2001*b*.)

A statement of a possession condition for a concept is responsible in the first instance to the epistemic phenomena involving possession of that concept. These phenomena involve the rationality or irrationality, in given circumstances, of judging certain contents containing that concept. The fact that it can be rational,

¹⁴ For further discussion, see Peacocke 2001*a*. I have not always held the views just stated in the text. For further discussion of the issues, see Peacocke (2003). The whole of the above discussion could also be adjusted to take account of the distinction between conceptual and nonconceptual content, without fundamental change to the essential elements of the account.

and correct, to apply an observational concept to an object even when the object is not perceived must be explained by the possession condition for the observational concept. The fact that we can rationally come to accept new axioms for some logical or mathematical concept, axioms that are not implied by what we previously accepted, also has to be explained by the possession conditions for the logical or mathematical concepts in question. If the identity of a concept is answerable to Frege's informativeness condition, and a possession condition individuates a concept, then these tasks of a theory of possession conditions are demanded simply by the nature of the subject matter of a theory of possession conditions.

These explanatory tasks are philosophical, and have an a priori character. The rationality or otherwise of judging something in given circumstances is an a priori matter. So these tasks have the characteristic epistemic status of much of philosophy that aims to be explanatory. The task is to explain a set of a priori truths—truths about what is informative in given circumstances, truths about what contents involving a given concept it is rational to judge in those circumstances—from more fundamental principles that individuate the concept in question.

Some of the phenomena displayed by possession of a given concept by actual human thinkers are, however, empirical phenomena that could not be excogitated simply from the a priori nature of the concept. If these phenomena are special to the concept in question, the possession condition for the concept should contribute to an explanation of how these phenomena are possible. One way in which this task can be implemented is illustrated by a treatment I will offer of some empirical phenomena involving possession of the concept of perception.

All of the following phenomena are displayed by children employing the concept of perception, and are well attested by psychological research. Some of these phenomena will be familiar to any parent.

(i) Toddlers between the ages of 24 and 30 months do not appreciate that they can see something that someone on the other side of an opaque screen cannot see.¹⁵ Asked to hide a toy from

¹⁵ For an engaging overview, see Gopnik, Meltzoff, and Kuhl (1999).

another person, who is on the other side of the screen, a child of this age will often put the toy in a position in which the child himself cannot see it, on the other side of the screen where the other person can see it. These are what Hughes has called "projective" errors, and Flavell calls "Level 1" errors.¹⁶

(ii) In playing hide-and-seek, a child of this age will be willing to hide under a table, in a location in which it is evident to any adult that he can be seen in the room, even though he himself cannot see the rest of the room. We can call this phenomenon "incompetent hide-and-seek". Incompetent hide-and-seek is plausibly an instance of the same inability displayed in projective errors. It is the special case in which there is failure to grasp the conditions under which the seeker sees something.

(iii) Somewhat older children, who do not make these errors, nonetheless make a different error. In a situation in which one of these older children sees an object, and appreciates that someone else also sees the same object, they nevertheless fail to appreciate that the other person will see a different side of the object than they themselves see—even though they know that the object is different on its two sides.¹⁷ Hughes calls these "perspective errors", and Flavell calls them "Level 2" errors.

(iv) Ordinary 3–4-year-olds are ignorant of the sources of their own and others' knowledge in the following respect. Suppose one child sees a second child look into a box. The first child denies that the second child knows what is in the box, even though he himself (a) looked in the box, (b) came thereby to know what was in the box, and (c) knows that he knew what was in the box. The explanation for the possibility of this state of affairs is that the first child does not know how he himself came to know what is in the box, even though he does have the concept of seeing. We can call this phenomenon "Ignorance of Sources".¹⁸

These phenomena (i)–(iv) are all empirical phenomena involving the concept of seeing, and we should be able to explain how they are possible by drawing on the possession condition for the concept of seeing. They can all be explained by drawing on the Core Rule,

¹⁶ See Masangkay *et al.* (1974) and Yaniv and Shatz (1988).

¹⁷ See Masangkay *et al.* (1974).

¹⁸ See Gopnik *et al.* 1999: 46–7 and the references therein; and especially Wimmer, Hogrefe, and Sodian (1988).

"Another I"

in both its positive and negative parts. I propose what I call "the Same Rule Hypothesis":

The child, in attributing seeings to others, applies the same Core Rule to others as he does in self-ascribing experiences, but does so taking as input to the Rule not another's seeing-that p, but his own.

That is, the child moves from his own

seeing that *p*

to

the other person sees that *p*.

A natural extension and partner of the Same Rule Hypothesis is that the child uses the same procedure in judging that the other person does not see as she uses in judging that she herself does not see.

The Same Rule Hypothesis is an example, as promised earlier, of one of the other ways in which one's grasp of a general property can be essentially first-personal without being of the sort Wittgenstein criticized. The thinker has some understanding of what it is for another to have some property because he knows that it can be attributed by applying the same rule for others as he employs in making self-attributions.

The Same Rule Hypothesis explains the "projective" errors in (i). If a child uses these procedures, he will judge that another person sees something, or sees something to be the case, precisely when he himself sees it, or sees something to be the case. Equally, if the child uses the negative part of the Core Rule in the same way, he will judge that the other does not see something in exactly the same conditions as he does not see something.

The Same Rule Hypothesis also explains the phenomenon of incompetent hide-and-seek. If our child judges, using the Core Rule, that he does not see anyone in the room, then if he uses the same rule in the way indicated to make judgments about the visual experience of the other player in the game, he will judge that the seeker equally does not see anyone in the room.

Perspectival errors can be explained using the same resource. Suppose the child sees one side of an object, and applies the Core Rule to judge that he himself sees that side. Applying the same rule in other-ascription would then yield the result that he ascribes to the other a view of that same side.

Christopher Peacocke

On this approach, increasing knowledge about the conditions under which others see, and what features of an object they see, is attained by the child's coming to qualify the conditions under which the Core Rule can be applied in other-ascription in this naïve way. The Core Rule works in other-ascription of seeings only for someone in roughly the same location as oneself, in the same conditions, with unobstructed sight. Eventually, for full knowledge, the child must come to appreciate that for a third-person ascription of seeing to be correct, the other person must be in the same state in relation to the world as he, the child, is when he sees. This is the end-state that has to be reached as the naïve applications of the Core Rule in other-ascriptions are progressively qualified by conditions on the other's relations to the objects and states of affairs perceived. The very qualifications on the use of the Core Rule in other-ascription also give a special role to the first person. The conditions under which the subject himself does not see are used in formulating the conditions under which the Core Rule is to be qualified in its use in ascriptions to others.

It may be objected that the empirical phenomena (i)–(iii) could equally be explained by the hypothesis that the child uses the Core Rule in self-ascribing experiences, and then infers to the occurrence of experiences in others by using the principle that others see something to be the case if and only if he himself does. I suggest, however, that at these early stages the child has no conception of what it is for another person to see something to be the case other than that such other-ascriptions can be reached by applying the Core Rule to others. If that is the child's conception of other-ascriptions and their correctness conditions, it will indeed be a consequence of the procedures for self- and other-ascription of experiences that he sees something to be the case if and only if another person does. But the child does not have an independent conception of perception for which this coincidence is believed to hold.

The empirical phenomenon of Ignorance of Sources would not be possible if to have the concept of seeing-that is to think of it as a kind of knowledge that has visual perception as its source. That way of thinking would make Ignorance of Sources for the case of knowing-that into some kind of conceptual impossibility. But it is not impossible under the Core Rule, for two reasons. First, to say that someone is following a rule does not imply that they can conceptualize the rule they are following, let alone that they know that they are following that rule. In fact, if someone could follow a rule only if he could conceptualize what rule it is that he is following, the Core Rule could not possibly be part of an account of what it is to have the concept of seeing, since under that condition the thinker would have to conceptualize the input conditions, which would require him already to have the concept of seeing. But that condition is much too strong on independent grounds. It would, for instance, rule out ever including introduction and elimination rules in the account of what it is to possess a logical concept, since logical concepts would have to be used in characterizing what the rules are.

The second reason why the Core Rule permits Ignorance of Sources is this. While seeing-that is indeed a kind of knowledge, the Core Rule does not imply that someone who thinks of his own seeings-that has to think of them as a kind of knowledge. The Core Rule requires someone who has the concept of seeing-that and applies it to himself to be sensitive to the boundary between those of his states that are seeings-that and those that are not. It does not require him to think of those states as states of knowledge. Doing that would require him to take a further step, to connect up his concept of seeing-that with his conception of knowledge. These two philosophical points bring the philosophical conception advocated here into very close alignment with the empirical explanation of Ignorance of Sources given by Wimmer, Hogrefe, and Sodian (1988). The Core Rule is a philosophical, constitutive account that makes intelligible their empirical explanations.

The fact that the Core Rule also provides a means for ascribing perceptions to other people is of more general philosophical significance than making some sense of the empirical data of acquisition. The fact that one can use the Core Rule in other-ascription shows how a possession condition for a concept, whilst being essentially first-personal, can nevertheless contain the seeds of a procedure for other-ascription. This is one very clear way in which a concept can be shown to be unambiguous as between first- and third-person applications, whilst still displaying an explanatory primacy for the first-person case in the account of possession. The possession condition for the concept of seeing that I have offered entails that one could not be capable of self-ascriptions of seeings without both having the resources to grasp and having the materials for a procedure for making other-ascriptions of seeing.

If someone other-ascribes in accordance with the Same Rule Hypothesis, he is taking one step to seeing the other person as "another I". In applying the same rule in attributing seeings to others as he applies to himself, he thinks of the other's seeings as states of the same kind that he himself enjoys.

Other-ascribing in accordance with the Same Rule Hypothesis does not, however, take a thinker the whole distance to thinking of the other person as "another I". Traveling the whole distance also involves thinking of the other as capable of self-ascribing too—that is, as capable of moving from the states ascribed in accordance with the Same Rule Hypothesis to self-ascriptions, in accordance with the Core Rule itself. Here the subject has to think of the other person as employing the first-person way of thinking. That is, the subject has to refer to, and not merely employ, the first-person way of thinking.¹⁹

The ways of coming to make other-ascriptions for which the Core Rule provides the starting point dovetail very smoothly with simulationist approaches to the procedures by which we make psychological attributions to others. In the case in which one attributes a perceptual state to someone standing right next to one, one may not even need to imagine being in some other state before being in a position to make an ascription to him of a seeing. Slight difference of angle, and individual differences aside, what you see to be the case is what he sees to be the case. When you are considering someone situated in space or time differently from you, you imagine what it would be like to be there. In imagining this, you are in fact imagining having certain perceptions, but you do not need to have the concept of perception in order to engage in such imagining. Making a transition from these imaginings-"there's a tree over there, and a gate to the left"—and applying the Core Rule in the imagined situation, you move to a self-ascription in the imagined situation, or even directly to an other-ascription, using the Core Rule as applied to another-"another I" again. Again, we prescind from individual differences in perceptual mechanisms, something whose intelligibility is made possible by the use of the relation of identity of psychological state, as

¹⁹ See Peacocke (1981).

"Another I"

discussed in section 3 above. In such ascriptions of seeings to others, one makes use of the Core Rule in ascriptions of seeings to others; and this is to be distinguished from formulating the rule, which is what one would be required to do at some level or other if one were ascribing by means of a theory-theory approach.²⁰

The Same Rule Hypothesis was put forward as an account that could explain children's developing understanding of another person's perception. But applying the Same Rule Hypothesis in other-ascription, however tempered with qualifications about same conditions, or perspective, can never capture our full, mature understanding of what it is for another person to be seeing. We understand the hypothesis of the inverted spectrum: that what I see as red, you see as green. We understand this hypothesis even in the case in which your situation and perspective on an object are precisely those in which I see it as red. Applying the Same Rule Hypothesis in other-ascription of experience builds in an implicit presumption of sameness of experience in two people in the same given external conditions and relations. No doubt in prephilosophical thought we rely on the Same Rule Hypothesis, which is why it is a surprise when the hypothesis of the inverted spectrum occurs to us, or is suggested to us. But we certainly understand it. A formulation of a possession condition for the concept of seeing that would make it unintelligible would be erroneous.

Actually there would already be an instability in a variant of the concept of seeing that treats first-person ascriptions in accordance with the Core Rule, and the surrounding conditions we noted for counterfactuals, and offered only the Core Rule applied in the third person for other-ascription. For an understanding of counterfactuals about one's own experience involves a grasp of a sameness relation for experiences. If one has some grasp of that sameness relation, it is, to say the least, not clear why the condition for another to be seeing that p should not simply be that the other is in the same state as one is in oneself when one sees that p, where one's understanding of the latter involves the Core Rule. In fact, if this identity condition were not met for correct other-ascriptions of seeing, then it would not after all be the same property that is

²⁰ This does not conflict with the point that the theory-theory may give the correct account of what it is to be in a given psychological state. Here we are just talking about the procedures we actually employ in other-ascription.

Christopher Peacocke

attributed in first-person and third-person ascriptions. And indeed it would not be the same property if the full account of our understanding of the third person were given by the Core Rule applied in the third person, with surrounding qualifications, since the property predicated in other-ascription would be one for which spectrum inversion is a priori impossible.

We should conclude that, contrary to Wittgenstein, grasp of a sameness relation plays a crucial part in the understanding of third-person ascriptions of seeings. I suspect that grasp of identity, applied to a given category of item, is not to be reduced to grasp of something else. The fact that the Core Rule applied in the third person cannot capture the full extent of our understanding is itself just one plank in support of the case for such irreducibility. Saving there is irreducibility is, however, consistent with saving much more about what such grasp involves. Further development of this position would need to supply this further elucidation. At the very least, the further account would have to say how the thinker latches onto the property itself, whose application in the third person is in question. A thinker's grasp of the Core Rule says how his own case provides a means of doing so. It is uniquely the property of seeing that p, nothing weaker or stronger, that is the basic one that the thinker must have for use of the Core Rule to result in a true self-predication. The claim that a thinker's understanding involves some grasp of an identity relation is also one answerable to what it can explain about his judgments and other actions. This explanatory power can be present without any reduction of identity to something else. All these telegraphic remarks would have to be expanded in a full treatment of this position.

5. THE MODEL GENERALIZED: AND ACTION AS AN INSTANCE

When the Core Rule is embedded in a possession condition for the concept of perception, the result is an instance of a general form of account. In the general account, the thinker is in some intentional state S with the content p. This state is one with representational content: in being in the state, it seems to the thinker that p holds of the world. The thinker then makes a transition from his

S-ing that p

to the self-ascription

I S that p.

We can call this general schema the "Outside-In" model. There are two variants of the general Outside-In model, according as the state *S* is factive or not. Are there any other instances of the Outside-In model, of either variant?

I suggest that certain concepts of action provide another instance of the Outside-In model, even in its stronger, factive variant. Suppose the thinker makes a self-ascription of the form

I am N-ing.

Instances of this will be "I am walking", "I am typing", "I am moving from the waiting room to the exit", "I am working out the sum of this column of numbers". These instances are not all ones that in themselves imply that the subject's *N*-ing is an action. You might be moving from the waiting room to the exit on a moving walkway, onto which you had stepped unintentionally. I do, however, want to suppose, as part of the specification of the range of cases about which I am talking, that the subject has the kind of distinctive awareness of *N*-ing that is made available by its in fact being an action on this particular occasion.

This kind of awareness can be present even when one is not perceiving that one is N-ing. One can be aware that one is raising one's arm, even when one's afferent nerves are severed, and there is no proprioceptive feedback, and one is turning one's head away from one's arm so that one cannot see it either. This distinctive phenomenology of action is what makes possible illusions that one has raised one's arm even when, unknown to oneself after some terrible accident, one has no arm. The phenomenology of action involves states with representational content.

A natural first suggestion would then be that another instance of the Outside-In model is one in which the thinker moves from 'I am N-ing' to

I am N-ing intentionally.

This, however, is too strong. We need to treat the cases of basic and non-basic actions differently. I may intentionally transfer one-third of my assets to my son. But there is no distinctive action-awareness of transferring one-third of my assets to my son. The actionawareness in such a case is action-awareness of moving my hand, and (say) of signing my name. These actions may be intentionally transferring one-third of my assets; but that does not give me actionawareness of conditions involving proportions of my assets.

Where *N*-ing is an action (type) that is basic for the subject, we have the Core Rule for the case of basic actions. From the subject's action-awareness

I am N-ing

for a basic-action type, the subject can rationally move to the judgment

I am *N*-ing intentionally.

There is a Core Rule for ascriptions of intentionality in the case of non-basic actions too. Suppose that P-ing is a non-basic action type for the agent. Suppose too that there is some basic action type N such that

- (i) the subject has an action-awareness of his N-ing, and
- (ii) in N-ing, the subject is P-ing, and
- (iii) the subject means to be P-ing as part of his plan in N-ing.

From the action-awareness in (i), and the conditions (ii) and (iii), the subject may make a rational transition to the self-ascription

I am P-ing intentionally.

(Here, known but unintended consequences must, for the purposes of (iii), not be understood as part of the agent's plan.)

The ability to follow the Core Rule in the case of non-basic action involves a sensitivity, on the part of the thinker, to the nature of his own plans. In this it takes a significant step beyond what is involved in following the Core Rule for the case of basic actions. The capacity to follow the Core Rule for non-basic actions thereby represents an intermediate state, located between that of having no sensitivity in one's judgments to one's own plans and decisions, at one extreme, and having full conceptualization of one's own plans, decisions, and intentions at the other extreme.

The Core Rules for the action case involve transitions a thinker is entitled to make. When these transitions are made from

action-based awareness that he is *N*-ing, these transitions will be truth-preserving. They will also be capable of yielding knowledge that he is *N*-ing intentionally when the further conditions (ii) and (iii), for the non-basic case, are also known.

One can also formulate an Extended Core Rule, to give a procedure for self-ascribing tryings. The Extended Core Rule states that a subject can self-ascribe a trying to N if it is for him subjectively, at least in respect of action-type awareness, as if he were in the state which is the input for the Core Rule for N-ing intentionally. The triggering condition for this Extended Core Rule can be met when the thinker is trying to N but is unsuccessful.

The explanatory attractions of the Outside-In model applied in the action case parallel some of those that are present for the Core Rule in the case of the concept of seeing. You do not have perceptions of your action-awarenesses-they do not become an object for you, any more than your seeings do. There is too a respect in which only someone who knows what it is like to act can have the concept of action. Action-awareness has the phenomenology of success. Perceptual experience can make rational judgment overrule this impression of success; but it is overruling that is required. Even in the case of actions that one can perceive that one is performing, there is a way of thinking of them made available by action-awareness that does not require one to perceive them. One can see oneself clenching one's own fist. But one can still think "this clenching", and refer thereby to one's action, provided that one tries and succeeds, even in a case in which one's arm, wrist, and hand are fully anaesthetized. One also does not perceive one's own mental actions, but can think of them demonstratively.

To all of this it may be objected that there is a fundamental asymmetry between the perception and action cases. There cannot be a case of seeing-that p without the subject being in a subjective, conscious state that can give him reason to self-ascribe a seeing that p. The phenomenology is intrinsic to the very nature of the state in question, and this makes possible the Core Rule. But, the objection runs, no such point holds for tryings. There can be events which are mental initiations of actions, caused appropriately by the subject's mental states, and which do not have the distinctive phenomenology of ordinary actions. Awareness, so the objection runs, is here an add-on, in the way it is not a mere add-on for the state of seeing something to be the case.

Christopher Peacocke

I reply that we must remember the informational states operative for blindsight subjects, the states which make them "guess" correctly when asked what shape that is in front of them, even though they have no visual experience as of anything in front of them. A sensitivity to such states should not be included in the possession condition for the concept of seeing! Similarly, I suggest, our conception of a trying is not merely something with a certain role in the production of bodily movements or mental states, stemming from the subject's own mental states. As I have repeatedly emphasized, there is a distinctive phenomenology of action that is produced by tryings. Events that fail to produce this phenomenology are not tryings, just as non-conscious informational states about one's environment are not seeings. Though this is not the place for a development of the case, this point does not seem to me to be a merely stipulative one, of no fundamental philosophical significance. Seeings, as conscious states, can have a rational bearing on the thought and action of a subject. They could not function as reasons without this conscious character. Similarly, the occurrence of tryings, as crucial in the phenomenology of action, is essential to the subject's non-inferential conception of some events in the world as those that he is controlling. Our conception of ourselves as agents would be very different, and would have a theoretical rather than an immediate character, if tryings lacked this conscious dimension. The phenomenology of action that they sustain also of course gives us reasons for making judgments about which actions we are performing in the normal, non-pathological cases. I therefore contend that the objector is mistaken in thinking that seeing and acting differ in respect of the importance of phenomenology to their individuation.

In the case of the concept of seeing, we identified a practice of using the same rule in other-ascriptions of seeing as is used in self-ascriptions. There is something analogous in the case of basic actions. You can perceive the movement of someone else as being of a kind that you yourself can perform. This is not a matter of personal-level inference, but is rather part of the content of your experience of the other person's movement. When you see someone waving in a certain way to hail a cab, you see his action as of a kind that you could perform. If asked to wave in the way he waves, you could do the same, without any inference or calculation. The famous "mirror" neurons identified by Rizzolati "Another I"

and his co-workers are likely to be involved in the possession of such capacities to act and to perceive. Such underlying representations are also the sort of thing required for the explanation of the ability, even of newborns, to imitate such gestures as sticking out one's tongue.²¹

Suppose a subject sees someone else as performing an action of kind φ , that is basic for the subject himself, and suppose too that this perception is of a sort that involves the subject experiencing the other's action as of a kind that he could perform himself. In such a case, our subject can move from the third-person content

That person is φ -ing

to the conclusion

That person is φ -ing intentionally

and also to the conclusion

That person is trying to φ .

In making these transitions, our subject applies a version of rule in the third-person case as he applies in self-ascription: that is, the same rule as he applies in the first-person case. The only differences between the first-person and third-person cases are that the awareness of action is as of another's action, and that "that person" is substituted for "I". Again, our subject thinks of the other person as "another I". The ability to see another's actions as ones of a sort one can perform oneself supplies the cantilevering from the case described in the Core Rules to the case of other-ascription.

The scope for errors in ascriptions of intentionality for basic actions is in certain respects narrower than the scope for errors in ascriptions of seeings. What a third person sees to be so depends on several of his relations to the state of affairs in question, particularly his spatial relations. But in the case of action, nothing plays quite the role of these spatial relations. There are no spatially wholly separate objects and events distinct from a subject and his own states and events in which the subject must stand in certain spatial relations for an event to be one of the subject's actions. Correspondingly, there is no scope for analogues of mistakes about such relations.

²¹ See Meltzoff and Moore (1977); Meltzoff (2002).

Christopher Peacocke

Again, as in the case of perception, our full understanding of intentionality in other-ascription is not exhausted by the use of the same rule. We know that an expert in yoga, in moving into exotic bodily configurations, is acting intentionally. We even see him as performing actions, even though we neither see nor think that he is doing something we could do ourselves. We do, however, think of those exotic movements and positions of his as being similar in certain respects to things we can do ourselves. We think of them as produced in the same sort of way as things we can do ourselves. Once again, it is grasp of a sameness relation that extends understanding beyond what can be reached by use of the same rule in self- and other-ascription. With grasp of this sameness relation, we can also make sense of the idea that organisms that are totally unlike humans also act intentionally. But we build out to our understanding of that case from the case in which we ourselves are intentional agents.

On this approach, the concepts of acting intentionally and of trying are first-personal in the deeper sense I tried to articulate in the case of seeing. A philosophical account of one's general understanding of what it is for an arbitrary person to be acting intentionally or trying makes reference to what is involved in making first-person ascriptions of these properties, so thought of.

6. CONCLUDING REMARKS: WIDER ISSUES

I conclude by remarking on two features present in this treatment of self- and other-ascription of certain psychological properties that are of wider application and interest.

The Core Rule in both the perception and the action cases is truth-preserving; it is so a priori; it is so as a result of the nature of the states and concepts involved in the respective rules. Other things being equal, it is adaptive to follow truth-preserving rules. It also, if the present approach is correct, comes with the very possession of the concept of perception that one follows a truthpreserving rule. So in acquiring the concept of perception, one has not only the ability to discriminate in thought between those situations in which someone is perceiving something to be the case and those in which he is not; one also has an ability to apply this distinction correctly. These points suggest a general account of the relations between grasp of transitions that are a priori and adaptive advantage. Some special form of truth-preservation comes with possession of the concept, and brings adaptiveness in its train.

The other feature of this treatment involves a connection between the external individuation of mental states, on the one hand, and epistemological relations, on the other. Both perceptual experiences and tryings are plausibly externally individuated. What gives them the content they have is constitutively dependent upon certain of their causes, in certain circumstances, in the case of perceptual experiences, and upon certain of their effects, in certain circumstances, in the case of tryings. There is, unsurprisingly, a connection between external individuation of a mental state and what enjoyment of that state entitles one to judge. At the first order, perceptual experiences and action-awarenesses entitle one to make judgments about the external world and about what one is doing, respectively. But if the present approach is correct, external individuation also bears upon the entitlement to make second-order judgments, about one's own mental states-about whether one is perceiving and whether one is acting. This follows immediately if the externally individuated mental states mentioned in the input to the Core Rules provide an entitlement to judge the contents that are the output of the Core Rules. The Core Rules show how the occurrence of mental states that are externally individuated can lead to knowledge of those very mental states that are externally individuated, and can do so in rational ways.

Space and Objective Experience

QUASSIM CASSAM

Ι

In his paper 'Things without the mind', Evans examines what he describes as 'the Kantian thesis that space is a necessary condition for objective experience' (1980a: 250). He reads the second chapter of Strawson's Individuals (1959) as a defence of a weakened version of this thesis, but maintains that Strawson's main line of argument is unsuccessful. This leads Evans to put forward his own arguments for the Kantian thesis. These arguments assume that this thesis is primarily concerned with the 'connection between the idea of an objective world and the idea of a spatial world' (1980a: 249). More specifically, they assume that what is at issue in the claim that space is necessary for objective experience is whether someone who has the idea of an objective world must 'thereby conceive of a system of spatial relations in which both he and the phenomena he experiences have a place' (ibid.). In this context, the idea of an objective world is the idea of a world that can be perceived and exist unperceived. Accordingly, Evans insists that 'to defend the Kantian thesis, the idea of space must be shown to be implicitly involved in the very idea of existence unperceived' (1980a: 261).

Evans's Kantian thesis is one of a range of Kantian theses about the necessary conditions of objective experience, although it is the only one discussed in 'Things without the mind'. I will refer to necessary conditions of objective experience as *epistemic conditions*, and to the thesis that space is an epistemic condition as the *Spatiality Thesis* (ST).¹ In general, claims about what is necessary for objective experience raise a number of basic questions. One is whether the various conditions that they identify are genuinely necessary conditions. I will call this the *question of necessity*. Another question concerns the ultimate basis or foundation of what is claimed to be necessary for objective experience. To enquire about the basis or foundation of a condition of objective experience is to ask why that condition obtains, or what makes it the case that it obtains. This might be described as the *question of foundations*. Finally, there is a question about the proper methodology or procedure for discovering or establishing necessary conditions of objective experience. I will refer to this as the *question of method*. These three questions are obviously related, and I will have more to say about the relations between them in due course.

It does not seem to me that Evans is successful in his attempts to establish that in order to have the idea of an objective world one must also have the idea of a spatial world. I also have doubts about his suggestion that the Kantian thesis cannot be defended without showing that the idea of space is implicitly involved in the very idea of existence unperceived. A different and better approach is suggested by Kant in the first *Critique*. According to Kant, it is the *perception* of space rather than the *idea* of space that should be seen as necessary for objective experience. I am sympathetic to this version of the Kantian thesis, and will attempt to bring out the respects in which it fares better than Evans's account of the link between space and objectivity. On the other hand, I reject Kant's idealist explanation of this link. The explanation that I favour, and which I will outline in the concluding section of this chapter, does not depend on Kant's idealism.

Π

According to Evans, objective experience requires the conception of a world whose existence and operations are independent of our

¹ The expression 'epistemic condition' is taken from Allison (1983). For Allison, an epistemic condition is one that is 'necessary for the representation of an object or an objective state of affairs' (1983: 10). Allison also assumes that such conditions 'must reflect the cognitive structure of the mind (its manner of representing) rather than the nature of the object as it is in itself' (1983: 27). I do not accept this assumption.

experience of it. To have objective experience, one must think of one's experience as experience of a world which is, in this sense, 'objective', and so must have the idea of an experience's being of something distinct from it. Assuming that features and processes can be conceived of as being distinct from our experiences of them, one can conceive of one's experience as experience of an objective world without conceiving of it as experience of a world of objective *particulars*. To this extent, the conception of an objective world need not be the conception of a world of *objects*, but it must be the conception of a world which can exist unperceived.

On this reading of 'objective experience', what would it be for 'space' to be among its necessary conditions? One thing that Evans might have in mind is:

(ST_e) The existence of space is necessary for objective experience.

I am going to take it that for 'space' to 'exist' is for spatial objects or phenomena to exist. A spatial object or phenomenon is something that has spatial properties. Extension, solidity, shape, and location are all examples of spatial properties. So (ST_e) is the thesis that objective experience requires the existence of objects or phenomena with at least some of these properties.

On a different reading, suggested by several of Evans's formulations, the Spatiality Thesis is the thesis:

(ST_i) The idea of space is necessary for objective experience.

I will assume that to have what Evans calls the 'idea' of space is to possess some spatial concepts and to be able to use them in spatial thinking or reasoning. Spatial concepts are concepts of spatial properties, and spatial thinking is thinking in which spatial concepts figure essentially. On this reading, the plausibility of the Kantian thesis depends on the plausibility of maintaining that spatial concepts and spatial thinking are needed to make adequate sense of the notion of existence unperceived.

A third reading of the Kantian thesis is suggested by the observation that spatial properties are not just ones that we are able to think about; they are also ones that we are able to perceive. We can perceive extension, solidity, shape, and location, and can perceive things as possessing such properties. If perception of spatial properties is described as the *perception of space*, then the Spatiality Thesis can then be understood as the thesis:

 (ST_p) The perception of space is necessary for objective experience.

Given Evans's conception of objective experience, the question that (ST_p) raises is whether the perception of space is required for possession of the idea of an objective world.

As I have already indicated, 'Things without the mind' is primarily concerned with (ST_i) . The question with which Evans begins his paper, and which he represents himself as trying to answer, is 'what is the connection between the idea of an objective world and the idea of a spatial world?' (1980*a*: 249). It is in the context of trying to answer this question that he suggests that, in order to defend the Kantian thesis, the idea of space must be shown to be implicitly involved in that of existence unperceived. He gives several reasons for thinking that these ideas are connected in this way, and I will have more to say about these reasons below. The important point for present purposes is that Evans's arguments for the Kantian thesis appear, at least in the first instance, to be arguments for (ST_i) .

This is not to say that these arguments have no bearing on (ST_e) and (ST_p) . It is important to bear in mind that Evans argues for (ST_i) by elaborating 'two different reasons for doubting whether a subject whose experience was wholly auditory could be regarded as having a conception of an independent reality' (1980a: 250-1). The hypothetical subject ('Hero') who figures in Evans's discussion is one whose experience is wholly auditory, at least in part because he 'inhabits a purely auditory universe' (1980a: 274). A purely auditory universe is one that is composed of sounds rather than material substances. Assuming that sounds have no intrinsic spatial characteristics, Hero's purely auditory universe will be what Strawson calls a 'No-Space world' (1959: 63). In part 3 of his paper, Evans argues that Hero would be unable to conceive of his world as an objective world. If, as Evans sometimes implies, the fact that Hero inhabits a No-Space world were at least partly responsible for depriving him of the conception of an independent reality, then it would seem that the existence of space is, just as (ST_e) claims, necessary for objective experience.

How is the fact that Hero inhabits what is in fact a No-Space world, a 'world without substance' (Evans 1980*a*: 268), supposed to explain his lack of spatial concepts, and therefore his inability to make sense of the same thing existing both experienced and unexperienced? A natural thought would be that Hero lacks those spatial concepts needed to make sense of existence unperceived because there would be no such thing as the perception of space in a No-Space world. The idea here is that the appropriate spatial concepts would not be available in the absence of spatial perception. So if objective experience requires spatial concepts, then it also requires the perception of space. On the present reading of 'Things without the mind', this is the force of (ST_p) , although it is not absolutely clear whether Evans himself would endorse this argument for (ST_p) .

It seems, then, that Evans can be seen as arguing for (ST_e) and (ST_p) , even though (ST_i) is undoubtedly his primary concern. In the light of this, what is the relationship between the Kantian thesis, as Evans understands it, and the Spatiality Thesis for which Kant argues in the first Critique? Unlike Evans, Kant is primarily concerned to argue for (ST_p) rather than for (ST_i) . In arguing for (ST_p) , Kant assumes that to have 'experience' is to have perceptual knowledge of objects, and that perceptual knowledge of objects involves both thought and perception. In this sense, all experience is objective experience, although Kant's conception of objective experience is narrower than Evans's. On Kant's conception of experience, the sense in which 'space' is an epistemic condition is that the perception of space or spatial properties is necessary for the perception of objects. If, as Kant maintains, we cannot perceive objects without perceiving them as shaped, extended, and located in space, then it follows straightforwardly that these forms of spatial awareness are necessary for perceptual knowledge of objects, and therefore for experience.

Let us now return to my three basic questions. So far, I have said that Kant and Evans are primarily concerned to argue for different versions of (ST), but I have not said what their arguments are. It remains to be seen, therefore, whether they provide satisfactory answers to the question of necessity. In the next two sections (III and IV below), I will examine some of Evans's arguments for (ST_i) and Kant's arguments for (ST_p) in more detail. These arguments will be easier to understand if we have a rough idea of what Kant and Evans have to say in response to the question of foundations and the question of method. The rest of this section will therefore be devoted to outlining and comparing their conceptions of the basis of

epistemic conditions and their views about the proper methodology for establishing necessary conditions for objective experience. My own views about these matters will be set out in section V.

As far as the question of foundations is concerned, the demand for an account of what makes space necessary for objective experience should be understood as the demand for an explanation of the fact that space is an epistemic condition in terms of something more basic. Intuitively, to explain a particular epistemic condition in terms of something more basic is to explain it by reference to independently intelligible factors or considerations, ones that can be grasped without a prior grasp of that very condition. One issue, therefore, is whether it is ever appropriate to try to explain epistemic conditions in this way. It might be argued, for example, that there is nothing more basic or fundamental in virtue of which space is an epistemic condition, no independently intelligible fact or facts by reference to which we can hope to explain the connection between space and objective experience. On this view, the appropriate reaction to the question of foundations is to reject it. All explanation comes to an end once space has been identified as an epistemic condition, and it is fruitless to ask *why* space is necessary for objective experience.

This is not Kant's view. His idea is that there is something more basic that makes it the case that space is an epistemic condition: namely, the 'subjective constitution of our mind' (A_{23}/B_{38}) .² To be more precise, Kant's proposal is that what makes space necessary for perceptual knowledge of objects is the fact:

(FS) Space is a form of human sensibility.

Kant's idea is that (FS) is intelligible independently of (ST_p) , and so can be used to explain in more basic terms why (ST_p) is correct. On this view, the point at which all explanation comes to an end is the point at which space is identified as a form of human sensibility. As Strawson remarks, Kant regards it as 'an ultimate *fact* about the cognitive equipment of us human beings—as something not capable of further explanation' (1997: 237) that we have just the forms of sensibility that we do have. Other beings might be different, but we are not in a position to judge whether our mode of perceiving objects is 'peculiar to us' (A42/B59). This means that we are also not in a position to judge whether space is only a necessary

² All references in this form are to N. Kemp Smith's translation of Kant (1787).

condition under which perceptual knowledge of objects is possible *for us* or a necessary condition for any perceptual knowledge of objects, human or otherwise.

What would it be for space to be a 'form' of human sensibility? On one interpretation, regarding space as a form of human sensibility is equivalent to regarding the perception of space as necessary for the perception of objects. On this interpretation, Kant would be left with no substantive account of the basis of (ST_p) . Instead, he would be forced to concede that (FS) and (ST_p) are notational variants, and that (FS) cannot be used to explain in more basic terms why (ST_p) is correct. Since Kant's aim is precisely to explain in more basic terms why (ST_p) is correct, he must show that regarding space as a form of human sensibility is not equivalent to regarding the perception of space as necessary for the perception of objects. He must show that there is some other way of understanding (FS), and must explain how (FS) can be intelligible independently of (ST_p) .

In representing space as an epistemic condition that is grounded in the forms of human sensibility, Kant is denying that the nature of the objects themselves can account for the connection between space and perceptual knowledge of objects. He is committed to denying this because he thinks that 'space does not represent any determination that attaches to the objects themselves' (A26/B42). Given the non-spatiality of the objects themselves, only our subjective constitution can account for the dependence of objective experience on space. This is intended as a substantive account of the basis or foundation of a particular epistemic condition. In general, Kant's view is that epistemic conditions have foundations that are entirely subjective. In other words, he thinks that they reflect the nature of our cognitive equipment rather than the nature of things or objects as they are in themselves. The conclusion he draws from this is that the proper procedure for establishing or discovering what is necessary for objective experience is to study our cognitive equipment. This is the gist of his answer to the question of method, although he would add that what is needed to establish epistemic conditions is a 'transcendental' rather than an empirical investigation of our cognitive equipment.³

³ Kant thinks that an empirical investigation of our cognitive equipment cannot establish epistemic conditions, because he is convinced that epistemic conditions are, by their nature, ones our knowledge of which must be a priori. There is more on this below.

Kant's conception of epistemic conditions is an idealist conception.⁴ To be an idealist about epistemic conditions is to view them as having entirely subjective foundations, and therefore as conditions that might not bind beings whose subjective constitutions are very different from ours. It remains to be seen whether there are any good arguments for idealism, but the lesson of the discussion so far is that idealism has no chance of being viable unless it is able to provide characterizations of our subjective constitution that are not simply equivalent to statements of those very epistemic conditions claimed to be grounded in our subjective constitution. I will have more to say below about the prospects for idealism and about the relationship between (FS) and (ST_p). The immediate priority is to bring the present phase of my discussion to a conclusion by giving an indication of Evans's approach to my second and third questions, and comparing his conception of the basis of the link between space and objectivity with Kant's conception of the basis of this link. The obvious question to ask in this connection is whether Evans's implicit conception of the basis of epistemic conditions is an idealist conception.

An immediate difficulty is that Evans makes no explicit attempt to explain the fact that objective experience requires the idea of space. At best, it is only by implication that he says anything that bears on this issue. One clue to his thinking is provided by his description of the connection between space and objectivity as a 'conceptual connection' that lies 'deep in our conceptual scheme' (1980a: 249-50). Further clues are provided by the concluding paragraphs of 'Things without the mind'. These paragraphs contain a ringing endorsement of Strawson's methodology in Individuals, a work in which Strawson represents himself as doing what he calls 'descriptive metaphysics'. Unlike other forms of metaphysics, descriptive metaphysics 'is content to describe the actual structure of our thought about the world' (1959: 9). Its aim is to bring to light certain very general features of our conceptual structure, and it proceeds by tracing the relations that obtain between the fundamental elements of this structure. This is what Evans is

⁴ See Allison (1983) for an extended defence of Kant's idealist conception of epistemic conditions. There are some telling criticisms of Allison's arguments in Guyer 1987: ch. 15.

alluding to when he remarks in his final paragraph that 'the connections between the fundamental concepts of our conceptual scheme are central objects of philosophical investigation' (1980*a*: 290). The implication is that the connection between space and objectivity is one such connection, and that the methodology of 'Things without the mind' is essentially the same as Strawson's in *Individuals*.

If this is right, then we already have Evans's answer to the question of method. More controversially, these remarks might also be thought to suggest an answer of sorts to the question of foundations. The answer they suggest is that our conceptual scheme, the structure of our thought about the world, is the basis or foundation of the conceptual connection between space and objectivity. To put it another way, the suggestion that is starting to emerge is that what makes it the case that space is necessary for objective experience is the fact:

(CS) The ideas of space and objectivity are connected in our conceptual scheme.

Two ideas are connected in our conceptual scheme just if one of them implicitly involves the other. As long as our conceptual scheme is such that the idea of space is implicitly involved in that of objectivity, we can conclude that the idea of space is conceptually necessary for objective experience.

This reading of Evans points to a kind of structural parallel between his view and Kant's. Just as Kant regards (FS) as the basis of (ST_p) , so Evans can be read as holding that (CS) is the basis of (ST_i) . More generally, just as Kant attributes epistemic conditions to our subjective constitution, so it now seems that Evans attributes them to our conceptual scheme. This is not to say that what Kant calls the 'subjective constitution of our mind' is the same as what Strawson and Evans refer to as 'our conceptual scheme'. For Kant, our subjective constitution consists of a set of interrelated cognitive faculties or capacities, such as sensibility and understanding. In contrast, a conceptual scheme is a set of interrelated concepts, or a set of propositions in which the concepts that constitute the scheme find essential employment.⁵ In this sense, there is a difference

⁵ This account of the notion of a conceptual scheme is drawn from Strawson (1982).

between regarding our subjective constitution and regarding our conceptual scheme as the basis of a particular epistemic condition. Nevertheless, the point of insisting on a structural parallel between the two approaches is that they both represent the fact that space is necessary for objective experience as a reflection of facts about *us*, facts about our perceptual or conceptual equipment.

Despite this parallel, it would be premature to conclude that Evans has an idealist conception of the basis of (ST_i) . There are two issues here. One is whether he really thinks that (CS) is the basis of (ST_i). The other is whether, if he does think this, his account would amount to a form of idealism. On the first of these issues, it may well be the case that the suggestion in 'Things without the mind' that the connection between space and objectivity is one that lies deep in our conceptual scheme is not intended as a substantive answer to the question of foundations. In order to think of (CS) as the basis of (ST_i) one would need to suppose that (CS) and (ST_i) are independently intelligible, but there is no evidence that Evans regards (CS) and (ST_i) as independently intelligible. His view might be that what it is for the ideas of space and objectivity to be connected in our conceptual scheme, or for one of these ideas implicitly to involve the other, just is for the idea of space to be necessary for objective experience. In that case, (CS) cannot be used to explain in more basic terms why (ST_i) is correct. And if Evans is not trying to explain in more basic terms why (ST_i) is correct, then he cannot reasonably be accused of being an idealist. The issue of idealism arises only for those who accept the question of foundations, but perhaps Evans would not regard this as a legitimate question.

Suppose, on the other hand, that we are not moved by these considerations, and insist on reading Evans as trying to explain in more basic terms what makes (ST_i) true. Given my characterization of idealism, the proposal that (CS) is the basis of (ST_i) would still only count as a form of idealism if two further conditions were met. First, it would have to be the case that conditions of objective experience that reflect our conceptual scheme are ones that *do not* reflect the nature of reality as it is in itself. Secondly, it would have to be the case that conditions of objective experience that reflect our solutions of objective experience that reflect scheme might not bind beings whose conceptual schemes are very different from ours. To put it another way, it would have to be the case that the epistemic conditions that apply

Quassim Cassam

to us might not be universal in scope. The best way of showing that Evans is not committed to some form of idealism would therefore be to show that his account of (ST_i) does not meet one or both of these conditions.

With regard to the first condition, the idea that reality as it is in itself is non-spatial plays no part in Evans's discussion. To this extent, it would be open to him to represent the link between space and objective experience as one that reflects the nature of reality as well as our conceptual scheme. The natural way of doing this would be to insist that 'our thinking is rooted in the nature of the world and in our own natures' (Strawson 1956: 107). The idea would be that we ourselves are spatial objects living in a spatial world, and that this somehow explains the connection in our conceptual scheme between the idea of an objective world and the idea of a spatial world. If this is right, then the proposal that our conceptual scheme is the basis of a particular epistemic condition such as (ST_i) does not imply that this condition has entirely subjective foundations.

On the question of whether space is an epistemic condition that might fail to be universal in scope, Evans concedes at one point that a spatial scheme is not the only scheme that is capable of making sense of the idea of existence unperceived. He is not troubled by this concession because, as a descriptive metaphysician, he is concerned primarily with our own conceptual scheme. As long as our scheme is one where the idea of space is implicitly involved in the idea of existence unperceived, the obvious way of acknowledging the possibility of alternative schemes without giving up (ST_i) would be to read this thesis as claiming only that the idea of space is, for us, necessary for objective experience. In this way, the concession that there might be non-spatial ways of making sense of existence unperceived leads smoothly to the conclusion that the spatial epistemic conditions under which objective experience is possible for us might not be universal in scope.

To sum up, I have argued that one way for Evans to counter accusations of idealism would be to deny that his remarks about the link between (CS) and (ST_i) constitute a response to the question of foundations. Another possibility would be for him to admit that he is addressing the question of foundations in these remarks, but to deny that his response is an idealist response. He could deny this on the basis that there are two conditions that need to be fulfilled

for an account of (ST_i) to count as an idealist account, and that his account does not fulfil both conditions. It is arguable that it fulfils the second condition, but it does not fulfil the first. Unless it can be shown that epistemic conditions that have their basis in our conceptual scheme are entirely subjective in origin, the structural parallels between Evans's account of (ST_i) and Kant's account of (ST_p) do not justify the conclusion that Evans is an idealist.

This discussion leaves many important questions unanswered. For example, it does not indicate whether Evans would be right to reject the question of foundations, if he does reject this question. It suggests various ways in which Evans could counter accusations of idealism, but it does not say which of these ways of proceeding is the most promising. I will argue below that the question of foundations is not one that should be dismissed as illegitimate or misguided. It is possible to explain in more basic terms the connection between space and objective experience, but the best explanation will not be one that represents our conceptual scheme as the basis of this connection. Even if we think of our conceptual scheme as rooted in the nature of the world, it is still misleading to attribute epistemic conditions to our conceptual scheme. It will become clearer why this is misleading once we have a better understanding of the precise sense in which space is necessary for objective experience. This is therefore the point at which we need to turn to the question of necessity. Once we have reflected on the plausibility of regarding space as necessary for objective experience, we should be in a better position to understand why we should neither reject the question of foundations nor respond to it in any of the ways which I have discussed so far.

III

Why should one suppose that space is a necessary condition for objective experience? Evans's first argument for the Kantian thesis is primarily an argument for (ST_i). It turns on the claim that 'the idea of existence now perceived, now unperceived, is not an idea that can stand on its own, stand without any surrounding theory' (1980*a*: 261). What provides the indispensable surrounding for this idea is a rudimentary theory of perception. Such a theory must incorporate an account of the enabling conditions of perception: that is,

Quassim Cassam

an account of the conditions that must be fulfilled for a perceiver to perceive what is there to be perceived. The required theory must also be able to explain why a perceptible phenomenon may not be perceived, and it must do so in a way that is subject to 'significant empirical control' (1980*a*: 268). If we are entitled to rely on spatial notions, then it is easy to explain why a perceptible phenomenon may not be perceived. For example, one might fail to perceive the ϕ -ing that is there to be perceived simply because one is in the wrong position to perceive it or because there is something in the way. The question, therefore, is whether we can 'find a non-spatial way of making sense of existence unperceived' (1980*a*: 263).

One proposal that Evans considers is that existence unperceived can be explained by citing deficiencies in the perceiver, such as his not being receptive in the appropriate modality. On this account, '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' (1980a: 265). This is similar to the way in which, in a spatial theory, one can tell that one's position has changed by the changing course of one's experience, but only when this is taken together with a map of the world, which must itself be constantly updated by adopting views as to when and where one is moving. Unlike the scheme using receptivity, a spatial theory employs a relativized notion of receptivity: one is receptive to (located at) this or that position, rather than absolutely receptive or unreceptive. What the spatial scheme and the scheme using receptivity have in common, however, is that they both take the condition that accounts for the presence or absence of perception to be 'a priori connected with, and, therefore, known to be satisfied only upon the basis of, propositions about the way the world is' (1980a: 266). According to Evans, this feature of the two schemes is not only permissible but necessary for a viable account of existence unperceived.

Why, then, does Evans object to the scheme using receptivity? His objection to it is that its explanation of existence unperceived is not subject to adequate empirical control, and that this is so because it does not employ a relativized notion of receptivity. Hero can suppose that if he were now to become receptive, he would now perceive ϕ -ing, but there is no criterion of his now becoming receptive other than his perceiving ϕ -ing. In contrast, it is not true in a spatial scheme that one's present experience is the only criterion

270

of one's being at a certain location. Given that one moves continuously through space, one can also fix one's location by reference to propositions about adjacent places. Evans's conclusion is that:

here, surely, are the materials for a possible 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. (1980*a*: 268)

Assuming that a spatial theory is one that employs spatial concepts and spatial reasoning, the version of the Kantian thesis supported by these considerations is (ST_i) . To have spatial concepts and to be able to engage in spatial reasoning of the kind required to explain why a perceptible phenomenon may not be perceived just is to have the 'idea' of space.

I will refer to this argument for (ST_i) as the *Perception Argument*. As I have stated it, this is not an argument for (ST_e) or (ST_p) . It is one thing to hold that only a spatial theory of perception can explain the possibility of existence unperceived, but it is a further question whether the ability to make sense of existence unperceived in spatial terms requires the existence of space or the perception of space. To turn the Perception Argument into an argument for (ST_e) or (ST_n), much more would need to be said about the conditions under which spatial thinking is possible. It would need to be argued, for example, that the spatial concepts that figure in such thinking would be available to one only in a spatial world or in a world perceived to have spatial properties. It is instructive, therefore, that Evans makes no attempt to argue along these lines, and thereby to relate the Perception Argument to (ST_e) or to (ST_p) . As far as he is concerned, the only Kantian thesis directly at issue in relation to this argument is (ST_i) .

How convincing is the Perception Argument? It is striking that Evans's own endorsement of it is distinctly half-hearted. After describing it as providing the materials for a possible defence of the Kantian thesis, he adds in a surprising and significant footnote that:

a spatial scheme is not the only scheme to employ a relativized receptivity condition, with the possibilities of additional empirical control that 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. (1980*a*: 268 n. 19)

Quassim Cassam

On the face of it, this concession is at odds with the claim in the text that only a spatial theory will do when it comes to making sense of the notion of existence unperceived. It now appears that there are other possibilities: that is, non-spatial but still empirically serviceable ways of making sense of existence unperceived. If Evans is right about this, then the Perception Argument fails to establish that the idea of space is necessary for objective experience.

One reaction to Evans's footnote would be to dismiss it as misguided. After all, the alternatives that Evans mentions are not spelled out in any detail, and perhaps detailed scrutiny will reveal each alternative to be unworkable. On the other hand, it is hard to see how one can know in advance that this will be the outcome. What guarantee is there that it will not turn out to be possible, with sufficient ingenuity, to find a viable non-spatial way of making sense of existence unperceived? As the complex discussion of the scheme using receptivity illustrates, it is not as if the thought that there might be non-spatial ways of making sense of existence unperceived leaves us blank. A great deal depends on the precise extent to which the factor accounting for the presence or absence of perception must be subject to empirical control, and this is surely a delicate matter. Viewed in this light, it is not obvious that the concession that Evans makes in his footnote is misguided.

A different reaction to the footnote, which is more in keeping with an idealist conception of epistemic conditions, would be to argue that it poses no threat to (ST_i) when this thesis is understood in the way that Evans understands it. The fact that a spatial scheme is not the only scheme to employ a relativized receptivity condition does not show that space and objectivity are not connected in our thinking, or that we have the option of making sense of existence unperceived in non-spatial terms. In other words, it might still be true that the idea of space is, as far as we are concerned, necessary for objective experience even if there are alternatives to the spatial scheme. As long as these alternatives are not alternatives for us, there is no reason to abandon something along the lines of (ST_i). The most that the footnote shows is that this thesis is correct only in a somewhat restricted sense, but this is in keeping with Evans's insistence that the primary aim of 'Things without the mind' is to illuminate 'the role of space in our thinking' (1980a: 268).

The problem with this line of argument is that it does nothing to justify the assertion that we do not have the option of making sense

of existence unperceived in non-spatial terms. If, as Evans concedes, there are alternatives to the spatial scheme, then we need to know why these are not alternatives for us. One thought might be that the alternatives that Evans has in mind are too complicated for us to grasp, or that there is some other reason why we are psychologically incapable of operating with anything other than a spatial scheme. It is not clear, however, why one should suppose that non-spatial ways of making sense of existence unperceived are bound to be any more complicated than spatial ways of doing this. Furthermore, even if it is true that we are psychologically incapable of operating with anything other than a spatial scheme, this is presumably not the sense in which Evans regards the idea of space as necessary for objective experience.

For these reasons, Evans's footnote must be regarded as posing a genuine threat to the Perception Argument for (ST_i). This makes it important to understand why Evans is apparently unconcerned about this threat. The explanation is that he is really engaged in two different projects but mistakenly runs them together. The first is the Kantian project of identifying necessary conditions for objective experience. The second is the Strawsonian project of describing and understanding our own conceptual structure. Yet understanding our own conceptual structure need not be a matter of showing that we must think about the world in the ways in which we do think about it, or that there are no alternatives to our perspective on the world. This is the sense in which the Kantian and Strawsonian projects are different, though not incompatible. If we employ certain concepts in our objective experience, that is one thing; it is a different matter entirely whether our employment of these concepts is strictly necessary for objective experience. If we want to show that we must think about the world in a certain way, then we cannot be content to describe the actual structure of our thought about the world. We must go beyond merely descriptive metaphysics.

These points also have a bearing on Evans's second main argument for the Spatiality Thesis. Since the question to which Evans represents this argument as providing an answer is whether there can be a world without substance, I will call it the *Substance Argument*. The substance that is at issue here is material substance, the kind of substance that is constituted by its possession of primary properties such as position, shape, size, and motion. These are properties of bodies or material substances 'immediately consequential upon the idea of space-occupation' (1980*a*: 269). Evans assumes that a purely auditory world would be a world without material substance, and also that one would lack the concept of substance in such a world. One would lack this concept because one would lack concepts of the primary properties of matter. The aim of the Substance Argument is to show that 'without ideas corresponding to our ideas of the primary properties of matter', one would not be able to make sense of 'the same thing existing both experienced and unexperienced' (1980*a*: 278). This amounts to an argument for the Spatiality Thesis, since primary properties are spatial properties, and ideas corresponding to our ideas of the primary properties of the primary properties of matter are spatial ideas.

Although the Substance Argument is primarily an argument for (ST_i) and (ST_e) , it can also be seen as providing some support for (ST_p) . It is an argument for (ST_i) because (ST_i) is just the claim that spatial ideas are needed to make sense of the notion of existence unperceived. What the Substance Argument adds to this claim is the requirement that the spatial concepts indispensable for objective experience are specifically concepts of primary properties. The sense in which the Substance Argument supports (ST_e) is that it represents the availability of such concepts as conditional on the existence of material substance. Finally, the case for reading the Substance Argument as providing some support for (ST_p) is that the mere existence of material substance cannot account for one's possession of concepts of primary properties. In order to have ideas corresponding to our ideas of the primary properties of matter, it is arguable that one must also be able to perceive the primary properties of matter, and so must be capable of spatial perception.

Is the Substance Argument successful in its attempt to show that a thinker who lacks concepts of primary or spatial properties could not make sense of the same thing existing both experienced and unexperienced? Evans writes that '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)' (1980a: 277). Suppose that we call this the *duality requirement* on objectivity. According to the Substance Argument, one must have concepts of primary properties in order to fulfil this requirement. In a purely auditory universe, in which one would not have such concepts, the thought that one would have certain auditory experiences if one were to go to a certain position would not be the thought of 'two states of affairs existing simultaneously and related causally' (ibid.). So it would not be the thought of one's experiences as experience of an objective world.

It is beyond dispute that the notion of objectivity involves the kind of duality that Evans describes, but it is a further question whether one must have concepts of primary properties in order to make room for this duality. Evans's idea is that, in order to make room for this duality, one must have the idea that the phenomena that one experiences have persisting or relatively abiding properties. Concepts of primary properties come into the picture because they are concepts of persisting or relatively abiding properties. So, for example, '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' (1980a: 278). Yet this supports the conclusion that one's thinking will satisfy the duality requirement if one possesses concepts of primary properties. It does not support the conclusion that one's thinking will satisfy the duality requirement only if one possesses concepts of primary properties. For concepts of primary properties to be strictly necessary for the notion of objectivity, it would need to be the case that one can only think of something as persisting or as relatively abiding by thinking of it as possessing primary properties. One way of undermining the Substance Argument would therefore be to dispute the assumption that only primary or spatial properties can coherently be thought of as persisting.

The thought that persistence and existence in space go together is one that Kant endorses when he writes in the General Note on the System of Principles that 'space alone is determined as permanent' (B291). In response, it might be argued that the most that can be claimed is that there is a connection 'in our conceptual scheme' between the idea of persistence and the idea of existence in space, but that there might be conceptual schemes in which sense is made in some other way of the idea of persistence. Given the general tenor of his discussion, and his own reaction to the Perception Argument, it is hard to see how Evans could object to this move. Indeed, even the extent to which persistence and existence in space are linked in our conceptual scheme is easily exaggerated. For example, it is arguable that we can conceive of a buzzing or a headache as persisting or relatively abiding without conceiving of it as having primary properties or as being produced by something with primary properties. So when Evans suggests that one must have the concept of space-occupying substance in order to satisfy the duality requirement, he is simply describing one very central and obvious way of satisfying this requirement. He cannot claim to be describing the only possible way of satisfying this requirement, even by the lights of our own conceptual scheme.

There are other aspects of the Substance Argument for (ST_i) that I have not discussed and which suggest alternative ways of linking the notion of objectivity with concepts of primary properties.⁶ For reasons that I do not have the space to go into here, I do not believe that these other aspects of the Substance Argument are any more convincing than the aspect that I have just discussed. But even if I am wrong about this, there is still a question about the role of the Substance Argument in 'Things without the mind'. For Evans elaborates this argument after conceding that a spatial scheme is not the only scheme to employ a relativized receptivity condition, with the possibilities of additional empirical control that this provides. I have already commented on the way that this concession undermines the Perception Argument, but does it also undermine the Substance Argument? One thought might be that it does undermine this argument, because a non-spatial scheme that employs a relativized receptivity condition would not be one that employs concepts of primary properties. So if such a scheme can explain why a perceptible phenomenon may not be perceived, then it follows that a grasp of the idea of existence unperceived does not require a grasp of concepts of primary properties.

In order to deflect this objection to the conclusion of the Substance Argument, one would need to show that one cannot make sense of the notion of existence unperceived simply by citing

⁶ For example, I have not discussed the idea that sounds and other secondary qualities are dispositions that require categorical grounds of a special sort. Furthermore, the Perception Argument and the Substance Argument are not the only arguments for the Spatiality Thesis in 'Things without the mind'. There is also the argument in part 4 of this paper to the effect that one must have 'simultaneous' spatial concepts in order to think of unperceived particulars existing simultaneously with perceived particulars. I am unable to discuss this argument here.

failures in receptivity, even if one has a relativized notion of receptivity. One would need to show that this explanation of the possibility of existence unperceived needs to be supplemented by an appeal to the idea that the perceptible phenomena to which one is sometimes receptive possess primary properties. The case for requiring this form of supplementation is that it is only on the basis of the idea that the world one inhabits is a material world that one can conceive of the phenomena to which one is receptive at various different times as persisting through gaps in observation. At this point, however, one would once again be drawing upon elements of the Substance Argument I have already called into question, and so would once again be faced with the objection that persistence and existence in space can be pulled apart.

Where does this leave the Substance Argument for (ST_e) and (ST_p) ? As far as (ST_e) is concerned, it is one thing to argue that the availability of concepts of primary properties is conditional on the existence of space, but this is not going to cut much ice if concepts of primary properties are not necessary for objective experience. By the same token, the failure of the Substance Argument for (ST_i) means that it is not going to be possible to argue for the indispensability of spatial perception on the basis that the perception of space is the source of the idea of space. Apart from anything else, this argument relies on an empiricist conception of conceptacquisition, but Evans is no empiricist when it comes to accounting for ideas of primary properties. He emphasizes that these concepts cannot be woven exclusively out of materials given in experience, and implies that it would be closer to the truth to describe them as innate or natively given.⁷ In that case, there is a sense in which, even by Evans's own lights, the perception of space is not the source of the idea of space, although there might still be a sense in which spatial perception is needed to trigger and justify one's innate spatial ideas.

These considerations suggest that the Substance Argument for (ST_e) and (ST_p) is as inconclusive as the Substance Argument for (ST_i) . Does this mean that we should reject the Kantian thesis? This would be an overreaction to what I have just been arguing. Aside from the fact that I have not been able to consider all of Evans's arguments for his preferred version of the Kantian thesis, it would

⁷ See, in particular, Evans 1980*a*: 271.

Quassim Cassam

not be right to reject this thesis without considering Kant's own arguments for it. Unlike Evans, Kant argues for (ST_p) rather than for (ST_i) , and he does not argue for (ST_p) by first arguing for (ST_i) . Instead, he argues directly for (ST_p) . If Kant's argument is successful, then one can agree that the *perception* of space is necessary for objective experience even if one is not persuaded by Evans that the *idea* of space is necessary for objective experience. The next question that needs to be addressed, therefore, is whether Kant makes a convincing case for (ST_p) .

IV

Kant's claim is that the perception of space or spatial properties is necessary for objective experience because it is necessary for the perception of objects. In the terminology of the Transcendental Aesthetic, it is a condition 'under which alone objects can be for us objects of the senses' (A29). The 'objects' at issue here are *objective particulars*: that is, particulars that can be perceived yet exist unperceived. The perception of space is necessary for the perception of objective particulars, because one cannot perceive an objective particular without perceiving it as having spatial properties such as shape, extension, and location. This is one sense in which space is one of the 'forms' of objective experience. Kant's idea is that space is necessary for objective experience because we can only perceive objective particulars by perceiving their spatial properties. This leaves it open whether we can think about objective particulars without thinking about their spatial properties.

Unfortunately, there are counter-examples to the proposal that we can only perceive objective particulars by perceiving their spatial properties. For example, sounds can be regarded as objective particulars, but it is false that one can only perceive sounds by perceiving their spatial properties. For a start, it is doubtful whether sounds have any intrinsic spatial properties or characteristics. It does not follow from this that sounds cannot be heard *as* having spatial properties, but it is independently implausible that one cannot hear a sound without hearing it as having any spatial properties.⁸ We do not hear sounds as shaped or extended or as

278

⁸ See Nudds (2001) for further discussion of this point.

occupying space. The only spatial property that we sometimes hear sounds as having is location, but it is possible to hear a sound without hearing it as located or as coming from any particular direction.⁹

Unless one is prepared to deny that sounds are objective particulars, one is going to have to concede that Kant is wrong to represent the perception of space as a necessary condition for the perception of objects. The reason why he makes this mistake is that the objects he has in mind throughout his discussion are *material* objects or bodies. He makes this explicit in the following important passage:

[I]f I take away from the representation of a body that which the understanding thinks in regard to it, substance, force, divisibility, etc., and likewise what belongs to sensation, impenetrability, hardness, colour, etc., something still remains over from this empirical intuition, namely, extension and figure. These belong to pure intuition, which, even without any actual object of the senses or of sensation, exists *a priori* in the mind as a mere form of sensibility. (A20–1/B35)

This passage suggests that it might be possible to defend a somewhat more restricted version of (ST_p) than the version undermined by reflection on the nature of auditory perception. Instead of taking objective experience to involve the perception of objective particulars, Kant could stipulate that objective experience involves the perception of material objects. According to the restricted version of (ST_p) , the perception of space is necessary for objective experience in so far as it is necessary for the *perception of bodies*. I will refer to this version of (ST_p) as (ST_{pb}) . The sense in which, according to (ST_{pb}) , the perception of space is necessary for the perception of bodies or material objects is that we can perceive bodies or material objects only by perceiving their spatial properties.

One of the attractions of restricting (ST_p) in this way is that sounds are no longer a problem. Sounds are not bodies, so the possibility of hearing a sound without hearing it as having any spatial properties is not a counter-example to (ST_{pb}) . Unlike sounds, bodies have 'extension and figure'. They also occupy space to the exclusion of other bodies. Locke describes this characteristic of bodies as their 'solidity' or 'impenetrability'. Bodies are

⁹ I am indebted at this point to discussion with John Campbell, and to Campbell (1997).

Quassim Cassam

an important subclass of objective particulars, and it is easy to forget that there are objective particulars that are not bodies. This is Kant's mistake. He claims that the perception of space is necessary for the perception of objective particulars because he assumes that all objective particulars are bodies. Once it is conceded that this is a mistake, it must also be conceded that the perception of space need not be a necessary condition for the perception of those objective particulars that are not bodies.

Although (ST_{pb}) is more plausible than the unrestricted version of (ST_p) , it still faces a number of challenges. Consider the case in which one hears a car alarm sounding in the middle of the night. A car alarm is a material object, but one can hear a car alarm sounding without hearing it as shaped or solid or even as spatially located. Does this not undermine (ST_{pb}) ? The reason that it does not is that the primary or direct object of audition in this case is not the car alarm but the sound it makes.¹⁰ In contrast, cases where it is plausible to insist that what one is perceptually aware of is not just a sound but a body making or producing a sound will not be ones where something is perceived without being perceived as having any spatial properties.¹¹

It is controversial whether the direct objects of audition are sounds or bodies, but it is not controversial that bodies are the primary direct objects of sight and touch. Although (ST_{pb}) draws attention to the fact that it is not possible to perceive a body by sight or touch without perceiving any of its spatial properties, this should not be taken to imply that it is not possible to perceive a body by sight or touch without perceiving *all* of its spatial properties. For example, one might see an object glinting in the far distance without seeing what shape it is or perceiving its solidity. One might believe that the object is shaped and solid, but that is a different matter. On the other hand, there is no question of the glinting object being seen without its being seen as located in relation to the perceiver and, perhaps, other objects. One might sometimes misperceive the locations of the things which one sees, but this is not to say that bodies can be seen without being seen as having a spatial location. The same goes for touch. When one perceives a

¹⁰ M. G. F. Martin also makes this point. See Martin 1997: 93.

¹¹ The cases that I have in mind are ones where one perceives a sound as being produced by something which one can also see. The significance of such cases is illuminatingly discussed in Nudds (2001).

body by touching it, one must be aware of it as located in space. Strictly speaking, therefore, it is the perception of location that is necessary for the perception of bodies.¹²

One reaction to what I have just been arguing would be to protest that it deprives the Spatiality Thesis of any interest. Since bodies are essentially spatial, is it not trivial that the perception of space is required for the perception of bodies? The answer to this question is that it is not trivial. The claim that bodies are essentially spatial is metaphysical or ontological. In contrast, the claim that the perception of space is required for the perception of bodies is epistemological rather than metaphysical. To argue from the premiss that bodies are essentially spatial to the conclusion that the perception of space is necessary for the perception of bodies would be to argue from a metaphysical premiss to an epistemological conclusion. As will become clearer below, it is certainly not beyond dispute whether it is legitimate to argue in this way, any more than it is beyond dispute whether (ST_{pb}) is plausible in its own right. The suggestion that (ST_{pb}) trivializes the Spatiality Thesis underestimates the amount of work that needs to be done to explain the meaning of (ST_{pb}) and to establish its correctness. If it were the case that (ST_{pb}) is not just true, but trivially true, it would be much less difficult to defend than it actually is.

Another anxiety about what I have just been arguing might be that it is unacceptable to assume that objective experience involves the perception of bodies or material objects. But in what sense is this unacceptable? One thought might be that it is unacceptable because it fails to address scepticism about the external world. Yet, if the object of the exercise were to respond to scepticism, then it would be no more acceptable to assume that we have experience of objects that can exist unperceived than that these objects are material objects. Scepticism, however, is not the issue when it comes to establishing a connection between space and objective experience. In this context, the object of the exercise is simply to determine the conditions under which perceptual knowledge of material objects is possible and whether the perception of space is one of these conditions. Sceptical doubts about the existence of

¹² Further discussion of this claim would need to take account of conditions, such as optic ataxia, which appear to call into question the existence of such a strong connection between the perception of bodies and the perception of location.

material objects need to be addressed, if they need to be addressed at all, in a different context.

If, as I have argued, the perception of location is necessary for the perception of bodies, then Evans is wrong to maintain that the Kantian thesis cannot be defended without showing that the idea of space is implicitly involved in the very idea of existence unperceived. For (ST_{pb}) is a version of the Kantian thesis, but it does not require one to suppose that the idea of existence unperceived implicitly involves the idea of space. As far (ST_{pb}) is concerned, the question is not whether objective experience requires the *idea* of space, but whether it requires a capacity for spatial *perception*. Assuming that one can perceive objects as spatially located even if one does not have any spatial concepts, the approach that I have been recommending does not amount to an argument for the version of the Kantian thesis that interests Evans. It is sufficient, however, that it amounts to an effective argument for the version of the Kantian thesis that interests Kant.

V

Why is the perception of location necessary for the perception of bodies? Kant's answer to the question of foundations is to represent the fact that space is a form of human sensibility (FS) as the basis of (ST_{pb}) . This is an attempt to explain an epistemic condition in more basic terms, so we need to ask whether (ST_{pb}) and (FS) are independently intelligible. Kant's answer to this question would be to argue that they are independently intelligible, and that the key to recognizing this is to adopt his distinction between the 'matter' and 'form' of 'appearance' or perception. The matter of a perceptual experience is sensation, while its form is the particular way that the matter is given or received in experience. For space to be a form of human sensibility is for the matter of perception to be given or received as spatially ordered, and therefore as spatially located.¹³ According to Kant, it is because the matter of perception is given in experience as spatially ordered that the perception of location is so fundamental to the perception of bodies. At the same time,

¹³ See Falkenstein (1995) for a detailed discussion and defence of this way of reading what I am calling (FS).

it would be open to him to argue that his story about the form of perception is intelligible independently of (ST_{pb}) .

This explanation of (FS) implies that it is sensations that are received as spatially ordered, but a different view is suggested by Kant's remark that the matter of appearance is what 'corresponds to sensation' (A_{20}/B_{34}). What corresponds to sensation consists of various sensible features of material objects such as shape and colour. This suggests that the form of appearance is the way in which, when one perceives a material object, its various sensible features are given to us in experience. Space can be described as a form of appearance or of human sensibility to the extent that these features are given to us in experience as spatially located. The fact that the perception of bodies requires the perception of location can then be seen as a consequence of the independently intelligible fact that space is, in this sense, the form of appearance.

The force of this suggestion can be brought out by considering the so-called Binding Problem. John Campbell gives the following account of this problem:

[T]here is much converging evidence that different properties of an object, such as colour, shape, motion, size, or orientation are processed in different processing streams. This means that the visual system has the problem of reassembling individual objects, as it were, from the results of these specialized processing streams....We do not have perception of an individual object until this Binding Problem has been solved, and various simple sensory properties have been put together as properties of a single object. (Campbell 2002: 30-1)

How, then, does the visual system solve the Binding Problem? Campbell suggests that:

[I]f the visual processing streams contain at least implicit information about the locations of the features being found—just where in the environment a particular colour or shape is, for example—then there is one effective strategy available. Features found at the same location could be put together as features of a single object. (Campbell 2002: 31)

The proposal that location is the key to solving the Binding Problem can be used to make sense of Kant's idea that space is one of the forms of human sensibility. The information provided by different processing streams is what Kant would call the 'matter' of appearance, and the form of appearance or human sensibility is the basis upon which different properties are put together as properties of a single object. If sameness of spatial location is the basis upon which different properties are put together as properties of a single object, then we can say that space is a form of human sensibility.

This account of (FS) makes sense of (ST_{pb}) . It explains why, and in what sense, the perception of location is necessary for the perception of material objects, at least as far as visual perception is concerned. We perceive material objects by perceiving their sensible properties, and we have just seen that different sensible properties will not be perceived as belonging to one and the same object unless they are perceived as having a common location. But if the sensible properties of material objects must be perceived as located, then the objects to which these properties are perceived as belonging must also be perceived as located. On this reading, the claim that the perception of location is a necessary condition for the perception of material objects gives expression to two distinct and plausible claims. The first is that the perception of different sensible properties as properties of a single material object is a necessary condition for the perception of material objects. The second is that the perception of location is a necessary condition for the perception of different sensible properties as properties of a single material object.

Since Kant is an idealist about epistemic conditions, he would maintain not only that (FS) makes sense of (ST_{pb}) , but also that (FS) is the sole basis of (ST_{pb}) . As I have already remarked, he also takes it to be an ultimate and inexplicable fact about our cognitive equipment that space is a form of human sensibility. His implicit argument for regarding epistemic conditions as having entirely subjective foundations is that epistemic conditions are 'a priori conditions' (A94/B126), and that conditions of knowledge or experience can be known a priori only if they are grounded in our subjective constitution. For the purposes of this argument, a priori conditions are ones that can be established without any empirical investigation. In connection with (ST_{pb}), therefore, his claim would be that we can only know without any empirical investigation that the perception of location is necessary for the perception of bodies, because we can know without any empirical investigation that space is a form of human sensibility.

This argument for idealism implies that Kant would not be entitled to defend (ST_{pb}) on the basis of what empirical psychologists say about the Binding Problem. After all, it is not as if we can know

without any empirical investigation how our visual systems solve the Binding Problem. It is doubtful, however, whether Kant is right to insist on the *a priority* of epistemic conditions. He does so on the basis that necessity is a mark of *a priority*, but it is a mistake to think that propositions that state necessary conditions for knowledge or experience must be necessary truths.¹⁴ Once we recognize that epistemic conditions need not be a priori, we are free to draw on empirical considerations in support of (ST_{pb}) . Even if we allow that empirical considerations have a bearing on (ST_{pb}), we would still be left with the idealist's core proposal that it is solely in virtue of the ultimate and inexplicable fact that space is a form of human sensibility that the perception of space is necessary for the perception of bodies.

On reflection, however, it is clear that this proposal is indefensible. For a start, it is scarcely inexplicable that the perception of spatial location is the basis upon which the Binding Problem is solved, given the nature of the objects from which most of our perceptual information derives.¹⁵ These objects are material objects, and to be a material object is to exist in space and time. Assuming that our perceptual systems have evolved in a world of spatio-temporal objects, it is hardly surprising that space and time are the forms of our sensibility. If this were not the case, we would not be able to differentiate objects in perception or identify different sensible properties as properties of one and the same object. This explanation of (FS) is not available to Kant because he thinks that space and time are themselves projections of human sensibility, but this view of space and time is one for which there are no good arguments, either in the first Critique or elsewhere.

It is one thing to explain (FS) by reference to the nature of the objects from which our perceptual information derives, but can (ST_{pb}) be explained in the same way? The thought would be that we cannot hope to understand why the perception of location is a necessary condition for the perception of material objects without taking account of the fact that spatial location is a fundamental and essential feature of material objects.¹⁶ It is fundamental in the sense that it is one of the key factors by reference to which material

¹⁴ I explain why this is a mistake in Cassam (1999).

¹⁵ Martin (1997) makes substantially the same point.

¹⁶ See Martin (1997) for a related line of thought.

Quassim Cassam

objects are individuated. It is essential in the sense that a material object cannot exist without being located somewhere. Some features of material objects, such as shape, are essential without being fundamental, but location is both essential and fundamental. A material object or body cannot exist without a spatial location, and the principle by reference to which material objects are individuated is that two material objects of the same kind cannot be in exactly the same place at the same time. The proposal that I am now considering is that at least part of what makes it the case that the perception of location is necessary for the perception of material objects is the fact that material objects, unlike sounds, are fundamentally and essentially spatially located.

If Kant is an idealist about epistemic conditions, then the position that I have just outlined might be described as amounting to a form of *realism* about epistemic conditions. Realism does not reject the question of foundations, but it denies that epistemic conditions have entirely subjective foundations or that they can be explained solely by reference to our cognitive constitution or conceptual scheme. Its central idea is that epistemic conditions are determined, at least in part, by the nature of the world.¹⁷ It also maintains that epistemic conditions that are rooted in the nature of the world are more likely to be universal in scope, to bind beings whose subjective constitutions are very different from ours, than ones that merely reflect our own subjective constitution.¹⁸ On a realist view, if the nature of material objects makes the perception of location necessary for the perception of material object, then any being capable of perceiving material objects must also be capable of perceiving these objects as spatially located.

I have already anticipated the most serious objection to realism. This is the objection that it moves illicitly from a metaphysical premiss to an epistemological conclusion.¹⁹ Spatial location might be a fundamental and essential feature of material objects, but how

¹⁷ Paul Guyer also draws attention to the possibility that 'something which is a necessary condition of knowledge may reflect the structure of *both* the epistemic subject and the object of knowledge, rather than of the former *instead* of the latter' (1987: 340). There are also traces of realism about epistemic conditions in Strawson (1997). I discuss the dispute between realism and idealism in more detail in Cassam (1999).

¹⁸ See Cassam (1998) for more on the thought that realism goes with the idea that epistemic conditions are universal in scope.

¹⁹ Daniel Warren (1998) presses an objection along these lines.

is it supposed to follow from this that the *perception* of location is necessary for the *perception* of material objects? An effective response to this question would be to point out that perceptual experiences of material objects must be able to represent material objects as distinct from one another. It is hard to see how our experience could be said to present us with material objects without presenting us with *discrete* material objects; but experiences that are capable of presenting us with discrete material objects cannot be ones that do not in any way reflect the basis upon which material objects are individuated. In Peacocke's terminology, they must be experiences that are 'suitably sensitive' (1993: 171) to what makes discrete material objects discrete from each other.

The principle that experiences which present us with discrete material objects must be suitably sensitive to the basis upon which material objects are individuated suggests one reasonable sense in which our epistemology must be governed by our metaphysics. Against this background, the realist's substantive proposal is that experiences of discrete material objects will be suitably sensitive to what makes these objects distinct from one another only if they represent them, together with their sensible properties, as being at different spatial locations. An inability to perceive discrete material objects as being at different locations would amount to an inability to perceive discrete material objects as discrete, and an inability to perceive discrete material objects as discrete would amount to an inability to perceive material objects.²⁰ That is why the perception of location is a necessary condition for the perception of material objects or bodies, and therefore a necessary condition for objective experience. The ultimate basis of this epistemic condition is partly metaphysical or ontological, but the realist's claim is that it is a mistake to think that the fact that the perception of location is necessary for the perception of material objects has nothing to do with the fact that spatial location is a fundamental feature of material objects.

Where does this leave the question of method? Once epistemic conditions are regarded as having an ontological or metaphysical

²⁰ Thought about material objects is subject to a parallel constraint. As Martin points out, 'if one is knowingly to pick out a physical object, then one will be rational only if one takes it that a different object must be in mind where one picks out something at a discrete location, unless one has specific reason to think otherwise' (1997: 90).

Quassim Cassam

foundation, it would be natural to conclude that the project of establishing what is necessary for the perception of bodies is continuous with the project of establishing the essential and fundamental properties of bodies. Although it is not a straightforwardly empirical matter how bodies are individuated, it is not difficult to see that empirical considerations might have a bearing on the principle that no two bodies can be in exactly the same place at the same time. As Leibniz points out, this principle gives expression to the 'reasonable assumption' that 'interpenetration is contrary to nature' (1982: 230), but it is not just a question for armchair philosophy whether this is a reasonable assumption. This suggests that a plausible methodology for establishing such principles, and the epistemic conditions which they ground, will be one that acknowledges a role for both armchair philosophy and empirical science. On a realist view, merely studying our own conceptual apparatus will not tell us how things are in the world, or how we must perceive them as being if we are to perceive them at all.

In my view, something along these lines is the best that can be done for the Kantian thesis. By reading this thesis in the way that the realist reads it, we avoid many of the pitfalls of idealism and are able to provide a comparatively straightforward account of the sense in which space is necessary for objective experience. In a more extended discussion, more would need to be said in defence of the principle that experiences of material objects must be suitably sensitive to the basis upon which these objects are individuated, and in support of the way in which realism exploits this principle in its account of (ST_{pb}) . It is worth remarking, however, that there is very little in the realist approach that Evans is committed to disputing. If what I have been arguing is correct, then it can only strengthen the case for the Kantian thesis to focus on (ST_{pb}) rather than (ST_i), and to avoid giving the impression that our conceptual scheme is the basis of the connection between space and objective experience. From a realist perspective, we can explain in more basic terms why space is necessary for objective experience, but the correct explanation will be one that appeals to independently intelligible facts about the nature of the objects of objective experience, rather than facts about our conceptual scheme or subjective constitution.

Why, then, does Evans interpret and defend the Kantian thesis in the way that he does, rather than in the way that I have been recommending? The answer to this question is that 'Things without the mind' is a deeply Strawsonian piece of work, and that Evans is, to some extent, led astray in this paper by his uncritical acceptance of Strawson's framework. If one approaches the Kantian thesis from the perspective of descriptive metaphysics, one is bound to be attracted by the thought that this thesis turns on whether there is a connection in our conceptual scheme between the idea of space and the idea of an objective world. The least that realism establishes is that there is no need to approach the Kantian thesis in this way, and that there is rather more to be said for this thesis than one might suppose on the basis of Evans's ingenious but ultimately inconclusive discussion.²¹

²¹ I thank John Campbell and Ciara Fairley for helpful discussions of the themes of this chapter and for comments on an earlier draft.

Identity, Vagueness, and Modality

E. J. LOWE

Gareth Evans's classic one-page paper in Analysis, 'Can there be vague objects?' (Evans 1978), must have generated more published commentary per word than just about any other philosophy article written in the twentieth century. This is partly because of the intrinsic interest and importance of its subjectmatter and conclusion, but no less because of the problems of interpretation that its readers are confronted with. I shall not dwell much on these problems here, but take it that Evans's purpose was to demonstrate, by means of a reductio ad absurdum proof, that there cannot fail to be a fact of the matter as to whether an object *a* is identical to an object *b*. My own purpose is to question the validity of the proof, and for this purpose I find it instructive to compare it with another notorious 'proof' of a metaphysically contentious doctrine, the Barcan-Kripke proof of the necessity of identity. More precisely, I shall compare it with a closely related proof of the non-contingency of identity. It seems not implausible, indeed, that Evans had the Barcan-Kripke proof partly in mind as a model for his own proof, given the obvious similarities between them and the notoriety of the Barcan-Kripke proof at the time at which he was writing.

As for my assumption that what is at stake in Evans's paper is the possibility of there *failing to be a fact of the matter* as to whether an object a is identical to an object b, I would simply point out that he opens the paper with the remark 'It is sometimes said

I should like to thank José Luis Bermúdez for his very helpful comments on a previous draft of this paper.

that the world might itself be vague' and contrasts vagueness of this supposed kind with 'vagueness being a deficiency in our mode of describing the world', with which he clearly has no quarrel. In other words, his target is what might be called ontic rather than semantic vagueness-though whether 'vagueness' is really an apt word in the ontic case is a moot point. It is also a moot point whether, in the light of its contents, the title of Evans's paper is apt in representing it as concerning the question of whether there can be vague objects. The real question, I would say, is whether there can be entities whose identities are ontically indeterminate: that is, whether there can ever fail to be a fact of the matter as to whether an entity *a* is identical to an entity *b*. I use the neutral term 'entity' here simply because if one were to judge, contrary to Evans-as I myself do-that some entities can indeed have ontically indeterminate identities in this sense, one might want to reserve the term 'object' for entities whose identities are not ontically indeterminate. This is in fact my own practice: according to my own preferred usage, an 'object' is an entity that has fully determinate identity conditions, so that if a and b are objects of any kind, it follows by definition that there is a fact of the matter as to whether a is identical to b (see Lowe 1998: ch. 3). Obviously, if Evans's argument is correct, such a restriction on the application of the term 'object' is vacuous, because if he is right, no entity can fail to have fully determinate identity conditions and so fail for that reason to be an 'object' in my sense. However, in what follows I shall not use the term 'object' in my own restrictive sense, as this would be confusing in the present context. Rather, I shall fall in with Evans's own practice and that of most philosophers, which makes no distinction of the sort that I wish to make between 'entity' and 'object'.

One other preliminary observation is in order before we turn to Evans's proof itself. This concerns Evans's remark that the idea whose coherence he seeks to call into question is 'the idea that the world might contain certain objects about which it is a *fact* that they have fuzzy boundaries'. This remark confirms that Evans's concern is with *ontic* rather than semantic vagueness, but it is puzzling in its suggestion that the idea of ontic indeterminacy of identity is necessarily connected with the idea of the possession of 'fuzzy boundaries'—by which I assume is meant 'fuzzy' *spatial or temporal* boundaries. It is true that cases of *semantic* vagueness

E. J. Lowe

frequently concern the drawing of such boundaries. For instance, it may be said that our use of the name 'Mount Everest' does not determine a precise spatial boundary between terrain that is part of the mountain so named and terrain that is not. In this case, there are many different ways of drawing a precise spatial boundary all of which are equally consistent with our use of the name: the 'fuzziness' lies not in any boundary that may be drawn, for each boundary that may be drawn is a precise one-rather, it lies in our language, which does not determine that any given one of these precise boundaries must be drawn in preference to any other. But it is far from obvious that ontic indeterminacy of identity would have to be grounded in a genuine 'fuzziness' in boundaries themselves, quite independent of language-as though boundaries could somehow *really* be 'smeared out'. It isn't even clear to me what could be meant by this. Fortunately for the advocate of ontic indeterminacy of identity, making sense of such a notion is not crucial to the position that he seeks to defend. As we shall see, the most plausible cases of ontically indeterminate identity do not turn on the issue of boundaries at all. However, I suspect that Evans's assumption-that ontic indeterminacy of identity would have to have something to do with 'really' fuzzy boundaries-is widely shared (see Keefe 2000: 15) and has done much to perpetuate scepticism about the possibility of such indeterminacy.

Now let us turn at last to Evans's proof itself. The proof contains just five lines. It begins with the following proposition, assumed for *reductio*:

(1) $\nabla(a=b)$

Evans indicates that (I) is to be understood as expressing the assumption that the sentence a = b is of indeterminate truthvalue, with the idea of indeterminacy being expressed by the sentential operator ∇ . So, it seems, (I) may read as 'It is indeterminate whether it is true that a = b', or, more concisely, 'It is indeterminate whether a = b'. And as I implied earlier, I take this to be another way of saying 'There is no fact of the matter as to whether a = b'. For the purposes of *reductio*, (I) is being assumed to be *true*, so it is being assumed that there *is* a fact of the matter as to whether there is no fact of the matter as to whether a = b. I shall return to this point later, since it bears on something that Evans says at the end of his paper. To explain and justify the next step of his proof, Evans says that '(1) reports a fact about *b* which we may express by ascribing to it the property " $\lambda x [\nabla(x = a)]$ " (I use the more familiar lambda symbol in place of Evans's circumflex). Because he takes (1) to report this (purported) fact and expresses the (purported) fact by

(2)
$$\lambda x [\nabla (x = a)] b$$

he takes it that (2) follows from (1). I shall assume that (2) may be read as 'b has the property of being such that it is indeterminate whether it is identical to a', or equivalently as 'b has the property of being such that there is no fact of the matter as to whether it is identical to a'.

Next, Evans asserts as a premise this:

(3)
$$\sim \nabla(a = a)$$

which I shall read as asserting 'It is *not* indeterminate whether *a* is identical to *a*', or equivalently as 'There *is* a fact of the matter as to whether *a* is identical to *a*'. Presumably, what justifies this premise is that it is *true* and so a *fact* that *a* is identical to *a*, whatever object *a* might be. For, surely, if it is indeed a fact that *a* is identical to *a*, then there is a fact of the matter as to whether *a* is identical to *a*.

Evans then supposes that, just as (2) follows from (1), so the following follows from (3):

(4)
$$\sim \lambda x [\nabla(x = a)]a$$

Modelling our reading of (4) on that of (2) above, (4) may be read as 'It is not the case that a has the property of being such that it is indeterminate whether it is identical to a', or equivalently as 'It is not the case that a has the property of being such that there is no fact of the matter as to whether it is identical to a'.

Finally, Evans says that 'by Leibniz's law, we may derive from (2) and (4)', the conclusion of his proof:

$$(5) ~(a = b)$$

Evans clearly has in mind here the version of Leibniz's law which asserts that if an object a is identical to an object b, then a has any property that b has, and vice versa. Contraposing, if a does not have some property that b has, then a is not identical to b. Now, in lines (4) and (2) respectively it is stated that a does not have

a certain property—the property of being such that it is indeterminate whether it is identical to *a*—and that *b does* have this property. Consequently, it may be inferred from (2) and (4) by the contrapositive of Leibniz's law—as above interpreted—that *a* is not identical to *b*, which is what (5) states.

(5) itself does not directly contradict (1), so we do not yet formally have a *reductio ad absurdum* proof of the falsehood of (1). To make good this seeming deficiency, Evans makes the following final remark, which has given rise to some puzzlement and a great deal of discussion:

If 'Indefinitely' and its dual, 'Definitely' (' Δ ') generate a modal logic as strong as S₅, (1)–(4) and, presumably, Leibniz's law, may each be strengthened with a 'Definitely' prefix, enabling us to derive

(5S) $\Delta \sim (a = b)$

which is straightforwardly inconsistent with (1).

The first oddity about this remark is that we were initially prompted to read the sentential operator ' ∇ ' *not* as 'indefinitely', but as something like 'it is indeterminate whether'. In fact, the nearest that Evans comes to spelling out exactly how we are to read a formula like (1) is when he says, by way of introducing (1) as an assumption for *reductio*, 'Let "*a*" and "*b*" be singular terms such that the sentence "*a* = *b*" is of indeterminate truthvalue'. This actually suggests a reading of (1) as 'The sentence "*a* = *b*" is of indeterminate truth-value'. However, this is a metalinguistic statement, whereas Evans quite explicitly intended his symbol ' ∇ ' to be a *sentential operator*—that is to say, an expression that forms a sentence of a given language when it is prefixed to another sentence of the same language. This is why it seems natural to read (1) as was proposed earlier: namely, as 'It is indeterminate whether (it is true that) *a* is identical to *b*'.

However, another possible reading would be something like 'It is indeterminately true that a is identical to b', where this is seen as being analogous to the modal statement 'It is contingently true that a is identical to b' (cf. Parsons 2000: 45 and 204 ff.). And indeed, this analogy might superficially seem advantageous if one wants, as I do, to draw certain parallels between Evans's proof and the Barcan–Kripke proof of the necessity of identity. But this reading requires us to understand (I) as expressing, so to speak, a *way* in which it is (supposedly) *true* that a is identical to b—to wit, 'indeterminately', as opposed to 'determinately'. It is not inconceivable that Evans himself did have something like this in mind (cf. again Parsons 2000: 204 ff.). And indeed, a reading like this might well be appropriate if semantic vagueness were at issue, because a 'supervaluational' treatment of such vagueness would supply a reading of 'It is indeterminately true that *a* is identical to b' as saying that the sentence 'a is identical to b' is true on some but not all precisifications of the references of the names 'a' and 'b' (cf. Lewis 1988). After all, being true on some precisifications is a way of being true. However, I am taking it that semantic vagueness is not what is at stake-and I can make no clear sense of an 'ontic' analogue of such 'indeterminate truth'. I am taking the interest of Evans's proof to lie in its apparent demonstration that there cannot fail to be a fact of the matter as to whether or not an object *a* is identical to an object b. And this is undoubtedly how most other philosophers have viewed it too. So I shall quite unashamedly carry on viewing it in this way.

But now the question is whether, if we view the proof in this way, we can make sense of Evans's final remark, quoted above. At first sight, at least, it doesn't look as though we can. For if ' ∇ ' is read as 'it is indeterminate whether (it is true that)', or 'there is no fact of the matter as to whether', how could this sentential operator be understood to have a *dual*, ' Δ ', in the sense familiar in modal logic? The modal operators \diamond and \Box are 'duals' in this familiar sense, with each being definable in terms of the other together with negation—so that (p) is equivalent to $(\neg p)$ and ' $\Box p$ ' is equivalent to ' $\sim \Diamond \sim p$ '. Obviously, Evans's remark, quoted above, that (5S) is 'straightforwardly inconsistent with (1)' presumes an analogous equivalence between ' ∇p ' and ' $\sim \Delta \sim p$ ', because only if (1) is thus equivalent to ' $\sim \Delta \sim (a = b)$ ' does it contradict (5S). But if we read Evans's other operator, ' Δ ', as 'it is not indeterminate whether (it is true that)', or 'there is a fact of the matter as to whether', do ' ∇ ' and ' Δ ' turn out to be suitably interdefinable with the help of negation? Is it the case that ' ∇p ' is equivalent to ' $\sim \Delta \sim p$ ' on this reading? That is to say, is 'It is indeterminate whether p' equivalent to 'It is not not indeterminate whether not p? (The double negation here is required, of course, since we have elected to read ' Δ ' as 'it is not indeterminate whether'.)

Well, 'It is not not indeterminate whether not p' is obviously equivalent, by double negation elimination, to 'It is indeterminate whether not p', so our question reduces to one of whether this is in turn equivalent to 'It is indeterminate whether p'. But then, surprising though this might have seemed prior to investigation, it turns out that our question does in fact have a positive answer. For it seems clear that 'It is indeterminate whether p' is true if and only if 'It is indeterminate whether not p' is true. That is to say, it seems clear that, as I have proposed otherwise to express it, 'There is no fact of the matter as to whether p' is true just in case 'There is no fact of the matter as to whether not p' is true. For if there was a fact of the matter as to whether not p, this would either be because it was a fact that not p or because it was a fact that p—and, either way, it would follow that there was likewise a fact of the matter as to whether p. We see, then, that even if Evans's operators ' ∇ ' and ' Δ ' are interpreted in the fashion that I have proposed, they do turn out to be interdefinable with the help of negation in a manner that exactly parallels the interdefinability of the dual modal operators ' \Diamond ' and ' \Box '.

Of course, this is by no means enough to confirm Evans's speculation, in his final remarks, that his two operators 'generate a modal logic as strong as S_5 '. So we are not in a position to endorse his attempt to turn his derivation of (5) from (1) and (3) into an argument with a conclusion that is 'straightforwardly inconsistent with (1)', namely (5S), by 'strengthening' (1)-(4) and Leibniz's law with the prefix ' Δ '. However, it also appears that nothing so ambitious as this is needed in order to turn his derivation of (5)from (1) and (3) into a formal reductio ad absurdum proof, given the interpretation of the operators ' ∇ ' and ' Δ ' now being proposed. For it appears that on this interpretation we can simply *extend* the existing derivation of (5) from (1) and (3) by going on to derive (5S) directly from (5). Recall once more that, as we are now proposing to interpret it, ' Δ ' may be read as 'it is not indeterminate whether (it is true that)', or equivalently as 'there is a fact of the matter as to whether'. Now, for the purposes of *reductio*, (1) is assumed to be true. As Evans himself says, it supposedly 'reports a fact'. And we may agree with Evans that premise (3) is trueindeed, that it is logically true. But if the derivation of (5) from these is valid, then it is *truth-preserving*, so that if (1) and (3) are true, so too is (5). But if (5) is true, then it is true and so a *fact* that *a* is not identical to *b*, in which case there *is* a fact of the matter as to whether a is not identical to b: which is what (5S) says. So we may extend Evans's original argument by deriving (5S) directly from (5). To be sure, to call (5S) a *strengthening* of (5), given our proposed reading of Evans's sentential operator ' Δ ', would be highly misleading. For on this interpretation it is not the case that (5S) entails but is not entailed by (5), and so (5S) is not in this sense 'stronger than' (5). The question at issue now, however, is whether the original derivation of (5) from (1) and (3) may legitimately be turned into a derivation of (5S) from (1) and (3), given the proposed interpretation of the operator ' Δ '—and it seems clear enough that it can. And then all that is further needed in order to turn Evans's original argument into a formal *reductio* of (1), on this interpretation, is the interdefinability of ' ∇ ' and ' Δ ' that we established earlier, for this allows us to derive the negation of (1) from (5S).

Let me now briefly sum up my conclusions so far. I consider that Evans's sentential operator ' ∇ ' can and should be interpreted as meaning 'it is indeterminate whether (it is true that)', or equivalently as 'there is no fact of the matter as to whether', and that on this interpretation it is, with the help of negation, interdefinable with his other sentential operator, ' Δ ', so that ' ∇p ' is logically equivalent to ' $\sim \Delta \sim p$ '. I also consider that, with ' ∇ ' and ' Δ ' thus interpreted, Evans has no problem in turning his original argument from (1) and (3) to (5) into a formal *reductio ad absurdum* proof of the impossibility of what I call ontic indeterminacy of identity, subject only to the following condition: that his original argument—which I shall henceforth refer to simply as 'Evans's argument'—is itself valid. My contention is that Evans's argument is not in fact valid, and I shall attempt to identify an invalid step in it in due course.

Before I do that, however, I shall explain why I think that we have reason to suspect, even before actually identifying an invalid step in Evans's argument, that it may indeed be invalid. The reason for this is that it appears, on closer inspection, that Evans's argument is subtly *question-begging*. By a 'questionbegging' argument I mean, roughly speaking, one which in some manner already assumes or presupposes something that it is supposed to establish. Now, of course, by no means every questionbegging argument can be convicted of containing an invalid step. An argument for a conclusion p that had p as its only premise would be blatantly question-begging, but it does not contain an invalid step: for p certainly entails p. However, an argument can be more subtly question-begging than this. It may be, for example, that the validity of a step in the argument depends in some way upon something that the argument is supposed to establish—and in this sort of case, its being question-begging may well be indicative of its being invalid. Such I believe to be the case with Evans's argument.

So why do I think that Evans's argument is subtly questionbegging? I have explained this elsewhere (Lowe 1994 and 1998: 63 ff.), but shall do so again now, for the point does not take much space to convey. The crux of Evans's argument is his use of Leibniz's law in an attempt to show that, on the supposition that (1) is true, *a* and *b* differ in their properties and hence that *a* is *not* identical to *b*. The property that *b* is supposed to possess but *a* to lack is symbolized by Evans as $\lambda x [\nabla(x = a)]$ '. That *b* possesses this property is asserted in (2), which is taken to follow from (1). But notice that if it is valid to derive (2) from (1), then it is equally valid, by parity of reasoning, to derive the following from (1):

(2S) $\lambda x [\nabla (x = b)] a$

(2S) asserts that *a* possesses the property of being such that it is indeterminate whether it is identical to b. But now we may ask the following question: is this property, $\lambda x [\nabla(x = b)]$, which has just been attributed to a, the same as or different from the property that was previously attributed to b, $\lambda x [\nabla(x = a)]$? These 'two' properties 'differ' only by permutation of a and b. So it would appear that, on the assumption that it is indeed indeterminate whether a is identical to b, it is by the same token indeterminate whether these properties themselves are identical-and thereby equally indeterminate whether they are different. (Recall that 'It is indeterminate whether p' is equivalent to 'It is indeterminate whether not p'.) But in that case it seems that the most that can be concluded is that it is *indeterminate* whether a and b differ in their properties, and hence not that a is not identical to b, but only that it is *indeterminate* whether a is identical to b—which is just what was originally assumed.

If this diagnosis is correct, Evans's argument is question-begging in the following way. The argument attempts to establish, through an application of Leibniz's law, the non-identity of a and b, by showing that b possesses a property that a lacks. And it attempts to derive this conclusion from the assumption that it is indeterminate whether a is identical to b. However, given that assumption, the very property in respect of which b is supposed to differ from ais one such that it is in fact indeterminate whether it is different from a property that *a* must equally be supposed to possess. Hence, the alleged difference in the properties of a and b, required to establish their non-identity, already presupposes their non-identity and hence cannot be used to establish it. The problem arises, of course, from a special feature of the properties concerned: namely, their *identity-involving* character. The properties in question are $\lambda x [\nabla(x = a)]$ and $\lambda x [\nabla(x = b)]$, which are 'identity-involving' in that each of them involves the identity of an object—a in the one case and b in the other. But since the properties 'differ' only by permutation of a and b, their own identity or distinctness turns entirely on the identity or distinctness of a and b themselves. Hence, if the latter is indeterminate, as has been assumed, so too is the identity or distinctness of these properties indeterminate-the consequence being that Leibniz's law is powerless to distinguish the objects by means of such properties.

However, although this diagnosis calls into question the ability of Evans's argument to establish its intended conclusion, it does not yet show where exactly the argument can be supposed to go wrong. But it will be noticed that in delivering this diagnosis I have said nothing about a crucial step in the argument: namely, the derivation of (4) from (3). (3) seems to be a perfectly uncontentious logical truth, but (4) is the line in which *a* is asserted *not* to possess the property attributed to b in line (2). Now, as we have seen, given that the inference from (1) to (2S) is valid—which it must be if the parallel inference from (1) to (2) is valid—(1)entails that a possesses a property that is not determinately *distinct from* the property that *a* is denied to possess in (4). The two claims (2S) and (4) are clearly in tension with each other, because the first attributes to an object a property that is not determinately distinct from a property that the second denies that object to possess. But that *a* possesses the property attributed to it in (2S) is not an inconsistent claim in itself, and cannot be inconsistent with the trivial logical truth (3). Hence, the inference from (3) to (4) must generate a tension between (2S) and (4) that did not exist between (2S) and (3). And this implies that the claim made in (4) goes beyond anything entailed by (3). I conclude that if Evans's argument is invalid, the most plausible place to locate its invalidity lies in the inference from (3) to (4). However, this is all that I shall say about the matter for the time being, because I want to turn now to the Barcan–Kripke proof of the necessity of identity. My intention is to show that a difficulty in that proof can be seen to parallel the difficulty in Evans's argument and confirm the source of the latter's invalidity.

As I remarked earlier, the true parallel is not exactly between Evans's argument and the Barcan-Kripke proof of the necessity of identity (for which see, e.g., Kripke 1971), but rather between the former and a closely related proof of the non-contingency of identity (on this parallel, cf. Keefe 1995). To say that objects a and b are *contingently* identical is to say that they are identical but might have been non-identical. This is a supposition that one might attempt to reduce to absurdity by means of an argument formally paralleling Evans's, simply by reading his operator ' ∇ ' as meaning 'it is contingent that', on the understanding that 'It is contingent that p' is equivalent to 'p and possibly not p'. Thus reinterpreted, Evans's argument may be paraphrased as follows. Suppose that (1) it is contingent that a is identical to b. Then it follows that (2) bpossesses the property of being such that it is contingently identical to a. However, (3) it is not contingent that a is identical to a. And from this it follows that (4) a does not possess the property of being such that it is contingently identical to *a*. But from (2) and (4) it follows by Leibniz's law that a is not identical to b, which contradicts our initial assumption that *a* is contingently identical to b (recalling here, once more, that 'a is contingently identical to b' means 'a is identical to b but a might not have been identical to b').

One thing to notice about this argument for the non-contingency of identity (hereafter 'NCI') is that it does not need to be supplemented in the way that Evans's argument had to be in order to turn the latter into a formal *reductio ad absurdum* proof, because when Evans's operator ' ∇ ' is read as 'it is contingent that', the conclusion (5) directly contradicts the assumption (1). Although the arguments are formally indistinguishable, then, their status as formal proofs is not the same.

Another important difference is that the charge of questionbegging that I raised against Evans's argument cannot apparently be transmuted into an exactly parallel charge against the argument for NCI. It is true that, with the latter argument as with Evans's, if (2) follows from (1), then, by parity of reasoning, so does (2S) follow from (1). In the argument for NCI, (2S) is interpreted as asserting that a possesses the property of being such that it is contingently identical to b. This property and b's supposed property of being such that it is contingently identical to *a* differ only by permutation of *a* and *b*. None the less, on the assumption that *a* and b are only contingently identical, it seems that the property of being contingently identical to *a* is indeed a different property from the property of being contingently identical to b, because it seems that something could possess one of these properties without necessarily possessing the other. (Consider, for example, an object that is identical to a in one of the worlds in which a exists and is not identical to b.) Now, I objected against Evans's argument that, in inferring that a and b differ in their properties, it already presupposed the non-identity of a and b, which it was meant to establish. But, apparently, it cannot in like manner be objected against the argument for NCI that, in inferring that a and b differ in their properties, it already presupposes what it is meant to establish: namely, that a is not contingently identical to b. For, as we have just seen, it seems that on the assumption that a is contingently identical to b, the properties in respect of which a and bdiffer according to the argument for NCI are indeed different properties. If there is a problem with the argument for NCI, then, it cannot be that it is question-begging in exactly the same way that Evans's argument is.

Despite these differences between Evans's argument and the argument for NCI, I believe that we may charge both with committing the same error of formal inference. Indeed, if both are guilty of such an error, it seems that they must be guilty of the same error, given that the arguments are formally indistinguishable. I consider that the error, if error it is, lies in the inference of (4) from (3). Elsewhere (Lowe 1982), I have charged (a version of) the Barcan–Kripke proof of the necessity of identity with an exactly similar logical error. In that case, the erroneous step, as I see it, is the inference of 'a possesses the property of being such that it is necessarily identical to a' from 'It is necessary that a is identical to a'. Since this step—the Barcan–Kripke step, as I shall call it—is much more familiar than, although formally exactly like, the step from

E. J. Lowe

(3) to (4) in Evans's proof and the argument for NCI, let us focus on it for the time being. Now, of course, a general complaint may be raised against the Barcan-Kripke step that it moves from a proposition ontologically committed merely to the existence of a certain object, a, to one ontologically committed in addition to the existence of a certain *property*—and indeed, to what may appear to be a very strange kind of property. However, general complaints of this sort, for what they are worth, do not at present concern me, either with regard to the Barcan-Kripke proof or with regard to Evans's argument and the argument for NCI. I have no hostility towards properties in general, and while I acknowledge that it cannot, on pain of paradox, be contended that every meaningful predicate expresses a property, I think that it would be at best tendentious to respond to the arguments now under consideration by contending that the properties that they invoke simply do not exist. Certainly, if one can find fault with the arguments without needing to deny the existence of the properties, this is a more satisfactory method of rebuttal, not least because it cannot be accused of using a sledge-hammer to crack a nut.

So what, if anything, is wrong with the Barcan-Kripke step, as I call it? The problem as I see it arises as follows. Even if it is conceded that 'It is necessary that *a* is identical to *a*' entails that *a* possesses some corresponding property, it may be disputed what property this is—and, of course, there might be more than one such property. One property that *a* might be thought to possess in virtue of the necessary identity of *a* with *a* is the property of being necessarily identical to itself or, more simply put, *the property of necessary self-identity*. This, clearly, is a property that *a* could share with many other things—plausibly, indeed, it is one which it does and must share with every other thing. Obviously, this is a quite different property from the property of being necessarily identical to a which, it seems evident, *a* alone can possess. The question then is whether *a* may be said to possess the latter property simply in virtue of the fact that it is necessary that *a* is identical to *a*.

To answer this question, we need to think about the grounds of necessary truths. Some necessary truths are grounded purely in the laws of logic, which are themselves necessary truths (cf. Lowe 1998: 13 ff.). An instance of a logical law need not itself qualify as a logical law, but it will inherit the necessity of the law of which it is an instance. The law of the reflexivity of identity—that everything is identical to itself-is a necessary truth. And an instance of the law, such as the singular proposition that a is identical to a, inherits that necessity. Hence, it is necessary that a is identical to a. Against this it may be objected that if a is a contingent being, then a does not exist in every possible world, whence it cannot be true in every possible world that *a* is identical to a. There are various ways to reply to this objection-for instance, by championing a kind of 'free' logic that allows a singular proposition to be true even if its singular terms are 'empty', thus denying that it entails the corresponding existential proposition. According to such an approach, that *a* is identical to *a* may be true even in a possible world in which *a* does not exist, so that even if *a* is a contingent being, it may none the less be affirmed that it is necessary that *a* is identical to *a*. Another strategy is to say that, where a is a contingent being, the proposition that a is identical to a is necessary in a restricted sense: to wit, in the sense that it is true in every possible world in which a exists. But whatever we say, it seems clear that we should say that some sort of necessity attaches to the fact that a is identical to a and that the ground of this necessity lies in the laws of logic.

What is by no means clear, however, is that the fact that a possesses the property of being necessarily identical to *a*—supposing there to be such a fact-is one whose ground could be held to lie solely in the laws of logic. The problem is that it would, it seems, be a substantive *metaphysical* fact of an essentialist character, whereas the laws of logic are properly conceived as being metaphysically neutral. No similar concern attaches to the thought that the laws of logic can ground the fact that a, like anything else, possesses the property of being necessarily self-identical. The laws of logic can ground facts about the properties of individuals, but only, it would seem, facts involving properties that are perfectly general in this way. The putative property of being necessarily identical to a is not, however, a perfectly general property. On the contrary, it is a property that, if it exists, a alone can possess. And the existence of such properties and their attribution to individual objects are matters for metaphysics, not logic. The problem with the Barcan-Kripke step, then, is that it purports to extract a metaphysical fact from a purely logical one.

My principal objection to Evans's argument and the argument for NCI is just the same: that each of them tries to pull a metaphysical

rabbit out of a purely logical hat. This, I think, is what is fundamentally objectionable about the inference from (3) to (4) in each case. At this point it may be protested that there can be nothing logically suspect about that inference because it simply exploits the formal device of so-called *property abstraction*, which is equally at work in the inference from (1) to (2). However, here we may pose a dilemma for the defendants of the arguments. Either property abstraction is simply a notational reformulation, so that $\lambda x[Fx]a$ is just an elaborate way of rewriting 'Fa', or else the property abstract $\lambda x[Fx]$ is seriously intended to denote a property, in a way in which the predicate in 'Fa' need not be supposed to do. It should be borne in mind here, as always, that not every predicate can automatically be taken to denote a property, on pain of contradiction. If so-called property abstraction is not understood necessarily to involve the denotation of a property, then it may indeed be no more than an elaborate rewriting device with a highly misleading name. But in that case lines (2) and (4) of Evans's argument and the argument for NCI are simply superfluous, and we should evaluate the arguments in the form in which they would be left without them. This I shall do in a moment. On the other hand, if property abstraction is understood necessarily to involve the denotation of a property, then neither the inference from (1) to (2) nor the inference from (3) to (4) can be construed as a perfectly innocent logical step that cannot be subject to the sort of objection I raised earlier. And as for my reason for regarding the second of these inferences more critically than the first, that is implicit in what I have already said. The point is that (4) is inferred from what seems to be a purely logical truth, (3), whereas what (2) is inferred from, (1), certainly does not have that status. That we may validly infer one metaphysically laden proposition from another is not problematic, but that we may infer a metaphysically laden proposition from a purely logical truth certainly is.

What, then, about the possibility of simply stripping down Evans's argument and the argument for NCI by removing lines (2) and (4)? The problem, of course, is that the arguments are supposed to involve an application of Leibniz's law, construed as the principle that if an object a is identical to an object b, then a has any property that b has, and vice versa. And this principle cannot be applied unless properties are invoked in the arguments. The best that one can do instead is to invoke the principle of the substitutivity of identity. But how could that possibly work in the case of Evans's argument? How are we supposed to derive (5) directly from (1)and (3) by means of this principle? It might be supposed that we could assert the following as an instance of the principle of the substitutivity of identity

(6)
$$a = b \rightarrow (\sim \nabla (a = a) \rightarrow \sim \nabla (a = b))$$

and contrapose this to give

(7)
$$(\sim \nabla (a = a) \otimes \nabla (a = b)) \rightarrow \sim (a = b)$$

Then, conjoining (1) and (3) applying *modus ponens* to their conjunction and (7), we might suppose that we could detach the consequent of (7), which is (5).

One apparent problem with this strategy is that we seem to be using classical truth-functional operators and classical bivalent logic, when the presence of the indeterminacy operator precludes us from doing that (cf. Parsons 2000: 47). Thus, for example, the contraposition of (6) to give (7) might be called into question. However, interesting though this line of objection may be, it has the drawback that it will appear question-begging to someone who has yet to be persuaded that the notion of ontic indeterminacy of identity is really intelligible. In any case, such an objection would obviously not be appropriate when the operator ' ∇ ' is interpreted as expressing contingency, as in the argument for NCI; so let us consider whether it would be legitimate to reformulate that argument in this stripped-down fashion. And here we may note that, in fact, the Barcan-Kripke argument in its original Kripkean formulation did not make use of property abstraction and proceeded along lines just like those now under consideration (see Kripke 1971: 136).

The answer that I shall give to this query recapitulates one that I have given elsewhere regarding Kripke's original argument for the necessity of identity (see Lowe 1982). In essence, it is this. The principle of the substitutivity of identity is in fact a schema, of the form

(*) $x = y \rightarrow (Fx \rightarrow Fy)$

where the predicate letter 'F' may be uniformly replaced throughout by any predicate and the variables bound by universal quantifiers or replaced by constants to give a logically true formula. In the case of the argument for NCI, the predicate that would need to be substituted for 'F' in (*) to deliver (6) as an instance of the principle is '~ $\nabla(a = \xi)$ ', where ' ξ ' marks an argument-place to be completed by the name of an object. (6) is obtained when 'x' and 'y' are replaced by 'a' and 'b' respectively. However, in order to derive (5) from (6), (1), and (3), we must discern this same predicate as present in (3), on pain of falling foul of a fallacy of equivocation. Now, of course, it is an article of faith of Fregean semantics that a proposition like (3), ' $\nabla(a = a)$ ', may be 'carved up' in different ways without this implying that it involves any kind of ambiguity. Thus it is assumed that (3) may equally well be characterized as saying of a that it is not contingent that a is identical to it and as saying of *a* that it is not contingent that it is identical to itself. However, that these really are just two ways of saying exactly the same thing is not, I think, as uncontentious as the Fregean orthodoxy assumes it to be. (I don't mean to imply that a broadly Fregean approach couldn't resist this assumption, only that it hasn't in fact been resisted in neo-Fregean circles.)

Even if we set aside the question of whether the predicates now at issue denote properties, it is clear that these predicates have different meanings-'is not contingently self-identical' and 'is not contingently identical to a' certainly do not mean the same, and so it is at least questionable whether, when they are predicates of the same subject term, 'a', the sentences thus formed have exactly the same meaning. When two expressions with different meanings are each combined with another univocal expression, to form in each case a meaningful sentence, it would seem surprising that this could result in their forming sentences with exactly the same meaning. It is certainly not obvious that 'a is not contingently self-identical' and 'a is not contingently identical to a' are synonymous, but both of these English sentences are supposed to be representable by the same symbolic formula, " $\nabla(a = a)$ ", which is assumed to be univocal. And the closest English equivalent to this formula, 'It is not contingent that *a* is identical to *a*', is assumed just to be another way of saying exactly the same thing. But all of this, it seems to me, is entirely open to debate. Indeed, returning to the business about 'property abstraction', it seems to me that one way of construing this technical device is precisely as a means of predicate disambiguation, rather than a means of denoting properties. The idea would be that a formula like ' $\nabla(a = a)$ ' is ambiguous, because it can be parsed as resulting from the combination of the name 'a' with either of two different predicates, with one parsing being read as ' $\lambda x [\nabla(x = x)]a$ ' and the other as ' $\lambda x [\nabla(a = x)]a$ '. The whole point of avoiding ambiguity in formal logic is that in such logic there should be a one-to-one correspondence between meaning and form, so that valid inferences can be identified as such purely in virtue of their form. The upshot of all this is that the 'stripped-down' version of the argument for NCI, invoking the principle of the substitutivity of identity in place of Leibniz's law, may be accused of involving a fallacy of equivocation which arises from an insufficiently perspicuous logical syntax.

Let me make it clear exactly what, on this construal, is objectionable about the 'stripped-down' versions of the arguments for NCI and against the indeterminacy of identity. The objection to the argument for NCI is that in order for the conclusion (5) to be derived from (1) and (3) by means of the principle of the substitutivity of identity, the monadic predicate chosen to replace the schematic letter 'F' in that principle will have to be ' $\nabla(a = \xi)$ ', rather than ' $\nabla(\xi = \xi)$ '. However, (3)'s status as a purely logical truth is plausible only if it is parsed as the result of filling both argument-places of the second of these predicates with the name 'a'—that is, as saying of a that it is not contingent that it is self-identical. Indeed, if (3) is instead parsed as saying of *a* that it is not contingent that it is identical to a-which it needs to be if the argument for NCI is not to involve a fallacy of equivocationthen it appears that the argument turns out to be question-begging in a perfectly straightforward way, because (3) so parsed is effectively nothing less than an assertion of the non-contingency of identity. Recall that *a* here is an arbitrarily chosen object. And what (3) so parsed says of this object-and so, in effect, of any object-is that it is not contingent that it is the very object that it is: in other words, that it could not have been any other object. But this is precisely what the doctrine of the non-contingency of identity amounts to. The alternative parsing of (3) is quite different in its metaphysical import, for on that parsing (3) merely says of any arbitrarily chosen object that it could not have failed to be self-identical. And an exactly parallel objection can be levelled at Evans's argument: namely, that his premiss (3), depending on how it is parsed, is either too weak to sustain his conclusion that identity

E. J. Lowe

is never indeterminate or else implicitly presupposes it. On the innocuous parsing, (3) says of an arbitrarily chosen object that it is not indeterminate whether it is self-identical, whereas on the question-begging parsing it says of an arbitrarily chosen object that it is not indeterminate whether it is just *that* object. But precisely what it *means* to assert that an object may have indeterminate identity is that an object may be such that it is indeterminate whether it is just *that* object, as opposed to another.

I want to draw matters to a close now by mentioning a putative example of ontically indeterminate identity that I have championed before (see Lowe 1994 and Lowe 1998: 62 ff.)-partly in order to illustrate my earlier point that such indeterminacy need have nothing to do with 'fuzzy' spatial or temporal boundaries and partly to provide material for some brief remarks about the notion of 'singular reference'. The example involves the capture of a free electron by a helium ion, which thus comes to have two orbital electrons, one of which is subsequently emitted. Throughout this episode there exist two electrons, neither of which begins or ceases to exist during the period of time involved. But, I maintain, there is no fact of the matter as to whether the electron that is emitted is identical to the electron that was captured. This is because, during the period in which both electrons are orbiting the helium nucleus, they are in a state of so-called quantum superposition. During this time, there are certainly two electrons orbiting the nucleus, each with a spin in a direction opposite to that of the other: but there is, it seems, no fact of the matter as to which electron has the spin in one of the directions and *which* has the spin in the other direction. In fact, nothing whatever differentiates one of the electrons from the other during this time. Suppose, now, that we call the captured electron 'a' and the emitted electron 'b'. Then the claim is that there is no fact of the matter as to whether a is identical to b.

One might imagine that there are in fact two alternative possible courses of events in this scenario. The first is that the captured electron continues to orbit the nucleus and the electron that was previously orbiting it is later emitted. The second is that the captured electron is later emitted and the electron that was previously orbiting the nucleus continues to do so. But the claim is that no fact of the matter can distinguish between these supposedly different courses of events. By this I mean not just that we cannot possibly *tell* which course of events actually occurred, but that it is a misconception to think that there really are these two distinct possibilities. The facts of the matter just amount to this and not more than this: that one electron was captured, two electrons orbited the nucleus for a while, and then one electron was emitted. There is simply no further fact of the matter as to the identity or distinctness of the captured electron and the emitted electron. That this is the proper way to characterize the situation seems to me to be not only perfectly intelligible but also almost certainly correct. If Evans's argument were correct, this could not be so. But now I think we have good reason not only to reject Evans's argument as fallacious, but also to reject the thesis that it is supposed to prove that ontic indeterminacy of identity is impossible. It is not only possible, but also very plausibly exemplified in the domain of sub-atomic particles.

Here, however, the following complaint may be raised. It may be urged that if one is to offer a genuine example of ontically indeterminate identity, then it is important that the singular terms employed—in this case, the names 'a' and 'b'—are not terms whose references are vague. They must be 'precise' designators, for if they are not, then it would appear that we are merely dealing with a case of *semantic* vagueness, not genuine ontic indeterminacy of identity. But is it not the case, in the foregoing example, that the names 'a' and 'b', introduced as names of the captured electron and the emitted electron respectively, are vague rather than precise designators? For isn't it the case that the manner in which these terms have been introduced leaves it indeterminate whether 'a' applies to the emitted electron and 'b' applies to the captured electron? So isn't it just this indeterminacy of *reference* that leaves it indeterminate whether the sentence 'a is identical to b' is true?

I think that this line of objection is entirely misplaced. It would not be misplaced if it were *correct* to suppose that there really are two distinct possible courses of events in the scenario, as outlined earlier. For in that case we could quite properly say that the name 'a' has been introduced in such a fashion that it is left undetermined whether it refers (I) to an electron that is captured and thereafter continues to orbit the nucleus, or (2) to an electron that is captured and is thereafter emitted, or indeed (3) to an electron that is captured and to *another* electron that is later emitted—and similarly with regard to the name 'b'. But my claim is that *there simply are no distinct possibilities* of the sort now being suggested.

E. J. Lowe

To suppose that there are is precisely to suppose that the example under discussion does *not* involve genuine ontic indeterminacy of identity—and as such entirely begs the question at issue. In other words, only if it is already *assumed* that the example does not really involve ontic indeterminacy of identity can it be classified as a case of semantic vagueness arising from our failure to fix precisely the references of the names involved. If I am right, we simply *couldn't* fix the references of these names any more 'precisely', because the facts themselves don't admit the distinctions that would be required for this.

The lesson is that some singular terms may necessarily fail to make determinately identifying reference. In our example, the name 'a' and the definite description 'the captured electron' are such terms. But this is not to say that they are 'vague designators' in the sense required by the preceding line of objection, for a vague designator in that sense is a singular term whose reference could be made determinate in principle, or which in other words is capable of 'precisification'. We might, of course, still call them 'vague designators' in another sense-implying thereby simply that statements containing them may be of indeterminate truth-value, without any presumption that their references could be precisified so as to eliminate such indeterminacy. (On these contrasting conceptions of a vague singular term, cf. Keefe 2000: 159-60.) It would be improper to complain, then, that my proposed counterexample to Evans's thesis defeats itself by turning into a harmless case of semantic vagueness, because it can only be seen in that light if it is already presumed that ontic indeterminacy of identity is not involved in the case. And it would be similarly questionbegging, of course, to raise a similar complaint in defence of the argument for NCI, by invoking the distinction between 'rigid' and 'non-rigid' designators. Both complaints attempt to rebut a metaphysical thesis by semantic sleight of hand. As such, they repeat the original error of Evans's argument and the parallel argument for NCI: the error of trying to establish substantive metaphysical claims by means of purely logical argument. Even so, because philosophy advances only through an appreciation of its past errors, we have much reason to be grateful to Evans for his classic one-page paper.

REFERENCES

- Allison, H. (1983), *Kant's Transcendental Idealism: An Interpretation and Defense* (New Haven and London: Yale University Press).
- Anscombe, G. E. M. (1975), 'The first person', in S. Guttenplan (ed.), *Mind and Language* (Oxford: Clarendon Press), 45–65.
- Aristotle (1984), On the Soul, in J. Barnes (ed.), The Complete Works of Aristotle (Princeton: Princeton University Press).
- Ayers, M. (1991), *Locke* (London: Routledge).
- Barwise, J., and Perry, J. (1983), *Situations and Attitudes* (Cambridge, Mass., and London: MIT Press).
- Bell, D. (1979), *Frege's Theory of Judgement* (Oxford: Oxford University Press).
- ----- (1990a), 'How 'Russellian' was Frege?', Mind, 99: 267-77.
- ----- (1990*b*), *Husserl* (London: Routledge).
- Bennett, J. (1971), Locke, Berkeley, Hume: Central Themes (Oxford: Oxford University Press).
- Bermúdez, J. L. (1998), *The Paradox of Self-Consciousness* (Cambridge, Mass.: MIT Press).
- (2002), 'Sources of self-consciousness', *Proceedings of the Aristotelian Society*, 102: 87–107.
- (2003), 'The elusiveness thesis, immunity to error through misidentification, and privileged access', in B. Gertler (ed.), *Privileged Access* (Aldershot: Ashgate), 213–31.
- Boolos, G. S. (1984), 'To be is to be the value of a variable (or some values of some variables)', *Journal of Philosophy*, 81: 430–49. Page references are to Boolos 1998: 54–87.
- (1985*a*), 'Nominalist platonism', *Philosophical Review*, 94: 327–44. Page references are to Boolos 1998: 73–87.

----- (1985b), 'Reading the Begriffsschrift', *Mind*, 94: 331-44. Page references are to Boolos 1998: 155-70.

- (1998), Logic, Logic, and Logic (Cambridge, Mass.: Harvard University Press).
- Bostock, D. (1988), 'Necessary truth and a priori truth', *Mind*, 97: 343-79.
- Bricker, P. (1989), 'Quantified modal logic and the plural *de re*', *Midwest Studies in Philosophy*, 14: 372–94.
- Bruce, V., Green, P. R., and Georgeson, M. A. (1996), *Visual Perception: Physiology, Psychology and Ecology*, 3rd edn (Hove, East Sussex: Psychology Press).

- Burge, T. (1974), 'Truth and singular terms', Noûs, 8: 309-25.
- —— (1977), 'A theory of aggregates', *Noûs*, 11: 97–117.
- Campbell, J. (1993), 'The role of physical objects in spatial thinking', in N. Eilan, R. McCarthy, and B. Brewer (eds.), *Spatial Representation* (Oxford: Basil Blackwell), 65–95.
- ----- (1994), Past, Space and Self (Cambridge, Mass.: MIT Press).
- (1997), 'Sense, reference and selective attention', *Proceedings of the Aristotelian Society*, suppl. vol. 71: 55–74.
- ----- (2002), Reference and Consciousness (Oxford: Clarendon Press).
- Carl, W. (1994), Frege's Theory of Sense and Reference: Its Origins and Scope (Cambridge: Cambridge University Press).
- Cartwright, R. L. (1994), 'Speaking of everything', Noûs 28: 1-20.
- Cassam, Q. (1994) (ed.), *Self-Knowledge* (Oxford: Oxford University Press).
- (1995), 'Introspection and bodily self-ascription', in J. L. Bermúdez,
 A. J. Marcel, and N. Eilan (eds.), *The Body and the Self* (Cambridge, Mass.: MIT Press), 311–26.
- ----- (1997), Self and World (Oxford: Oxford University Press).
- (1998), 'Mind, knowledge and reality: Themes from Kant', in A. O'Hear (ed.), *Current Issues in Philosophy of Mind* (Cambridge: Cambridge University Press), 321–48.
- (1999), 'Self-directed transcendental arguments', in R. Stern (ed.), *Transcendental Arguments: Problems and Prospects* (Oxford: Clarendon Press), 83–110.
- Caston, V. (2002), 'Aristotle on consciousness', Mind, 111: 751-815.

Chierchia, Gennaro (1989), 'Anaphora and attitudes *de se*', in R. Bartsch, J. van Benthem, and P. van Emde Boas (eds.), *Semantics and Contextual Expression* (Dordrecht: Foris), 1-31.

- Chomsky, Noam (1971), *Problems of Knowledge and Freedom* (New York: Random House).
- (1976*a*), 'Conditions on rules of grammar', *Linguistic Analysis*, 2: 303–51.
- (1976b), Knowledge of Language (New York: Pantheon).
- ----- (1981), Lectures on Government and Binding (Dordrecht: Foris).
- Coltheart, Max (1999), 'Modularity and cognition', *Trends in Cognitive Sciences*, 3: 115–20.
- Davidson, D. (1967), 'Truth and meaning', Synthese 7: 304-23; repr. in *Inquiries into Truth and Interpretation* (Oxford: Clarendon Press, 1984), 17-36.
- (1970), 'Mental events', in L. Foster and J. Swanson (eds.), *Experience and Theory* (Amherst, Mass.: University of Massachusetts Press), 79–101; repr. in *Essays on Actions and Events* (Oxford: Clarendon Press, 1980), 207–25.

- and Harman, G. (1972) (eds.), *Semantics of Natural Language* (Dordrecht: Reidel).
- Donnellan, K. S. (1966), 'Reference and definite descriptions', *Philosophical Review*, 75: 281–304.
- Dretske, F. (1981), *Knowledge and the Flow of Information* (Cambridge, Mass.: MIT Press).
- Dummett, M. (1973), *Frege: Philosophy of Language* (London: Duckworth).
- (1976), 'What is a theory of meaning? II', in Evans and McDowell 1976: 67-137.
- (1981), Truth and Other Enigmas (London: Duckworth).
- (1991a), Frege: Philosophy of Mathematics (London: Duckworth).
- ----- (1991*b*), 'The relative priority of thought and language', in *Frege* and Other Philosophers (Oxford: Oxford University Press), 315-24.
- ----- (1994), Origins of Analytical Philosophy (London: Duckworth).
- Elbourne, Paul (2001), 'E-type anaphora as NP-deletion', *Natural Language Semantics*, 9: 241–88.
- Evans, G. (1973), 'The causal theory of names', *Proceedings of the Aristotelian Society*, Suppl. vol. 47: 187–208. Page references are to Evans 1985*a*: 1–24.
- (1975), 'Identity and predication', *Journal of Philosophy*, 72: 343–63. Page references are to Evans 1985*a*: 25–48.
- (1976), 'Semantic structure and logical form', in Evans and McDowell 1976: 199–222. Page references are to Evans 1985*a*: 49–75.
- (1977), 'Pronouns, quantifiers, and relative clauses I–II', *Canadian Journal of Philosophy*, 7 (3): 467–536 and 7 (4): 777–97. Page references are to Evans 1985*a*: 76–175.
- ----- (1978), 'Can there be vague objects?', *Analysis*, 38: 208. Page references are to Evans 1985*a*: 176–7.
- ----- (1979), 'Reference and contingency', *The Monist*, 62 (2): 161-89. Page references are to Evans 1985*a*: 178-213.
- (1980), 'Pronouns', *Linguistic Inquiry*, 11: 337–62. Page references are to Evans 1985*a*: 214–48.
- (1980a), 'Things without the mind: a commentary upon chapter two of Strawson's *Individuals*', in Z. Van Straaten (ed.), *Philosophical Subjects: Essays Presented to P. F. Strawson* (Oxford: Clarendon Press), 76–116. Page references are to Evans 1985*a*: 240–90.
- (1981*a*), 'Semantic theory and tacit knowledge', in Holtzman and Leich 1981: 118–37. Page references are to Evans 1985*a*: 322–42.
- (1981*b*), 'Understanding demonstratives', in H. Parret and J. Bouveresse (eds.), *Meaning and Understanding* (Berlin: W. de Gruyter) 280–303. Page references are to Evans 1985*a*: 291–321.

- Evans, G. (1982), The Varieties of Reference (Oxford: Clarendon Press).
- —— (1985*a*), *Collected Papers* (Oxford: Clarendon Press).
- —— (1985*b*), 'Molyneux's question', in Evans 1985*a*: 362–99.
- and McDowell, J. (1976) (eds.), *Truth and Meaning: Essays in Semantics* (Oxford: Oxford University Press).
- Falkenstein, L. (1995), *Kant's Intuitionism: A Commentary on the Transcendental Aesthetic* (Toronto: University of Toronto Press).
- Fiengo, R., and May, R. (1994), *Indices and Identity* (Cambridge, Mass.: MIT Press).
- Fodor, J. (1983), The Modularity of Mind (Cambridge, Mass.: MIT Press).
- Fox, D. (1998), 'Locality in variable binding', in P. Barbosa, D. Fox, P. Hagstrom, M. McGinnis, and David Pesetsky (eds.), *Is the Best Good Enough*? (Cambridge, Mass.: MIT Press), 129–55.
- Frege, G. (1884), Die Grundlagen der Arithmetik (Breslau: Koebner).
- (1892), 'On sense and reference', in P. Geach and M. Black, *Translations from the Philosophical Writings of Gottlob Frege* (Oxford: Blackwell, 1960), 56–78.
- ----- (1903), Grundgesetze der Arithmetik II (Jena: Pohle).
- ----- (1914), 'Logic', in Frege 1969: 126-51.
- (1918), 'The thought: a logical inquiry', in P. F. Strawson (ed.), *Philosophical Logic* (Oxford: Oxford University Press, 1966), 17-38.
- (1969), *Posthumous Writings*, ed. H. Hermes, F. Kambartel, and F. Kaulbach; page references are to the English trans. by P. Long and White (Oxford: Basil Blackwell, 1979).
- Furth, M. (1968), 'Two types of designation', in N. Rescher (ed.), *Studies in Logical Theory*, American Philosophical Quarterly Monograph, 2: 9–45.
- Garnham, A. (1989), 'A unified theory of the meaning of some spatial relational terms', *Cognition* 31: 45-60.
- Geach, P. T. (1962), *Reference and Generality* (Ithaca, NY: Cornell University Press).
- Gibson, J. J. (1979), *The Ecological Approach to Visual Perception* (Hillsdale, NJ: Erlbaum).
- Goodman, N. (1966), *The Structure of Appearance*, 2nd edn. (Cambridge, Mass.: Harvard University Press).
- Gopnik, A., Meltzoff, A., and Kuhl, P. (1999), *How Babies Think: The Science of Childhood* (London: Weidenfeld & Nicolson). Also published as *The Scientist in the Crib* (New York: Harper Collins Publishers, 2000).
- Gregory, R. (1970), *The Intelligent Eye* (London: Weidenfeld & Nicolson). — (1974), *Concepts and Mechanisms of Perception* (London: Duckworth).

- Grice, H. P. (1962), 'Some remarks about the senses', in R. J. Butler (ed.), *Analytical Philosophy* (Oxford: Blackwell), 133–53.
- Grodzinsky, Y., and Reinhart, T. (1993), 'The innateness of binding and coreference', *Linguistic Inquiry*, 24: 69–102.
- Gunther, Y. (2003) (ed.), *Essays on Nonconceptual Content* (Cambridge, Mass.: MIT Press).
- Guyer, P. (1987), *Kant and the Claims of Knowledge* (Cambridge: Cambridge University Press).
- Heim, Irene (1993), 'Anaphora and semantic interpretation: a reinterpretation of Reinhart's approach', SfS-Report-07-93 (Tübingen: University of Tübingen).
- Higginbotham, J. (1983), 'Logical form, binding and nominals', *Linguistic Inquiry*, 14: 395–420.

—— (1985), 'On semantics', *Linguistic Inquiry*, 16: 547–93.

- Holtzman, S., and Leich, C. (1981), Wittgenstein: To Follow a Rule (London: Routledge & Kegan Paul).
- Hornstein, Norbert (2001), Move! A Minimalist Theory of Construal (Oxford and Malden, Mass.: Blackwell).
- Huang, C.-T. J., and Luther Liu, C.-S. (2001), 'Logophoricity, attitudes and *ziji* at the interface', in P. Cole, G. Hermon, and C.-T. James Huang (eds.), *Syntax and Semantics*, *33: Long Distance Reflexives* (New York: Academic Press), 141–95.
- Jackendoff, R. (1992), 'Madame Tussaud meets the binding theory', Natural Language and Linguistic Theory, 10: 1–31.
- (1997), The Architecture of the Language Faculty (Cambridge, Mass.: MIT Press).
- Kant, I. (1787), *Critique of Pure Reason*, trans. N. Kemp Smith (London: Macmillan, 1929).
- Kaplan, D. (1979), 'On the logic of demonstratives', *Journal of Philosophical Logic*, 8: 81–98.
- (1989), 'Demonstratives', in J. Almog *et al.* (eds.), *Themes from Kaplan* (Oxford: Oxford University Press), 481-563.
- Kayne, R. S. (2002), 'Pronouns and their antecedents', in S. D. Epstein and T. Daniel Seely (eds.), *Derivation and Explanation in the Minimalist Program* (Oxford and Malden, Mass.: Blackwell), 133–66.
- Keefe, R. (1995), 'Contingent identity and vague identity', *Analysis*, 55: 183–90.

----- (2000), *Theories of Vagueness* (Cambridge: Cambridge University Press).

- Koopman, H., and Sportiche, D. (1981), 'Variables and the bijection principle', in *Linguistic Review*, 2: 139–60.
- Kripke, S. (1971), 'Identity and necessity', in M. K. Munitz (ed.), *Identity and Individuation* (New York: New York University Press), 135–64.

References

Kripke, S. (1980), Naming and Necessity (Oxford: Blackwell).

- Lakoff, G. (1968), 'Counterparts, or the problem of reference in transformation grammar', MS, Harvard University; distributed by the Indiana University Linguistics Club.
- (1996), 'Sorry, I'm not myself today: The metaphor system for conceptualizing the self', in G. Fauconnier and E. Sweetser (eds.), *Spaces, Worlds, and Grammar* (Chicago: University of Chicago Press), 91–123.
- Langacker, R. (1969), 'On pronominalization and the chain of command', in W. Reibel and S. Schane (eds.), *Modern Studies in English* (Englewood Cliffs, NJ: Prentice-Hall), 160–86.
- Larson, R., and Segal, G. (1995), *Knowledge of Meaning: An Introduction to Semantic Theory* (Cambridge, Mass.: MIT Press).
- Lasnik, H. (1976), 'Remarks on coreference', *Linguistic Analysis*, 2: 1–22; repr. in Lasnik 1989: 90–109.

----- (1989), Essays on Anaphora (Dordrecht: Kluwer).

- Lees, R. B., and Klima, E. S. (1963), 'Rules for English pronominalization', *Language*, 39: 17–28.
- Leibniz, G. W. (1982), *New Essays on Human Understanding*, trans. P. Remnant and J. Bennett (Cambridge: Cambridge University Press).
- Levinson, S. (1987), 'Pragmatics and the grammar of anaphora: A partial pragmatic reduction of binding and control phenomena', *Journal of Linguistics*, 23: 379–434.

— (1991), 'Pragmatic reduction of the binding conditions revisited', *Journal of Linguistics*, 27: 107–62.

- Lewis, D. (1979), 'Attitudes *de dicto* and *de se*', *Philosophical Review*, 88: 513-43.
- (1988), 'Vague identity: Evans misunderstood', *Analysis*, 48: 128-30.
- —— (1991), Parts of Classes (Oxford: Blackwell).
- Locke, J. (1975), *An Essay Concerning Human Understanding*, ed. P. H. Nidditch (Oxford: Oxford University Press).
- Logan, Gordon (1995), 'Linguistic and conceptual control of visual spatial attention', *Cognitive Psychology*, 28: 103–74.
- Lowe, E. J. (1982), 'On the alleged necessity of true identity statements', *Mind*, 91: 579-84.
- (1994), 'Vague identity and quantum indeterminacy', *Analysis*, 54: 110–14.
- (1998), *The Possibility of Metaphysics: Substance, Identity and Time* (Oxford: Clarendon Press).
- McDowell, J. (1977), 'On the sense and reference of a proper name', *Mind*, 86: 159-85.
- McGinn, C. (1982), 'The structure of content', in A. Woodfield (ed.), *Thought and Object* (Oxford: Clarendon Press), 207–58.

- Martin, M. G. F. (1997), 'The shallows of the mind', *Proceedings of the Aristotelian Society*, suppl. vol. 71: 75–98.
- Masangkay, Z., McCluskey, K., McIntyre, C., Sims-Knight, J., Vaughn, B., and Flavell, J. (1974), 'The early development of inferences about the visual percepts of others', *Child Development*, 45: 357–66.
- Meltzoff, A. (2002), 'Elements of a developmental theory of imitation', in A. Meltzoff and W. Prinz (ed.), *The Imitative Mind: Development, Evolution and Brain Bases* (Cambridge: Cambridge University Press), 19-41.
- and Moore, M. K. (1977), 'Imitation of facial and manual gestures by human neonates', *Science*, 198: 179–92.
- Mill, J. S. (1843), A System of Logic (London).
- Milner, A. D., and Goodale, M. A. (1995), *The Visual Brain in Action* (Oxford: Oxford University Press).
- Nudds, M. (2001), 'Experiencing the production of sounds', *European Journal of Philosophy*, 9: 210–29.
- Oliver, A. D., and Smiley, T. J. (2001), 'Strategies for a logic of plurals', *Philosophical Quarterly*, 51: 289–306.
- Parsons, T. (2000), *Indeterminate Identity: Metaphysics and Semantics* (Oxford: Clarendon Press).
- Peacocke, C. (1981), 'Demonstrative thought and psychological explanation', *Synthese*, 49: 187–217.
- (1992), A Study of Concepts (Cambridge, Mass.: MIT Press).
- (1993), 'Intuitive mechanics, psychological reality, and the idea of a material object', in N. Eilan, R. McCarthy, and B. Brewer (eds.), *Spatial Representation* (Oxford: Basil Blackwell), 162–76.
- (1999), Being Known (Oxford: Oxford University Press).
- ---- (2001*a*), 'Does perception have a nonconceptual content?', *Journal* of *Philosophy*, 98: 239–64.
- (2001*b*), 'Theories of concepts: A wider task', in J. Branquinho, J. Saagua, and A. Marques (eds.), *Foundations of Cognitive Science* (Oxford: Oxford University Press), 157–82.
- ----- (2003), 'Postscript', in Gunther 2003: 318-22.
- ----- (2004), The Realm of Reason (Oxford: Oxford University Press).
- (2005), 'Explaining perceptual entitlement', in R. Schantz (ed.), The Challenge of Externalism (Berlin: W. de Gruyter).
- Perry, J. (1977), 'Frege on demonstratives', *Philosophical Review*, 86: 474–97.
- (1979), 'The problem of the essential indexical', *Noûs*, 13: 3-21. Page references are to Cassam 1994: 167-83.
- Postal, P. (1971), Crossover Phenomena (New York: Holt, Rinehart and Winston).
- Prawitz, D. (1965), *Natural Deduction* (Stockholm: Almqvist and Wiksell).

- Prior, A. N. (1955), Formal Logic (Oxford: Clarendon Press).
- Quine, W. V. (1972), 'Methodological reflections on current linguistic theory', in Davidson and Harman 1972: 442–54.
- Rayo, A., and Yablo, S. (1999), 'Nominalism through de-nominalization', *Noûs*, 35 (1): 436–64.
- Reinhart, T. (1983*a*), *Anaphora and Semantic Interpretation* (Chicago: University of Chicago Press).
- (1983*b*), 'Coreference and bound anaphora: A restatement of the anaphora questions', *Linguistics and Philosophy*, 6: 47–88.
- ----- (1999), 'Strategies of anaphora resolution', OTS working paper, Utrecht Institute of Linguistics.
- Reuland, E. (2001), 'Primitives of binding', *Linguistic Inquiry*, 32: 439–92.
- Russell, B. (1905), 'On denoting', *Mind*, 14 (56): 479–93. References to the reprinted version in R. C. Marsh (ed.), *Logic and Knowledge* (London: George Allen and Unwin, 1956), 41–56.
- (1912), *The Problems of Philosophy* (Oxford: Oxford University Press).
- —— (1917), 'Knowledge by acquaintance and knowledge by description', in *Mysticism and Logic* (London: George Allen & Unwin), 152–67.
- and Whitehead, A. N. (1910–13), *Principia Mathematica* (Cambridge: Cambridge University Press).
- Safir, K. (1984), 'Multiple variable binding', *Linguistic Inquiry*, 15: 603-38.
- (2003), 'Anaphors, movement and coconstrual', in L.-O. Delsing,
 C. Falk, G. Josefsson, and H. Sigurdsson (eds.), *Grammar in Focus Grammatik i Fokus*, vol. 2, *Festschrift for Christer Platzack* (Lund: Lund University Press), 283–303.
- (2004*a*), *The Syntax of Anaphora* (Oxford and New York: Oxford University Press).
- (2004*b*), *The Syntax of (In)dependence* (Cambridge, Mass.: MIT Press).
- ----- (2004*c*), 'On person as a model for logophoricity', in A. Akinlabi (ed.), *Proceedings of the Fourth World Congress of African Linguistics* (Cologne: Ruediger Koeppe Verlag).

Sag, Ivan (1976), 'Deletion and logical form' (doctoral dissertation, MIT).

- Sainsbury, R. M. (1998), 'Indexicals and reported speech', in Sainsbury 2002: 137–58.
- (1999), 'Names, fictional names, and "really"', in Sainsbury 2002: 159–80.
- (2002), Departing from Frege: Essays in the Philosophy of Language (London: Routledge).
 - ----- (2005), Reference without Referents (Oxford: Clarendon Press).

- Salmon, N. (1986), Frege's Puzzle (Cambridge, Mass.: MIT Press).
- Schein, B. (1993), Plurals and Events (Cambridge, Mass.: MIT Press).
- Schlenker, P. (2000), 'Propositional attitudes and indexicality: A crosscategorial approach' (revised doctoral dissertation, MIT).
- Searle, J. (1958), 'Proper names', Mind, 67: 166-73.
- (1983), Intentionality: An Essay in the Philosophy of Mind (Cambridge: Cambridge University Press).
- Sharvy, R. (1980), 'A more general theory of definite descriptions', *Philosophical Review*, 89: 607–24.
- Shoemaker, S. (1968), 'Self-reference and self-awareness', Journal of *Philosophy*, 65: 555-67.
- ----- (1970), 'Persons and their pasts', *American Philosophical Quarterly*, 7: 269–85.
- Sloman, A. (1965), 'Functions and rogators', in J. Crossley and M. Dummett (eds.), Formal Systems and Recursive Functions (Amsterdam: North Holland), 55–174.
- Soames, S. (2002), *Beyond Rigidity* (New York: Oxford University Press).
- Stalnaker, R. C. (1995), 'The interaction of modality with quantification and identity', in W. Sinnott-Armstrong *et al.* (eds.), *Modality, Morality, and Belief: Essays in Honour of Ruth Barcan Marcus* (Cambridge: Cambridge University Press), 12–28.
- Stich, S. (1979), 'Beliefs and sub-doxastic states', *Philosophy of Science*, 45 (4): 499–518.
- Strawson, P. F. (1950), 'On referring', Mind, 59: 320-44.
- (1956), 'Construction and analysis', in A. J. Ayer *et al.* (eds.), *The Revolution in Philosophy* (London: Macmillan), 97–110.
- (1959), *Individuals* (London: Methuen).
- (1971), 'Identifying reference and truth-values', in *Logico-Linguistic Papers* (London: Methuen), 75–95.
- ----- (1982), 'Review of Bieri *et al.*, Transcendental Arguments and Science', *Journal of Philosophy*, 79 (1): 45–50.
- (1997), 'Kant's new foundations of metaphysics', in *Entity and Identity and Other Essays* (Oxford: Clarendon Press), 232–43.
- Streri, A. (1993), Seeing, Reaching, Touching: The Relations between Vision and Touch in Infancy (Cambridge, Mass.: MIT Press).
- and Pêcheux, M. G. (1986), 'Cross-modal transfer of form in 5-month-old infants', *British Journal of Developmental Psychology*, 4: 161–7.
- Thompson, J. J. (1974), 'Molyneux's problem', *Journal of Philosophy*, 71: 637–50.
- Warren, D. (1998), 'Kant and the apriority of space', *Philosophical Review*, 107: 179-224.

References

- Wasow, T. (1979), Anaphora in Generative Grammar (Ghent: E. Story-Scientia).
- Wiggins, D. (1975), 'Identity, designation, essentialism and physicalism', *Philosophia*, 5: 1-30.
- Williams, E. (1977), 'Discourse and logical form', *Linguistic Inquiry*, 8: 103-39.
- Williamson, T. (1996), 'The necessity and determinacy of distinctness', in S. Lovibond and S. G. Williams (eds.), *Essays for David Wiggins: Identity, Truth and Value* (Oxford: Blackwell), 1–17.
- Wimmer, H., Hogrefe, J., and Sodian, B. (1988), 'A second stage in children's conception of mental life: Understanding informational accesses as origins of knowledge and belief', in J. Astington, P. Harris, and D. Olson (eds.), *Developing Theories of Mind*, (Cambridge: Cambridge University Press), 173–93.
- Wittgenstein, L. (1958), *Philosophical Investigations*, trans. G. E. M Anscombe, 2nd edn (Oxford: Blackwell).
- Wright, C. (1981), 'Rule-following, objectivity and theory of meaning', in Holtzman and Leich 1981: 99–117.
- Yaniv, I., and Shatz, M. (1988), 'Children's understanding of perceptibility', in J. Astington, P. Harris, and D. Olson (eds.), *Developing Theories* of *Mind* (Cambridge: Cambridge University Press), 93–108.
- Yi, B.-U. (1999), 'Is two a property?, *Journal of Philosophy*, 96: 163–90. — (2002), *Understanding the Many* (London: Routledge).

INDEX

Allison, H. 259 n. 1, 265 n. 4 anaphora see coconstrual and coreference Anscombe, G. E. M. 190 Aristotle 222 asymmetric cross-modal transfer 206 Ayers, M. 179 n. 12 Barcan, R. 2, 40 Barwise, J. 55-6 Bell, D. 10 n. 6, 62 n. 33, 64 n. 36, 93 n. 5 Bennett, J. 216 Bermúdez, J. L. 28–31 binding problem 283-5 Boolos, G. S. 23, 84, 91, 100-4 Bricker, P. 112 Bruce, V. 214 Burge, T. 67, 88–9 Campbell, J. 31-3, 182 n. 14, 193 n. 18, 279 n. 9, 283 cantilevering problem 233-6 Carl, W. 56 n. 23, 61 n. 30, 64 n. 36 Cartwright, R. L. 92 Cassam, Q. 37-9, 179 n. 12, 182 n. 14 Caston, V. 230 n. 8 causal theory of names 5-6, 11-12, 51, 80-1 c-command 25-6, 127, 129, 135; dependency 142; noncoreference 129-31 Chierchia, G. 162 n. 24 Chomsky, N. 128, 130, 131, 135, 138, 144, 153, 155 n. 18, 157, 158

coconstrual 26, 125, 131, 155-62; enforced vs. pragmatic 16, 134, 152-5; see also coreference Coltheart, M. 204 concept of action 251; action awareness 251-2; ascriptions of intentionality 252; perception vs. action cases 253 - 4concept of perception 220-1; developmental phenomena 243-50; first-person concept 226-7; internal and external individuation 229; modal contexts 233-6; possession conditions 224, 228-9; sameness of experiences 234–6; seeing an object 222; seeing-that 222-3. vs. observational concepts 224, 229, 234; see also core rule conceptual scheme 266-88 Core Rule 34, 222-3; actions, See concept of action; epistemic consequences 230-1; extended 224-5; negative component of 223-4; otherascriptions 245, 247-50, see also same rule hypothesis; possession conditions of concepts 224; rationality of self-ascriptions 227-8, 230, 242-3; same rule hypothesis, see same rule hypothesis vs. inside-in rule coreference 8-9, 25, 124-5; covaluation 133, 140, 149-50; cross-sentential

Index

- dependencies 148–52; Dependent reference 126, 131–7, *see also* independence principle and form-tointerpretation principle; expected noncovaluation 132, 138–48; noncoreference 126, 128–30, 135–7, *see also* c-command; rules pronominalization and reflexivization 127–8
- Davidson, D. 2, 12–15, 48 n. 11, 71
- definite descriptions 7, 10; *see also* Russell's theory of descriptions
- demonstratives 6, 70–2; *see also* plural terms
- descriptive names 9–10, 66, 72–3, 77–8, 88; classical logic 19–20, 67; collective names 88; vs. Russellian names, *see* Russellian names
- discriminating knowledge 4, 10, 169–70; demonstrative identification 6, 169–70; descriptive identification 168; perceptual identification 169–70
- Donellan, K. S. 87
- Dretske, F. 17
- Dummett, M. 2–4, 10 n. 6, 23, 96–100, 102, 113–14, 237
- egocentric space 6, 31–3, 192, 195–6; behavioural space 199–201; modularity 214–16; nonconceptual content, *see* Nonconceptual content; shape recognition 213–14; unity of 199–203, 210–11; vs. deictic and intrinsic frames 202, 214–15 Elbourne, P. 154 n. 17

epistemic conditions for objective experience 258; binding problem 283–5; conceptual scheme 266–8; duality requirement 274–5; forms of sensibility 263–4, 267, 282–4; idealism 265, 267; question of foundations 259, 263–6; question of method 259, 264–6; question of necessity 259, 269; realism 285–7

- Falkenstein, L. 282 n. 13
- Fiengo, R. 133 n. 4, 140
- Fodor, J. 204
- form-to-interpretation principle 27, 145-8
- Fox, D. 139 n. 11
- free logic 20, 67–80; *see also* Russellian names and Truth-theories
- Frege, G. 2, 3, 4, 9, 18, 28, 43 n. 2, 46–51, 55, 59–65, 90, 97, 102–4, 166, 178, 182–5 Furth, M. 53 n. 18

Garnham, A. 202 Geach. P. T. 8, 152 generality constraint 209–10 Gibson, J. J. 200 Goodman, N. 89 Gopnik, A. 243 n. 15, 244 n. 18 Gregory, R. 225 n. 3 Grice, H. P. 3, 197 Grodzinsky, Y. 143, 151, 155 n. 18 Gunther, Y 17 n. 8 Guyer, P. 265 n. 4, 286 n. 17 Heim, I. 133 Hero 261 Higginbotham, J. 135, 140

- Hornstein, N. 128 n. 3
- Huang, C.-T. J. 162 n. 24

"I" - sense of, see "I"-thoughts identity 290; metaphysical vs. logical facts 303-4; necessity of (Barcan-Kripke proof) 290, 300-2; non-contingency of 290, 300-2; substitutivity of 305-7 ignorance of sources 244; see also concept of perception; same rule hypothesis immunity to error through misidentification 29-30, 178 independence principle 26-7, 135-6, 143 indexicals 28, 165, 171-2, 183; indexical beliefs 175; see also "I"-thoughts information-processing 197-8; and nonconceptual content 207-11; vs. conceptual content 203-6 interpretational semantics 14, 66 n. 1 intuitive criterion of difference 11, 32, 49, 205-6 "I"-thoughts 28-31, 177-82; action component 180; egocentric space 192; information component 178-9, 187-8; "I"-utterances 189-93; location component 180-2; privacy and objectivity T82-6

Jackendoff, R. 158-9

Kant, I. 37, 49 n. 12, 65, 259, 262–6, 275, 278–81, 284, 286 Kaplan, D. 164, 175 Kayne, R. S. 128 n. 3 Keefe, R. 292, 300 Koopman, H. 148 Kripke, S. 2, 5, 7, 9–10, 11, 40, 50, 121, 169, 300 Lakoff, G. 159, 161 n. 23 Langacker, R. 127 Larson, R. 67 n. 4, 76 n. 10 Lasnik, H. 25, 125-32, 133, 135, 136, 140, 157 Lees, R. B. 127 Leibniz, G. W. 288 Levinson, S. 145 n. 14 Lewis, D. 89 n. 3, 295 Logan, G. 215 Lowe, E. J. 39-41 McDowell, J. 12, 18–19 McGinn, C. 51 n. 17 Martin, M. G. F. 280 n. 10, 285 nn. 15–16, 287 n. 20 Masangkay, Z. 244 nn. 16-17 Meltzoff, A 255 n. 21 Mill, J. S. 90 Milner, A. D. 218 mode of presentation 4, 50; see also sense (Sinn)

Molyneux's question 17, 31, 195-6, 212, 209, 212, 217

Nonconceptual content 16–17, 33; conceptual and information processing content 207–11; egocentric space 203, 206–7; perceptual experience 196–7 Nudds, M. 278 n. 8, 280 n. 10

one-off referential devices 10 outside-in model 251; *see also* core rule

Parsons, T. 294, 305 Peacocke, C. 33–6, 287 Penrose's impossible object 225 perceptual experience 31–5, 194–7; categorical concepts 216–19; concept of, *see* concept of perception; core Rule; extraneous information

236-41; other-ascriptions 245, 247–50, *see also* same rule hypothesis; primacy of veridical cases 228; sameness of 234-6; self-ascriptions of 226, 230, 234, see also core rule; self-ascriptions of experiences vs. self-ascriptions of beliefs 241-2; space, see spatiality thesis; specificity in modality 212-16; vs. conceptual and information processing content 211-12; see also egocentric space, see egocentric space; nonconceptual content, see nonconceptual content Perry, J. 29, 164, 165, 171 Perspective (level 2) errors 244; see also concept of perception; same rule hypothesis photograph model 6, 54 plural terms 22-4; modal contexts 111-16; as predicative terms 96-100; rigidity of collective names 120; rigidity of compound names 117-20; rigidity of plural demonstratives 120–1; second-order quantification 100–11; as semantically singular 89-92; truth theory for 94-6 Postal, P. 128, 143 Pratwitz, D. 105 Prior, A. N. 115 projective (level 1) errors 244; see also concept of perception and same rule hypothesis Pronouns 8, 24-5; e-type pronouns 25, 152-5; proxy

readings of 27

Proper names 3–12; and/vs. definite descriptions 44–5, 50–1, 169; externalism vs. internalism 81–3; non descriptive and non-Russellian 78–83; *see also* descriptive names; Russellian names; sense (*Sinn*) psychologism 48–9

Quine, W. V. O. 15

Rayo, A. 101 Reinhardt, T. 133, 143 n. 14, 149 n. 15, 155 n. 18 Reuland, E. 145 n. 13

rigid designators 118; *see also* plural terms

Rumfitt, I. 22–4

- Russell, B. 4, 7, 42-4, 52, 69 n. 5
- Russell's paradox 89, 91
- Russell's principle 6, 29, 169, 181
- Russell's theory of descriptions 42–6
- Russellian names 4, 18, 42–4, 52, 170–1; and a theory of truth 19–22, 66–70, 72–3; vs. descriptive names 9, 19, 66–7, 77–81

Safir, K. 24–8 Sag, I. 133 Sainsbury, M. 19–22, 45, 46 n. 7, 50 n. 14, 53 n. 18, 58 n. 26, 60 n. 29, 167 n. 4, 185 Salmon, N. 58 n. 24 same rule hypothesis 245–50 Schein, B. 91 Schlenker, P. 161 Searle, J. R. 46 n. 6, 54 n. 20, 78 n. 12 self-consciousness 28 sense (*Sinn*) 2–4, 10, 46–64; descriptivism 50–1, 169;

non-referring expressions 58-64; rationality 19, 47-9, 57; rigid designators 9; see also thoughts Sharvey, R. 87, 99 Shoemaker, S. 29, 178, 179 simulation 248-9 Sloman, A. 114 Smiley, T. 91 Spatiality thesis 37–9, 258–61; argument from perception 38, 271–3; conceptual scheme (CS) 266-8; epistemic conditions, see epistemic conditions for objective experience; forms of sensibility (FS) 263-4, 267, 282-4; perception of particulars (ST_P) 260-4, 278-9; spatial concepts (ST_i) 260-2, 266-8; spatial properties (ST_e) 260-2; substance argument 38, 273-7 Stich. S. 17 Strawson, P. 2, 4-6, 36-7, 258, 261, 265, 266 n. 5, 268 streri, A. 206 subdoxastic states 17 subpersonal information processing 33 symmetry constraint 30, 166, 185 tacit knowledge 15-16, 234-6 theory of mind 35 theory of perception 269-70 theory of truth 12–14, 20–1, 68–83; compositionality 73-4; demonstratives 70-2; derivation of theorems 74-7; plural terms 94-5; Russellian vs. descriptive names 68-70, 72-3 theory of understanding 2-3, 12-13, 14-15

Thompson, J. J. 217 thoughts 18-19, 46-9, 54-8, 166-8; as contents of propositional attitudes 48-9, 166; explanation of behavior 175-7; illusions of 7, 44; rationality 19, 47-9, 57; as senses of sentences 167-8; singular thoughts 46-7, 52-3, see also "I"-thoughts; structure of 16-17; truth-values 54-8, 166-7; vs. senses in indexical expressions 172-5 transcendental 264 translational semantics 12

vagueness 39–41, 308–10; ∇ as a sentential operator 294–6; argument against 292–4; argument against the indeterminacy of identity 304–7; Leibniz's law 298–9; non-contingency of identity, *see* identity; ontic vs. semantic 19, 291; problems with the argument against 298–9, 303–4, 307–8 vision, *see* core Rule; perceptual

experience

Warren, D. 286 n. 19 Wasow, T. 143 weak crossover effect 143 Wiggins, D. 79 Williams, E. 133 Wimmer, H. 244 n. 18 Wittgenstein, L. 235-6 Wright, C. 15

Yaniv, I. 244 n. 16 Yi, B.-U. 101