ROUTLEDGE

Philosophy of Language

Alexander Miller

THIRD EDITION

Philosophy of Language

Philosophy of Language provides a comprehensive, meticulous survey of twentieth-century and contemporary philosophical theories of meaning. Interweaving the historical development of the subject with a thematic overview of the different approaches to meaning, the book provides students with the tools necessary to understand contemporary analytic philosophy. Beginning with a systematic look at Frege's foundational theories on sense and reference, Alexander Miller goes on to offer a clear exposition of the development of subsequent arguments in the philosophy of language. Communicating a sense of active philosophical debate, the author confronts the views of the early theorists, taking in Frege, Russell, and logical positivism and going on to discuss the scepticism of Quine, Kripke, and Wittgenstein. The work of philosophers such as Davidson, Dummett, Searle, Fodor, McGinn, Wright, Grice, and Tarski is also examined in depth.

The third edition has been fully revised for enhanced clarity and includes:

- a short introduction for students, outlining the importance of the philosophy of language and the aims of the book;
- two substantial new sections on Philip Pettit's "ethocentric" account of rule-following and on Hannah Ginsborg's "partial reductionism" about rule-following and meaning;
- the addition of chapter summaries and study questions throughout, designed to promote greater understanding and engagement;
- updated guides to further reading at the end of every chapter.

This well-established and sophisticated introduction to the philosophy of language is an unrivalled guide to one of the liveliest and most challenging areas of philosophy and is suitable for use on undergraduate degrees and in postgraduate study.

Alexander Miller is Professor of Philosophy at the University of Otago, New Zealand. He is author of *Contemporary Metaethics: An Introduction* (Polity Press, second edition, 2013), co-editor of *Rule-Following and Meaning* (Acumen, 2002) and co-editor of *A Companion to the Philosophy of Language* (second edition, Wiley-Blackwell, 2017).

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Philosophy of Language

Third Edition

Alexander Miller

Routledge
Taylor & Francis Group
LONDON AND NEW YORK

Third edition published 2018
by Routledge
2 Park Square, Milton Park, Abingdon, Oxon, OX14 4RN

and by Routledge
711 Third Avenue, New York, NY 10017

Routledge is an imprint of the Taylor & Francis Group, an informa business

© 2018 Alexander Miller

The right of Alexander Miller to be identified as author of this work has been asserted by him in accordance with sections 77 and 78 of the Copyright, Designs and Patents Act 1988.

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

Trademark notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

First edition published by Routledge 1998
Second edition published by Routledge 2007

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Cataloging-in-Publication Data
Names: Miller, Alexander, 1965- author.
Title: Philosophy of language / Alexander Miller.
Description: Third edition. | New York : Routledge, 2018. | Includes bibliographical references and index.
Identifiers: LCCN 2017033619 | ISBN 9780415718950 (hardback : alk. paper) | ISBN 9780415718974 (pbk. : alk. paper) | ISBN 9781351265522 (e-book)
Subjects: LCSH: Language and languages--Philosophy. | Meaning (Philosophy)
Classification: LCC P107 .M547 2018 | DDC 401—dc23
LC record available at https://lccn.loc.gov/2017033619

ISBN: 978-0-415-71895-0 (hbk)
ISBN: 978-0-415-71897-4 (pbk)
ISBN: 978-1-351-26552-2 (ebk)

Typeset in Century Schoolbook and Futura
by Swales & Willis Ltd, Exeter, Devon, UK

Contents

Preface to the first edition *xi*

Preface to the second edition *xiv*

Preface to the third edition *xv*

Acknowledgements, first edition *xvi*

Acknowledgements, second edition *xvii*

Acknowledgements, third edition *xviii*

Introduction **1**

1 Frege: Semantic value and reference **8**

1.1 Frege's logical language 8

1.2 Syntax 14

1.3 Semantics and truth 16

1.4 Sentences and proper names 18

1.5 Function and object 21

1.6 Predicates, connectives, and quantifiers 22

1.7 A semantic theory for a simple language 26

Chapter summary 30

Study questions 30

Further reading 31

2 Frege and Russell: Sense and definite descriptions 34

2.1 The introduction of sense 34

2.2 The nature of sense 39

2.3 The objectivity of sense: Frege's critique of Locke 48

2.4 Four problems with Frege's notion of sense 55

2.5 Kripke on naming and necessity 65

2.6 A theory of sense? 71

2.7 Force and tone 73

2.8 Russell on names and descriptions 77

2.9 Scope distinctions 82

2.10 Russell's attack on sense 85

2.11 Russell on communication 91

2.12 Strawson and Donnellan on referring and definite descriptions 93

2.13 Kripke's causal-historical theory of reference 96

2.14 Appendix: Frege's theses on sense and semantic value 99

Chapter summary 101

Study questions 101

Further reading 102

3 Sense and verificationism: Logical positivism 112

3.1 From the *Tractatus* to the verification principle 112

3.2 The formulation of the verification principle 117

3.3 Foster on the nature of the verification principle 123

3.4 The *a priori* and the linguistic theory of necessity 128

3.5 Carnap on internal and external questions 137

3.6 Logical positivism and ethical language 142

3.7 Moderate holism 145

Chapter summary 146

Study questions 147

Further reading 148

4 Scepticism about sense (I): Quine on analyticity and translation 154

4.1 Quine's attack on the analytic/synthetic distinction: Introduction 155

4.2 The argument of "Two Dogmas" (part I) 155

4.3 Criticism of "Two Dogmas" (part I) 161

4.4 The argument of "Two Dogmas" (part II) 164

4.5 Criticism of "Two Dogmas" (part II) 167

4.6 Quine on the indeterminacy of translation: Introduction 169

4.7 The argument from below 171

4.8 Evans and Hookway on the argument from below 177

4.9 The argument from above 185

4.10 Conclusion 192

Chapter summary 192

Study questions 193

Further reading 193

5 Scepticism about sense (II): Kripke's Wittgenstein and the sceptical paradox **199**

5.1 The sceptical paradox 200

5.2 The sceptical solution and the argument against solitary language 210

5.3 Boghossian's argument against the sceptical solution 212

5.4 Wright's objections to the sceptical solution 217

5.5 Zalabardo's objection to the sceptical solution 219

5.6 The normativity of meaning? 223

5.7 "Factualist" interpretations of Kripke's Wittgenstein 226

Chapter summary 237

Study questions 238

Further reading 238

6 Saving sense: Responses to the sceptical paradox **247**

6.1 Linguistic meaning and mental content 248

6.2 Sophisticated dispositionalism 251

6.3 Lewis-style reductionism and ultra-sophisticated dispositionalism 256

6.4 Fodor's "asymmetric dependency" account of meaning 260

6.5 McGinn on normativity and the ability conception of understanding 265

6.6 Wright's judgement-dependent conception of meaning 270

6.7 Pettit's "ethocentric" account 278

6.8 Wittgenstein's dissolution of the sceptical paradox? 289

6.9 Ginsborg's "partial reductionism" 297

Chapter summary 304

Study questions 304

Further reading 305

7 Sense, intention, and speech-acts: Grice's programme 316

7.1 Homeric struggles: Two approaches to sense 316

7.2 Grice on speaker's-meaning and sentence-meaning 319

7.3 Searle's modifications: Illocutionary and perlocutionary intentions 323

7.4 Objections to Gricean analyses 328

7.5 Response to Blackburn 334

7.6 Strawson on referring revisited 337

Chapter summary 339

Study questions 340

Further reading 340

8 Sense and truth: Tarski and Davidson 344

8.1 Davidson and Frege 345

8.2 Davidson's adequacy conditions for theories of meaning 346

8.3 Intensional and extensional theories of meaning 348

8.4 Extensional adequacy and Tarski's Convention (T) 351

8.5 Tarskian truth-theories 356

8.6 Truth and translation: Two problems for Davidson 363

8.7 Radical interpretation and the principle of charity 366

8.8 Holism and T-theorems 373

8.9 Conclusion: Theories of meaning and natural language 376

Chapter summary 377

Study questions 378

Further reading 378

9 Sense, world, and metaphysics 382

9.1 Realism 383

9.2 Non-cognitivism and the Frege–Geach problem 383

9.3 Realism and verification-transcendent truth 386

9.4 Acquisition, manifestation, and rule-following: The arguments against verification-transcendent truth 389

9.5 Twin-Earth, meaning, mind, and world 399

9.6 Grades of objectivity: Wright on anti-realism 406

9.7 Two threats of quietism 411

Chapter summary 414

Study questions 414

Further reading 415

Bibliography *422*

Index *439*

Preface to the first edition

To the student, philosophy of language can seem a bewilderingly diverse and complex subject. This is not an illusion, since philosophy of language deals with some of the most profound and difficult topics in any area of philosophy. But beneath the diversity and complexity, there is some unity. In this book I have concentrated on exhibiting this unity, in the hope that it might make some of the more profound and difficult questions a little more approachable to the student. I have adopted an approach which is broadly thematic, but also (up to a point) historical. If there are two main themes in twentieth-century philosophy of language, they could perhaps be termed *systematicity* and *scepticism*. Ordinarily, we would say that speakers of a language *understand* the expressions of that language, or *know* their *meanings*. Philosophers have been motivated by a desire to say something systematic about these notions of linguistic understanding, meaning, and knowledge. One way in which this can be done is to give some *informal theory of meaning*: this is a theory which attempts to analyse and elucidate our ordinary, pre-theoretic notion of meaning. In Chapters 1 and 2 we begin with Frege's informal theory of meaning, and his analysis of the intuitive notion of meaning in terms of the notions of sense, semantic value, reference, force, and tone. Another way in which philosophers attempt to say something systematic about the notion of meaning is via the construction of *formal theories of meaning*. A formal theory of meaning is, roughly, a theory which generates, for each sentence of the language under consideration, a theorem

which in some sense or other states the meaning of that sentence. Philosophers have attempted to get clear on the notion of meaning by asking about the nature of such a formal theory. Again, the starting point here is Frege, and in Chapters 1 and 2 we look briefly at a simple example of a Fregean formal theory of meaning. The main notion discussed in the book is that of *sense*. After an extensive discussion of Frege's notion of sense in Chapter 2, we move on in Chapter 3 to look at the logical positivists' views on sense: what constraints are there on the possession of sense? We'll look at the logical positivists' answer to this question, and show how it impacts on issues in *metaphysics*. In Chapters 4 and 5 we look at the second main theme in twentieth-century philosophy of language, that of *scepticism about sense*. Are there facts about meaning, and if there are, how do we know them? We'll look at arguments from Quine and Kripke's Wittgenstein which attempt to argue that there are no facts about meaning, that the notion of meaning, as Kripke puts it, "vanishes into thin air". These attacks on the notion of meaning have been enormously influential, and much of contemporary philosophy of language can be viewed as an attempt to rehabilitate the notion of meaning in the face of these attacks. We look at some of these attempts to rehabilitate the notion of meaning in Chapter 6, and, *inter alia*, show that there are important and close connections between issues in the philosophy of language and issues in the *philosophy of mind*. The question of the relationship between mind and language is discussed further in Chapter 7, when we give a brief, critical account of Grice's attempt to analyse the notion of linguistic meaning in terms of the notion of intention. In Chapter 8, we return to the systematicity theme, and look at Donald Davidson's views on the construction of formal theories of meaning for natural languages. We finish in Chapter 9 by returning to a theme which loomed large in Chapter 3, the relevance of questions about meaning to issues in metaphysics. I try to provide a rough map of the current debate between realism and anti-realism, displaying the relevance to this debate of the issues discussed in the previous chapters.

Obviously, in a book of this length, many important topics in the philosophy of language have had to be ignored, and the discussion

of chosen topics has sometimes had to be drawn to a premature close. I hope, though, that although the map provided in this book is incomplete, it is detailed enough to allow the student undertaking further study to work out where these other topics should be located, and to continue the discussion from where I have left off. Guides to further reading are provided at the end of each chapter. Likewise, it is my hope that teachers of the philosophy of language will be able to use this book in their courses, filling out the map as they go along, according to their own interests in the philosophy of language.

The book has been written to be accessible to second- or third-year undergraduate students, or to anyone with a basic knowledge of the language of elementary logic, such as that taught in first-year university courses. Some knowledge of elementary general philosophy, such as that taught in first-year courses on metaphysics and epistemology, would be useful, though, I hope, not essential. Some parts of the book are more demanding than others. For readers entirely new to the philosophy of language, §3.3, §§5.3–5.5, §§6.3–6.9, and §8.5 could be left out on a first reading, and returned to later. Postgraduates and more advanced undergraduates should note, though, that in many ways §§6.3–6.9 constitute the heart of the book.[1] It is my hope that these sections, and indeed the rest of the book, may also be of use to professional philosophers with an interest in the philosophy of language.

ALEXANDER MILLER
Birmingham March 1997

Note

1 I've altered the section numbers here to take account of the new third edition sections – AM (June 2017).

Preface to the second edition

In this second edition I have added several new sections, cleaned up the original text considerably, and updated the guides to further reading at the end of each chapter. The presentation of Kripke's Wittgenstein, in particular, has been modified to take into account the complexities brought to light by the "factualist" interpretation pioneered by George Wilson and David Davies (although in the end I argue against the factualist interpretation in §5.7). Since the preparation of the first edition, three excellent resources for the philosophy of language have been published: *A Companion to Philosophy of Language* (Oxford: Blackwell 1997), edited by Bob Hale and Crispin Wright, *The Blackwell Guide to the Philosophy of Language* (Oxford: Blackwell 2006), edited by Michael Devitt and Richard Hanley, and *A Handbook of Philosophy of Language* (Oxford: Oxford University Press 2006), edited by Ernest LePore and Barry C. Smith. I mention a few of the constituent articles in these volumes in the Further reading and in the footnotes, but I'd like to take this opportunity to recommend them generally: they are the essential next port-of-call following the present text for all serious students and researchers in the philosophical study of language.

ALEXANDER MILLER
Birmingham January 2007

Preface to the third edition

The main additions to this third edition are the new sections (6.7 and 6.9) discussing the responses to the rule-following paradox that have been developed by Philip Pettit and Hannah Ginsborg. In addition to many small corrections and improvements throughout the text, I have added a brief introduction, as well as chapter summaries and study questions at the end of each chapter. I have also updated the guides to further reading.

I'd like to take this opportunity to enter a plug for the second edition of *A Companion to the Philosophy of Language* (Wiley-Blackwell 2017), co-edited by me, Bob Hale, and Crispin Wright, and now expanded to two volumes. Anyone wishing to continue in the philosophy of language beyond the scope of the present book will find the new edition of the Blackwell *Companion* an excellent place to start.

ALEXANDER MILLER
Dunedin, June 2017

Acknowledgements, first edition

In writing this book, I have had the benefit of many comments on the preliminary draft from a number of colleagues and friends: these have resulted in many improvements and have saved me from many errors. For this, I would like to thank Michael Clark, Nick Dent, John Divers, Jim Edwards, Brian Garrett, Chris Hookway, Iain Law, Greg McCulloch, Duncan McFarland, Elizabeth Mortimore, Stephen Mumford, Philip Pettit, Jim Stuart, Mark Walker, Alan Weir, and Stefan Wilson. Thanks also to the students in my Honours philosophy of language classes of 1994–1997 in Nottingham and Birmingham, who acted as guinea pigs for much of the material. I would also like to thank my series editor, John Shand, for useful comments on the typescript and for encouragement and advice throughout. Thanks also to Mina Gera-Price at UCL for editorial assistance with the preparation of the final version. Most of all, I am grateful to Fisun Güner for her help, encouragement, and tolerance of my bad temper during completion of the book.

Acknowledgements, second edition

In addition to thanking again the colleagues and friends mentioned above, I'd like to thank Tama Coutts, Ed Dain, Koje Tanaka, Alessandra Tanesini, Ieuan Williams, and the students in my philosophy of language classes in Sydney between 2003 and 2005. I'm also grateful to Tony Bruce at Routledge for his enthusiasm for a second edition, and to the five anonymous referees who gave it the go ahead. Back at the University of Birmingham, I'm grateful to the School of Social Sciences for giving me a term's study leave during which the second edition was prepared, and to my colleagues in the Philosophy Department for their support. I'm especially grateful to Janet Elwell in the Philosophy Office for her invaluable assistance. For indispensable help of a philosophical and extra-philosophical nature at key moments since I wrote the first edition, I thank John Divers, Fisun Güner, Bob Kirk, Martin Kusch, Brian Leiter, Philip Pettit, Mr. A. R. Walsh, Alan Weir, Crispin Wright, Mark Walker, and – most of all – Jean Cockram and Rosa Miller.

Acknowledgements, third edition

In addition to thanking again everyone named in the previous acknowledgements, I'd like to thank the five Routledge referees for their feedback on the proposal for a third edition (and for spotting a few gremlins in the earlier editions). Thanks, too, to the University of Otago for a period of study leave in which the third edition was completed, and to the students at Otago who have taken Phil 236/336 and Phil 415/458 in recent years. Special thanks to Philip Pettit for some very stimulating conversations about rule-following during my visit to the Australian National University in May 2017, and thanks to Brian Garrett and the other philosophers who attended my seminar there. I'd also like to thank Peter Hanks for his invitation to participate in a session on rule-following at the Central Division APA Meeting in Kansas City in March 2017, and to Hannah Ginsborg for making the session so productive. I'm also grateful to Joe Mendola, Robert Myers, Olivia Sultanescu, Claudine Verheggen, and seminar audiences at the University of Nebraska, Lincoln and York University, Toronto. Many thanks indeed to Jinho Kang for some very helpful feedback on Chapter 7. I am very grateful to Marguerite Nesling for her expert help in correcting the proofs and constructing the index. As always, my greatest debt is to Jean Cockram and Rosa Miller for their love and support.

Introduction

Language is the main medium through which we express our thoughts, and, more generally, our minds. It is also the main medium through which we attempt to represent the world. Given philosophy's traditional preoccupation with fundamental questions about the nature of mind and world and the relationship between them, it is unsurprising that language has also been a major topic of philosophical concern. Although language has been a concern of philosophers from at least as far back as Plato and Aristotle, with the exception of some brief discussion of John Locke's empiricist theory of meaning, this book has the limited aim of introducing some of the main figures in modern philosophy of language, starting with Gottlob Frege's seminal papers from the early 1890s in which modern – post-Aristotelian – logic is for the first time brought to bear on the study of language. The book is limited along another dimension, in that it is concerned almost exclusively with what is sometimes called "cognitive" meaning, the sort of meaning in virtue of which sentences are capable

of being true or false (what Ayer termed "literal significance"). Apart from a brief mention of Frege's notion of *tone*, we abstract away from whatever differences in "expressive" meaning there are between

(a) Fido is a cur

and

(b) Fido is a dog

and between

(c) May is a Conservative and she is a good loser

and

(d) May is a Conservative but she is a good loser.

Given philosophy's preoccupation with the notion of truth, this limitation should not occasion surprise. So rather than attempt to justify the limitation further, in the remainder of this short introduction I'll try to give a brief indication of issues about cognitive meaning – hereafter simply "meaning" – that the book will cover.

Consider the sentences on the following list:

(1) $68 + 57 = 125$
(2) $68 + 57 = 5$
(3) Jones is a bachelor
(4) Jones is an unmarried male of marriageable age
(5) Grass is green
(6) Snow is white
(6a) Schnee ist Weiss
(6b) La neige est blanche
(7) Reality is One, and not Many
(8) Every even number is the sum of two primes
(9) Julius Caesar sneezed exactly ten times on his tenth birthday
(10) The River Clyde is full of water.

One the face of it, all of these (with the possible exception of (7)) are meaningful, all capable of being true or false. Two central questions in the philosophy of language are:

(A) What is meaning?

and

(B) What makes it the case that a linguistic expression means what it does?

In this book, we'll look at answers to these questions provided by a range of the most important philosophers of language from the end of the nineteenth century onwards. We begin, in Chapters 1 and 2, with Frege's attempt to answer question (A) by distinguishing between sense (*Sinn*), semantic value/reference (*Bedeutung*), force, and tone. Frege's work sets the scene for the rest of the book, starting with Bertrand Russell, who argues that it is a mistake to introduce the notion of sense in addition to that of semantic value/reference, and that theorising about language can (and should) make do with the notion of semantic value/reference.

As noted above, one of the notions traditionally at the heart of philosophical thinking is the notion of truth (and its companion, falsity). These notions can be brought to bear on the sentences on our list. (1), (3), (4), (5), (6), (6a), (6b), and (10) are all true; (2) is false; intuitively (9) and (10) are either true or false, though we don't – and perhaps cannot – know which; philosophical monists will hold that (7) is true, pluralists will hold that it is false, and many of the rest of us will deny that it is either. And a direct connection between the philosophy of truth and the philosophy of language is forged by Frege's influential idea that the sense of a sentence is its truth-condition, and that since understanding a sentence is a matter of knowing its sense, understanding a sentence is a matter of knowing its truth-conditions. Historically, one important school of thought – logical positivism – placed tight constraints on which sentences can legitimately be regarded as having a truth-condition: if a sentence is not analytic (i.e. true (or false) purely in virtue of meaning), it only has a genuine truth-condition if it is in principle possible for us to verify whether or

not that truth-condition obtains. In other words, all sentences that are genuinely meaningful are either analytic or in principle empirically verifiable. On our list, (1) to (4) qualify as meaningful in virtue of their analyticity; (5), (6), (6a), (6b), and (10) qualify as meaningful in virtue of being in principle empirically verifiable; (7) has no meaning; and the status of (8) and (9) is at least unclear. Logical positivism's contribution towards an answer to question (A) is the topic of Chapter 3, where we focus mainly on the work of Alfred Ayer.

Intuitively, (3) and (4) are analytic, true purely in virtue of the meanings of the expressions out of which they are constructed. Frege, Russell, and the logical positivists would have held that (1) is analytic also (and that (2) is analytically false). Moreover, (3) and (4) are *synonymous*, and (6a) and (6b) are correct translations of (6), in that they possess the same meaning. These ideas come under pressure in Chapter 4, where we consider the arguments of one of the twentieth century's great sceptics about the notion of meaning, Willard van Orman Quine. We'll see that, according to Quine, the notions of analyticity and synonymy make no sense, and there is no such thing as correct translation as intuitively conceived. In fact, there is no such thing as meaning, no fact of the matter as to the meanings of the sentences on our list. Scepticism about meaning (or sense) is taken further in Chapter 5, where we look at the sceptical arguments about rule-following and meaning developed in the work of Ludwig Wittgenstein, at least as seen through the eyes of Saul Kripke. Intuitively, given what we mean by the "+" sign and so on, and given the facts of elementary arithmetic, (1) is true and (2) is false. We'll see that the sceptic about rule-following holds that all of the facts about our mental lives and behavioural dispositions are consistent with the hypothesis that given what we mean by "+", and given the facts of elementary arithmetic, (1) is false and (2) is true! This startling conclusion leads to a wholly negative answer to question (B): nothing makes it the case that a linguistic expression has a meaning, because there are no facts in virtue of which an expression means one thing rather than another. Meaning, as Kripke's sceptic puts it, "vanishes into thin air". In Chapter 5, we'll also consider some of the potential difficulties that would be involved in embracing this

conclusion, and in Chapter 6 – the central chapter in the book – we'll trace out some of the implications of the meaning sceptical arguments for issues in the philosophy of mind and consider the positive responses to question (B) developed by a range of recent and contemporary philosophers in addition to that arguably present in the work of Wittgenstein himself.

In Chapters 7 and 8 we put scepticism about sense to one side and consider contributions towards questions (A) and (B) from two of the most important philosophers of language from the twentieth and twenty-first centuries. Suppose I utter sentence (6) with the intention of getting my audience to believe that snow is white in virtue of their recognition of that very intention. According to Paul Grice, this suffices for my utterance having the speaker's-meaning *that snow is white*, and the sentence-type "Snow is white" means that snow is white in virtue of its conventional association with that speaker's-meaning.[1] In contrast, Donald Davidson accounts for the meaning of "Snow is white" in terms of the fact that a formal theory of truth which satisfied the relevant constraints on radical interpretation would include as a theorem the biconditional

(i) "Snow is white" is true-in-English if and only if snow is white.

Likewise, for (6a) and (6b):

(ii) "Schnee ist Weiss" is true-in-German if and only if snow is white

and

(iii) "La neige est blanche" is true-in-French if and only if snow is white.

Davidson, like Frege, thinks of understanding a sentence as a matter of knowledge of truth-conditions, and this idea figures prominently in Chapter 9, where we consider Michael Dummett's idea that there is a tension between it and certain realist views in metaphysics. As intuitive realists about arithmetic and the past, we are prepared to assert that sentences (7) and (8) are determinately either true or false even though for all we know we may never be in a position

to say which. Yoking together Wittgensteinian and Fregean ideas about linguistic understanding, Dummett's semantic anti-realist suggests that we should refrain from asserting that (7) and (8) are determinately either true or false. In addition to considering semantic anti-realist arguments, we also consider Hilary Putnam's famous "Twin-Earth" argument that our understanding of sentences – such as (10) – which contain natural kind terms is essentially determined by the nature of the physical environment that we inhabit. Chapter 9 ends with a brief description of some "quietist" threats to Crispin Wright's critical development of Dummett's approach to debates between realists and their anti-realist opponents.

Overall, then, by the time you get to the end of this book, you will have had an introduction to the basic ideas in the philosophy of language propounded by Frege, Russell, Ayer, Quine, Kripke, Wittgenstein, Grice, Davidson, Dummett, and Wright and a number of leading recent and contemporary thinkers. Although this list is – as noted in the Preface – far from complete, a basic knowledge of the views of these philosophers is necessary for any kind of understanding of the philosophy of language, and (hopefully) sufficient to enable you to go on to extend your studies beyond the approaches to the subject considered in this book.

Note on use and mention

When referring to linguistic expressions, we use *quotation marks*. This also signifies that the quoted expression is being *mentioned* rather than *used*. Thus

(i) "Neil Armstrong" has thirteen letters

is an example of a case in which the expression is mentioned, and in which the first expression in the sentence stands for a linguistic expression, while

(ii) Neil Armstrong was the first man to step foot on the moon

is an example of a case in which the expression is *used*, and in which the first expression in the sentence stands for a particular man.

Note on types and tokens

In the course of the book, we sometimes make use of what is known as the *type–token distinction*. Very roughly, this marks a distinction between sorts (i.e. types) of things, and instances (i.e. tokens) of sorts of things. Thus in

(iii) blue
(iv) red
(v) Michael
(vi) blue

we have four word *tokens*, but three word *types*. (iii) and (vi) are *tokens of the same type*. Likewise, if Smith believes that Edinburgh is the capital city of Scotland and Jones believes that Edinburgh is the capital city of Scotland, we can say that Smith and Jones both token a belief of the same type.

Note

1 For the notion of a type (as distinct from a token), see the note on types and tokens.

Chapter 1

Frege

Semantic value and reference[1]

Philosophy of language is motivated in large part by a desire to say something *systematic* about our intuitive notion of meaning, and in the preface (to the first edition) we distinguished two main ways in which such a systematic account can be given. The most influential figure in the history of the project of systematising the notion of meaning (in both of these ways) is Gottlob Frege (1848–1925), a German philosopher, mathematician, and logician, who spent his entire career as a professor of mathematics at the University of Jena. In addition to inventing the symbolic language of modern logic,[2] Frege introduced some distinctions and ideas which are absolutely crucial for an understanding of the philosophy of language, and the main task of this chapter and the next is to introduce these distinctions and ideas, and to show how they can be used in a systematic account of meaning.

1.1 Frege's logical language

Frege's work in the philosophy of language builds on what is usually regarded as his greatest achievement, the invention of the

language of modern symbolic logic. This is the logical language that is now standardly taught in university introductory courses on the subject. As noted in the introduction, a basic knowledge of this logical language will be presupposed throughout this book, but we'll very quickly run over some of this familiar ground in this section.

Logic is the study of *argument*. A *valid argument* is one in which the premises, if true, guarantee the truth of the conclusion: i.e. in which it is impossible for all of the premises to be true and yet for the conclusion to be false. An *invalid argument* is one in which the truth of the premises does not guarantee the truth of the conclusion: i.e. in which there are at least some possible circumstances in which all of the premises are true and the conclusion is false.[3] One of the tasks of logic is to provide us with rigorous methods of determining whether a given argument is valid or invalid. In order to apply the logical methods, we have first to translate the arguments, as they appear in natural language, into a formal logical notation.[4] Consider the following (intuitively valid) argument:

(1) If Jones has taken the medicine then he will get better;
(2) Jones has taken the medicine; therefore,
(3) He will get better.

This can be translated into Frege's logical notation by letting the capital letters "P" and "Q" abbreviate the whole sentences or propositions out of which the argument is composed, as follows:

P: Jones has taken the medicine
Q: Jones will get better

The conditional "If . . . then . . ." gets symbolised by the arrow ". . .→. . .". The argument is thus translated into logical symbolism as:

$P \rightarrow Q$, P; therefore, Q.

The conditional "→" is known as a *sentential connective*, since it allows us to form a complex sentence (P → Q) by connecting two simpler sentences (P, Q). Other sentential connectives are: "and", symbolised by "&"; "or", symbolised by "v"; "it is not the case that", symbolised by "–"; "if and only if", symbolised by "↔". The capital letters "P", "Q", etc. are known as *sentential constants*, since they are abbreviations for whole sentences. For instance, in the example above, "P" is an abbreviation for the sentence expressing the proposition that Jones has taken the medicine, and so on. Given this vocabulary, we can translate many natural language arguments into logical notation. Consider:

(4) If Rangers won and Celtic lost, then Fergus is unhappy;
(5) Fergus is not unhappy; therefore
(6) Either Rangers didn't win or Celtic didn't lose.

We assign sentential constants to the component sentences as follows:

P: Rangers won
Q: Celtic lost
R: Fergus is unhappy.

The argument then translates as:

(P & Q) → R, –R; therefore –P v –Q.

Now that we have translated the argument into logical notation we can go on to apply one of the logical methods for checking validity (e.g. the truth-table method) to determine whether the argument is valid or not.

The logical vocabulary described above belongs to *propositional logic*. The reason for this tab is obvious: the basic building blocks of the arguments are sentences expressing whole propositions, abbreviated by the sentential constants "P", "Q", "R" etc. However, there

are many arguments in natural language which are intuitively valid, but whose validity is not captured by translation into the language of propositional logic. For example:

(7) Socrates is a man;
(8) All men are mortal; therefore
(9) Socrates is mortal.

Since (7), (8), and (9) are different sentences expressing different propositions, this would translate into propositional logic as:

P, Q; therefore, R.

The problem with this is that whereas the validity of the argument clearly depends on the *internal structure* of the constituent sentences, the formalisation into propositional logic simply ignores this structure. For example, the proper name "Socrates" appears both in (7) and (9), and this is intuitively important for the validity of the argument, but is ignored by the propositional logic formalisation which simply abbreviates (7) and (9) by, respectively, "P" and "R". In order to deal with this, Frege showed us how to extend our logical notation in such a way that the internal structure of sentences can also be exhibited. We take capital letters from the middle of the alphabet "F", G", "H", and so on, as abbreviations for *predicate expressions*; and we take small case letters "m", "n", and so on, as abbreviations for *proper names*. Thus, in the above example we can use the following translation scheme:

m: Socrates
F: . . . is a man
G: . . . is mortal.

(7) and (9) are then formalised as Fm and Gm respectively. But what about (8)? We can work towards formalising this in a number of stages. First of all, we can rephrase it as:

For any object: if it is a man, then it is mortal.

Using the translation scheme above we can rewrite this as:

For any object: if it is F, then it is G.

Now, instead of speaking directly of objects, we can represent them by using *variables* "x", "y", and so on (in the same way that we use variables to stand for numbers in algebra). We can then rephrase (8) further as:

For any x: if x is F, then x is G

and then as

For any x: $Fx \rightarrow Gx$

The expression "For any x" (or "For all x") is called the *universal quantifier*, and it is represented symbolically as $(\forall x)$. The entire argument can now be formalised as:

Fm; $(\forall x)(Fx \rightarrow Gx)$; therefore, Gm.

The type of logic which thus allows us to display the internal structure of sentences is called *predicate logic*, for obvious reasons (in the simplest case, it represents subject-predicate sentences *as* subject-predicate sentences). Note that predicate logic is not separate from propositional logic, but is rather an extension of it: predicate logic consists of the vocabulary of propositional logic

plus the additional vocabulary of proper names, predicates, and quantifiers. Note also that in addition to the universal quantifier there is another type of quantifier. Consider the argument:

(10) There is something which is both red and square; therefore
(11) There is something which is red.

Again, the validity of this intuitively depends on the internal structure of the constituent sentences. We can use the following translation scheme:

F: . . . is red
G: . . . is square.

We'll deal with (10) first. Following the method we used when dealing with (8) we can first rephrase (10) as:

There is some x such that: it is F and G.

Or,

There is some x such that: Fx & Gx.

The expression "There is some x such that" is known as the *existential quantifier*, and is symbolised as (∃x). (10) can thus be formalised as (∃x)(Fx & Gx), and, similarly, (11) is formalised as (∃x)Fx. The whole argument is therefore translated into logical symbolism as:

(∃x)(Fx & Gx); therefore (∃x)Fx.

That, then, is a brief recap on the language of modern symbolic logic, which in its essentials was invented by Frege. The introduction of this new notation, especially of the universal and existential quantifiers, constituted a huge advance on the syllogistic logic which had dominated philosophy since the time of Aristotle. It allowed logicians to formalise and prove intuitively valid arguments whose form and validity could not be captured in the traditional Aristotelian logic. An example of such an argument is:

(12) All horses are animals; therefore,
(13) All horses' heads are animals' heads

1.2 Syntax

A *syntax* or *grammar* for a language consists, roughly, of two things: a specification of the vocabulary of the language, and a set of rules which determine which sequences of expressions constructed from that vocabulary are grammatical (syntactically well-formed) and which are ungrammatical (syntactically ill-formed). For example, in the case of the language of propositional logic, we can specify the vocabulary as follows:

Sentential connectives: expressions having the same shape as "→" or "–" or "&" or "v" or "↔"

Sentential constants: expressions having the same shape as "P", "Q", "R", and so on.

It is important to note that when we are working at the level of syntax, the only properties of expressions that are mentioned in the specifications of the vocabulary are *formal properties*, such as shape. This is clearly the case in the specification of the vocabulary of propositional logic above: in principle, even someone who had no knowledge whatsoever of what the various bits of vocabulary

mean could separate expressions into those that belong to the vocabulary and those that do not. In this regard, syntax is prior to semantics, the study of meaning. This is true also of the syntactical rules: these determine, in terms of purely formal properties of the expressions concerned, whether a given sequence of expressions drawn from the vocabulary counts as grammatical or not. For example, the syntactical rules for propositional logic can be stated very simply as follows:

(i) Any sentential constant is grammatical.
(ii) Any grammatical expression preceded by "–" is grammatical.
(iii) Any grammatical expression followed by "→" followed by any grammatical expression is grammatical.
(iv) Any grammatical expression followed by "&" followed by any grammatical expression is grammatical.
(v) Any grammatical expression followed by "v" followed by any grammatical expression is grammatical.
(vi) Any grammatical expression followed by "↔" followed by any grammatical expression is grammatical.
(vii) Any sequence of expressions which does not count as grammatical in virtue of (i)–(vi) is not grammatical.[5]

Again, someone with no knowledge of what the expressions concerned mean (e.g. that "&" means *and*, that "v" means *or*, and so on) could use these rules to determine whether an arbitrary sequence of marks counts as a grammatical expression of the language of propositional logic. To see this, consider how we could use the rules to show that e.g. "(P & Q) v R" is grammatical. First of all, on the basis of shape properties, we would identify P, Q, and R as sentential constants, and that "&" and "v" count as sentential connectives. On the basis of rule (i), we would then identify "P", "Q", and "R" as grammatical. Then, on the basis of (iv), we would identify "(P & Q)" as grammatical (in terms of purely formal properties, such as the shape and ordering of the constituent expressions). Finally, on the basis of (v) we would identify "(P & Q) v R" as grammatical (again, in terms of purely formal properties).

We can do the same thing for the language of predicate logic. We can specify the vocabulary of predicate logic – proper names, predicate expressions, variables, and quantifiers – in purely formal terms, and then give formal rules which determine which sequences of marks count as grammatical. The details of this needn't concern us here. What is important for present purposes is simply to note that Frege discerns the following *syntactical categories* in his logical language: proper names, predicates, declarative sentences, sentential connectives, and quantifiers.

1.3 Semantics and truth

In dealing with the syntax of a language, we are dealing only with the purely formal properties of its constituent expressions. But, of course, in addition to those formal properties, the expressions can also possess *semantic* properties: they *mean* this, or *refer* to that, and so on. In semantics we move from considering the purely formal properties of linguistic expressions to considering their meaning and significance.

Let's start by thinking a little more about arguments in propositional logic, and how we determine their validity. Consider another very simple argument:

(14) Beethoven was German and Napoleon was French; therefore
(15) Beethoven was German.

This formalises as P & Q; therefore, P. Now, how do we determine whether this argument is valid or not? Recall that an argument is said to be valid if there are no possible circumstances in which all of its premises are true and its conclusion is false. One way to determine whether an argument is valid, then, is simply to enumerate the various possible distributions of truth and falsity over the premises and conclusion, and check whether there are any such that the premises all come out true and the conclusion comes out false. If there are, the argument is invalid; if there are not, the argument is valid. This, of course, is just the familiar *truth-table* method of determining validity. The truth-table for the argument above is as follows:

P	Q	P & Q	P
T	T	T	T
T	F	F	T
F	T	F	F
F	F	F	F

There are four possible distributions to the constituent sentences P and Q, and these are enumerated on the four lines on the left-hand side of the table, with T representing truth and F representing falsity. Given this, we can work out the possible distributions of truth and falsity to the premise and conclusion: this is done in the third and fourth columns. We see that there is only one circumstance in which the premise is true – when both P and Q are assigned the value true – and that in this case, the conclusion is also true. So there are no possible cases in which the premise is true and in which the conclusion is false. So the argument is valid.

What does the question about the validity of an argument have to do with semantics? Intuitively, the validity of an argument is going to depend on the meanings of the expressions which appear in it. That is to say, the validity of an argument is going to depend on the semantic properties of the expressions out of which it is constructed. In the argument above the basic expressions out of which the argument is constructed are sentences. What properties of the sentences are relevant to determining the validity of the inference? In the first instance, it seems as if it is the properties of *truth* and *falsity*. After all, the truth-table method works by determining the possible distributions of these very properties. So, truth and falsity look like good candidates for the semantic properties in question. Given all of the possible assignments of truth and falsity to P and Q, we can work out the various assignments of truth and falsity to the premises and conclusion, and this allows us to say whether or not the argument is valid. So, validity is determined by the possible distributions of truth and falsity to the premises and conclusion, and this in turn is determined by the possible distributions of truth and falsity to the constituent sentences. Let's define the notion of *semantic value* as follows:

Definition: The semantic value of any expression is that feature of it which determines whether sentences in which it occurs are true or false.[6]

In the case we have just looked at, the constituent expressions of the argument are the sentences P, Q. Which features of P, Q are relevant to determining whether the sentences in which they occur are true or false? Well, their truth or falsity: as shown in the truth-table, the distributions of T and F to P and Q determine the truth or falsity of the complex sentence P & Q which forms the premise of the argument. Given the definition above, then, it follows that the semantic value of a sentence is its truth-value.

We have here the beginnings of a semantic theory: an assignment of a semantic property (truth or falsity) to the sentences of a language, in terms of whose distribution the validity of the inferences in which those sentences appear as constituents is determined. In the next section, we develop this theory further.

1.4 Sentences and proper names

Frege's name for the semantic value of an expression, as defined in the previous section, was *Bedeutung*.[7] According to Frege, then, the semantic value of a sentence is one of the truth-values, true or false. Note that in the case above, the semantic value of the complex expression P & Q – its truth-value – is determined by the truth-values of the constituent sentences P, Q and the way they are put together (this is shown by the relevant truth-table). In general, the semantic value of a complex expression is determined by the semantic values of its parts and the way they are put together. This is known as the *principle of compositionality*. Thus far, then, we can discern two theses in Frege's semantic theory:[8]

Thesis 1: The semantic value of a sentence is its truth-value (true or false).

Thesis 2: The semantic value of a complex expression is determined by the semantic values of its parts and the way they are put together.

From this, we can derive a third thesis. Since the semantic value of a complex expression is determined by the semantic values of its parts, substituting one expression in a sentence with another which has the same semantic value will leave the semantic value (truth-value) of the whole sentence unchanged:

Thesis 3: Substitution of a constituent of a sentence with another expression which has the same semantic value will leave the semantic value (i.e. truth-value) of the sentence unchanged.

So far, though, we have only considered expressions from one of the syntactic categories introduced in §1.2, declarative sentences. Frege extends this semantic theory to cover expressions from the other syntactic categories: proper names, sentential connectives, predicates, and quantifiers. The idea is to assign a type of semantic value to each type of expression: as in the case of declarative sentences, this will be the property of the type of expression which determines the contribution of instances of that type to the truth or falsity of the sentences in which they appear.

Let's begin with the case of proper names. Consider the sentence "Cicero is Roman". What feature of the proper name "Cicero" is relevant to determining whether this sentence is true or false? Intuitively, the fact that it stands for the individual *object* which is the man Cicero: if the proper name stood for some other individual (e.g. Plato) the sentence in question might have a different truth-value from the one it actually has. So, just as the semantic value of a declarative sentence is a truth-value, the semantic value of a proper name is an object. This allows us to state the fourth thesis of Frege's semantic theory:

Thesis 4: The semantic value of a proper name is the object which it refers to or stands for.[9]

This might seem a little odd. Isn't it just a platitude that proper names refer to objects? And if it is a platitude, how can it be a thesis of a substantial semantic theory? The important thing to appreciate here is that Frege is using the notion of "semantic value" in a technical way: the notion of semantic value has its content fixed by the definition above. Given the definition, it can emerge as a *discovery* that the semantic value of a proper name is the object which it refers to. That this corresponds with our intuitive use of "reference" as applied to proper names is all to the good. However, this led Frege to some strange-sounding views. Just as Cicero is an object, and is the reference of the proper name "Cicero", Frege construed the semantic values of sentences, the truth-values true and false, as objects also, and this led him to construe sentences as a kind of *proper name* for these objects, which he called the True and the False:

> Every assertoric sentence concerned with the [semantic value] of its words is therefore to be regarded as a proper name, and its [semantic value], if it has one, is either the True or the False.[10]

Now this sounds bizarre! Frege himself realised that his characterisation of truth-values as objects is apt to evoke this sort of reaction, saying that "The designation of the truth-values as objects may appear to be an arbitrary fancy or perhaps a mere play on words".[11] In what follows, we'll simply ignore this strange-sounding doctrine. The thing to bear in mind is that the notion of semantic value is a technical term, whose content is given by our definition: sentences can be assigned semantic values in this technical sense, and so can proper names, but the fact that the semantic values of the latter are objects needn't force us into accepting that the semantic values of the former are also objects of a special and mysterious kind.

Theses 1 and 4 specify the semantic values of declarative sentences and proper names, that is, the semantic properties of those expressions in virtue of which sentences containing them are determined as true or false, and, in turn, in terms of which arguments containing those sentences as constituents are determined as valid or invalid. But what about the expressions in the

other syntactic categories discerned by Frege: connectives, predicates, and quantifiers? Before answering this question, we need to prepare by considering what Frege says about mathematical functions.

1.5 Function and object

The semantics which Frege provides for the connectives, predicates, and quantifiers stems from an analogy with *mathematical functions*. The idea of a functional expression will be familiar to anyone who has studied elementary mathematics. Take the functional expression "y = 2x". Here y is said to be a function of x: we get different values for y as we insert different numerals for x. The numbers which the variable x stands for are called the *arguments* of the function (this must not be confused with the notion of argument used in logic, as in "valid argument"). Thus, for the argument 1, we get the value 2, for the argument 2, we get the value 4, for the argument 3, we get the value 6, and so on. We can thus represent the function as a set of ordered pairs, in each of which the first member corresponds to the argument of the function and the second member corresponds to the value which the function delivers for that argument. Thus, the function y = 2x can be represented as {(0, 0), (1, 2), (2, 4), (3, 6), . . .}.[12] Call this the *extension* of the function. Now "y = 2x" stands for a *function of one argument*: there is only one variable, so only one numeral can be "slotted in" to deliver a value for the function. There can also be *functions of two arguments*. For example, "z = 2x + 5y" stands for such a function. Here we need to slot in two numerals in order to obtain a value for the function: e.g. the value of the function for x = 1 and y = 1 is 7, and for x = 1 and y = 2 its value is 12. We can represent a function of two arguments as a set of ordered triples, with the first member of the triple representing the arguments for x, the second member the arguments for y, and the third member the value delivered by the function for those arguments. Thus, the function just given has the extension {(0, 0, 0), (1, 0, 2), (0, 1, 5), (1, 1, 7), (1, 2, 12), . . .}.

Now, consider the process by means of which we determine the values of the function which "y = 2x" stands for. We slot in

the arguments and calculate the values as follows: 2·0 (= 0), 2·1 (= 2), 2·2 (= 4), 2·3 (= 6), and so on. This talk of "slotting in" arguments suggests that the expression which stands for a function must have a "gap" into which expressions standing for the arguments can be slotted in: so we might represent the functional expression in this case as "y = 2( )", where the brackets show that there is an empty space in the functional expression which must be filled by an expression of the appropriate sort in order for a value to be computed. In fact, representing the function as "y = 2x" does this just as well, since the variable "x" does not stand for a specific number, but only serves to indicate the place where a numeral standing for a particular number may be inserted to obtain a value. Frege represents this feature of functions by saying that they are *incomplete* or *unsaturated*:

> I am concerned to show that the argument does not belong with a function, but goes together with the function to make up a complete whole; for a function by itself must be called incomplete, in need of supplementation, or unsaturated.[13]

This contrasts with the case of proper names (including numerals, which are the proper names of numbers) and sentences, which have no such gap: in contrast to functional expressions, the objects they stand for are complete or saturated.

In the case of the functions above we have functions from numbers to numbers: both functions take numbers as arguments and yield a number as value. The insight of Frege's which led to his semantics for predicates, connectives, and quantifiers was the realisation that there can be functions which take things *other* than numbers as arguments and values.[14]

1.6 Predicates, connectives, and quantifiers

Consider the predicate expression ". . . is even". Like the functional expressions discussed in the previous section, this has a gap into which a numeral can be slotted. What is the result of slotting a given numeral into the gap? It will be a *true* sentence, if the number denoted by the numeral is even; it will be a *false* sentence otherwise.

Thus, we can view the predicate ". . . is even" as standing for a function from numbers to truth-values. But there are also functions which take objects other than numbers as their arguments. Consider ". . . is round". This has a gap into which a proper name may be slotted, and the value delivered will be true if the object denoted by that proper name is round, false otherwise. Thus ". . . is round" can be viewed as standing for a function from objects to truth-values. *In general, a predicate expression will stand for a function from objects to truth-values. Frege reserves the term "concept" for a function whose value is always a truth-value.*

This allows us to state the fifth thesis of Frege's semantic theory:

Thesis 5: The semantic value of a predicate is a function.

By analogy with the examples in the previous section, the extension of the function denoted by ". . . is even" is the set of ordered pairs {(1, false), (2, true), (3, false), (4, true), . . .}.

Intuitively, it is the extension of a predicate which determines the truth-value of sentences in which it appears. Take a subject-predicate sentence like "4 is even". That this is true is determined by two things: firstly, that the numeral "4" stands for the number 4, and secondly, that the number 4 is paired with the value true in the extension of the function denoted by ". . . is even". Also, thesis 3 states that the substitution, in a complex expression, of a part with some other part having the same semantic value, leaves the semantic value (truth-value) of the whole unchanged. We can see that this condition is met if we identify the semantic value of a predicate with a function, understood in extensional terms: the substitution of a predicate having the same extension as the predicate " . . . is even" will leave the truth-value of "4 is even" unchanged, since the identity in extension will ensure than the number 4 is still paired with the value true.[15] This leads us to:

Thesis 6: Functions are extensional: if function f and function g have the same extension, then f = g.[16]

We can also include the logical connectives and the quantifiers within the scope of our semantic theory, since these too can be viewed as standing for functions. Indeed, the logical connectives that we introduced above are often called "truth-functions" or "truth-functional connectives". *The reason is that these can be viewed as standing for functions from truth-values to truth-values*. Take the negation operator "–. . .". This can be viewed as standing for a function of one argument, which has the following extension: {(T, F), (F, T)}. For the argument true, the value false is delivered, and for the argument false, the value true is delivered. Likewise, the connective for conjunction, ". . . & . . ." can be viewed as standing for a function of two arguments, which has the following extension: {(T, T, T), (T, F, F), (F, T, F), (F, F, F)}.

Note that this allows us to respect the thesis that the semantic value of a complex expression is determined by the semantic values of its parts. Consider a complex sentence such as "Beethoven was German and Napoleon was French". This is formalised as P & Q. It is true if and only if the truth-values of P, Q are paired with T in the extension of the function denoted by ". . . & . . .". P is T and Q is T, and (T, T, T) is included in the extension of the function. So P & Q is true.

What about the universal and existential quantifiers? Frege treats these as standing for a special sort of function: *second-level functions*. A first-level function is a function which takes objects (of whatever sort) as arguments. A second-level function is a function which takes *concepts* as arguments. Frege viewed the universal and existential quantifiers as standing for second-level functions, taking concepts as arguments and yielding truth-values as values. Let's deal with the universal quantifier first. As will be familiar, whenever we are formalising parts of natural language by using quantifiers, we have to specify a universe of discourse: this is the group of objects which our variables are taken to range over. Suppose that we select the group of humans {Hillary Clinton, Vladimir Putin, Theresa May, Donald Trump} as our universe of discourse. Now consider the universally quantified sentence "Everyone is mortal". We can formalise this, taking "G" to abbreviate ". . . is mortal", as follows: $(\forall x)Gx$. Frege suggested

that we view the quantifier as standing for a function (∀x)( ), which takes a concept Gx as argument and yields the truth-value T if the concept G is paired with T in its extension. The concept G will be paired with T in the extension of the quantifier if every object in the universe of discourse is paired with T in the extension of G. Similarly (∀x)Gx yields the truth-value F if the concept G is paired with F in the extension of the quantifier. And the concept G is paired with F in the extension of the quantifier if at least one object in the universe of discourse is paired with the value F in the extension of G. Thus, consider "Everyone is mortal". (∀x)( ) is a second-level function, from concepts to truth-values. If the argument is the concept Gx, then the function (∀x)( ) yields the value T if G is paired with T in its extension. In turn, G will be paired with T in the extension of (∀x)( ) if *every* object in the universe of discourse is paired with T in the extension of G. In the case at hand, the extension of G is {(Hillary Clinton, T), (Vladimir Putin, T), (Theresa May, T), (Donald Trump, T)}. We see that every object *is* paired with T in the extension of G, so that G will be paired with T in the extension of (∀x)( ). So, finally, "(∀x)Gx" is true. Note that this shows that the semantic value (truth-value) of the sentence "(∀x)Gx" is determined by the semantic values of its parts, namely, the extension of the function (∀x)( ), and the extension of the concept G.

Likewise, consider the existentially quantified sentence "Someone is Russian", keeping the universe of discourse the same as in the example above. We can formalise this as (∃x)Hx, taking "H" to abbreviate ". . . is Russian". We can then spell out how the semantic value of the existentially quantified sentence is determined by the semantic values of its parts as follows. (∃x)( ) is a second-level function, from concepts to truth-values. If the argument is the concept Hx, then the function (∃x)( ) yields the value T if H is paired with T in its extension. In turn, H will be paired with T in the extension of (∃x)( ) if *at least one* object in the universe of discourse is paired with T in the extension of H. In the case at hand, the extension of H is {(Hillary Clinton, F), (Vladimir Putin, T), (Theresa May, F), (Donald Trump, F)}. We see that at least one object is paired with T in the extension of H (Vladimir Putin), so

that H will be paired with T in the extension of (∃x)( ). So, finally, "(∃x)Hx" is true.

It might be useful to summarise these points about predicates, connectives, and quantifiers in a separate thesis:

> **Thesis 7:** The semantic value of a predicate is a first-level function from objects to truth-values; the semantic value of a sentential connective is a first-level function from truth-values to truth-values; the semantic value of a quantifier is a second-level function from concepts to truth-values.

1.7 A semantic theory for a simple language

The above considerations provide us with one way of attaining a systematic perspective on the semantics of a given language. We identify a range of syntactic categories to which the various expressions in that language belong, and to each category we give an account of the semantic property (semantic value) in virtue of which instances of that category impact upon the truth of the sentences in which they appear. All this is done in such a way that the principle of compositionality is respected: the semantic values of complex expressions are determined by the semantic values of their parts.[17] Let's run through this idea in application to a very simple language. This language consists of two predicates, "G" and "H", which abbreviate ". . . is Greek" and ". . . is Scottish" respectively; four proper names, "a", "b", "c", and "d", standing for Plato, Socrates, Hume, and Reid respectively. We'll suppose that the language contains just one sentential connective, negation, and, in addition, the universal quantifier and a stock of variables which range over the four objects in the quantifier's universe of discourse.

First of all, we spell out the semantic properties (semantic values) of the primitive expressions of the language in a series of *axioms*:

Axiom 1: "a" refers to Plato

Axiom 2: "b" refers to Socrates

Axiom 3: "c" refers to Hume

Axiom 4: "d" refers to Reid

Axiom 5: the extension of "G" is {(Plato, T), (Socrates, T), (Hume, F), (Reid, F)}[18]

Axiom 6: the extension of "H" is {(Plato, F), (Socrates, F), (Hume, T), (Reid, T)}

Axiom 7: the extension of "–" is {(T, F), (F, T)}.

We'll also add three *compositional axioms*:

Compositional axiom 1: A sentence coupling a proper name with a predicate is true if and only if the object referred to by the proper name is paired with T in the extension of the predicate.

Compositional axiom 2: The negation of a sentence is true if and only if the truth-value of that sentence is paired with T in the extension of "–. . .".

Compositional axiom 3: A universally quantified sentence is true if and only if the predicate involved is paired with T in the extension of the quantifier; the predicate involved is paired with T in the extension of the quantifier if and only if every object in the universe of discourse is paired with T in the extension of the predicate.

Given these compositional axioms, and the axioms 1–7, we can work out the semantic values (i.e. truth-values) of the complex *sentences* of the language. We'll look at three examples:

Example 1: "Plato is Greek". This is translated into logical language as Ga.

(i) "Ga" is true if and only if the object referred to by the proper name "a" is paired with T in the extension of "G" (compositional axiom 1)
(ii) "a" refers to Plato (from axiom 1)
(iii) the extension of "G" is {(Plato, T), (Socrates, T), (Hume, F), (Reid, F)} (axiom 5)
(iv) "Ga" is true (from (i), (ii), and (iii)).

Example 2: "Everyone is Scottish". This is translated into logical language as (∀x)Hx.

(i) "(∀x)Hx" is true if and only if H is paired with T in the extension of the quantifier; H is paired with T in the extension of the quantifier if and only if every object in the universe of discourse is paired with T in the extension of H (from compositional axiom 3)
(ii) the extension of "H" is {(Plato, F), (Socrates, F), (Hume, T), (Reid, T)} (axiom 6)
(iii) not every object in the universe of discourse is paired with T in the extension of H (e.g. Plato and Socrates are paired with F) (from (ii))
(iv) H is not paired with T in the extension of (∀x)(. . .x) (from (i))
(v) "(∀x)Hx" is false (from (i) and (iv)).

Example 3: "Not everyone is Scottish". This is translated into logical language as – (∀x)Hx.

As for example 2, followed by:

(vi) "– (∀x)Hx" is true if and only if the truth value of "(∀x) Hx" is paired with T in the extension of "–. . ." (from compositional axiom 2)
(vii) The truth-value of "(∀x)Hx" (F) is paired with T in the extension of "–. . ." (from axiom 7)
(viii) "– (∀x)Hx" is true (from (vi) and (vii)).

These are examples of how we might derive, via the semantic values assigned to the primitive expressions of the language by Frege, the truth-values of the complex sentences of that language. In addition, we can also use these as examples of how the *truth-conditions* of sentences might be derived on the basis of the assignments of semantic values to their parts. The truth-condition for a sentence is not a truth-value like T, but rather the condition which must obtain in the world in order for the sentence to be true. What condition must obtain in the world in order for the sentence "Plato is Greek" to be true? Intuitively, *that Plato is Greek*. This intuition is captured by Frege's theory as follows:

(i) "Ga" is true if and only if the object referred to by the proper name "a" is paired with T in the extension of "G" (compositional axiom 1)
(ii) "a" refers to Plato (from axiom 1)
(iii) "Ga" is true if and only if Plato is paired with T in the extension of "G" (from (i) and (ii))
(iv) Plato is paired with T in the extension of "G" if and only if Plato is Greek (from the meanings of "true" and "extension")
(v) "Ga" is true if and only if Plato is Greek (from (i) and (iv))
(vi) "Plato is Greek" is true if and only if Plato is Greek (from (v) and the formalisation of "Plato is Greek").

Just as we expected! (vi) is called a *homophonic* statement of the truth-condition of "Plato is Greek", since the very same sentence is used on the right-hand side of the biconditional (vi). The notion of truth-conditions – and the idea of a semantic theory which systematically generates statements of truth-conditions for the sentences of a language on the basis of semantic properties assigned to the language's primitive expressions – will figure prominently later this book.

Chapter summary

In this chapter, we began to outline Frege's informal and formal theories of meaning. The fundamental notion is that of *Bedeutung*, or *semantic value*, which we defined as that feature of an expression which determines its contribution to the truth-value of sentences in which it appears. We saw that for Frege, the semantic value of a sentence is its truth-value, while the semantic value of a proper name is the object which is its bearer. Predicates, logical connectives, and quantifiers all have functions of one sort or another as their semantic values. We also saw that semantic value is *compositional*, in that the semantic value of a complex expression is determined by the semantic values of its parts and the way they are put together.

Study questions

1. Use the truth-table method to show that the argument

$$(P \,\&\, Q) \rightarrow R,\ -R;\ \text{therefore}\ -P \vee -Q$$

   is valid.

2. Make up an example of your own of an argument that is intuitively valid due to the internal structure of its premises and conclusion, but which would count as invalid if propositional logic were all that we had at our disposal.
3. Formalise the following argument in the language of predicate logic:

   All horses are animals; therefore, all horses' heads are animals' heads.

4. Is Frege's view that sentences are proper names standing for the True or the False really as weird as it initially sounds?
5. What are the extensions of the logical connectives "v" ("or") and "↔" ("if and only if")?

6. Show how the truth-value of the sentence "Everyone is Russian" is determined by the semantic values of its parts.
7. Derive statements of the truth-conditions of "Everyone is Scottish" and "Someone is not Scottish".
8. According to Frege, proper names stand for objects which, unlike functions, are *complete* or *saturated*. The expression "the concept *horse*" appears to be a proper name: it is complete, in that there is no gap into which another expression can be inserted. So if it refers to anything, it refers to an object. But according to Frege, concepts are functions which, unlike objects, are incomplete or unsaturated. How much of a problem is this for Frege?

Further reading

There are many good introductions to the modern form of symbolic logic which Frege invented. The best of these is probably Tomassi, *Logic*. Other textbooks which can be recommended are Lemmon, *Beginning Logic*; Newton-Smith, *Logic*; Hodges, *Logic*; and Jeffrey, *Formal Logic*. Acquaintance with at least the elementary sections of one of these textbooks is essential for following the rest of this book. For a useful discussion of the central philosophical issues surrounding logic, see Read, *Thinking about Logic*. On the paradox of "the concept *horse*" (study question 8), see Kenny, *Frege*, pp.122–125; and Textor, *Routledge Philosophy Guidebook to Frege on Sense and Reference*, pp.254–261). Further reading on Frege's notion of reference/semantic value is given in the Further reading section at the end of Chapter 2.

Notes

1 Note that in this chapter and the next I have concentrated on presenting the core of Frege's philosophy of language in a simple and straightforward manner. The interpretation of Frege's philosophy is controversial, and no short presentation can do justice to the complexities which detailed exegesis would involve. Forced to choose between presenting Frege's views in a short and accessible manner, and respecting the texts down to the last letter, I have opted for the former. Obviously, like any philosopher, Frege's views evolved in the course of his career: in these chapters, I aim

to outline Frege's "mature" (post-1892) philosophy. Those wishing to follow up exegetical questions should consult the works listed under Further reading for Chapter 2.

2 Although other philosophers and logicians – such as Boole – contributed to the invention of modern logic, Frege is usually viewed as its most important founding father. Note that Frege's logical *notation* (as developed in his *Begriffschrift* (or "concept script")) is much more cumbersome than the standard notation which is taught nowadays and used in this book (which is more like that in Russell and Whitehead's *Principia Mathematica*). For some useful discussion of the *Begriffsschrift*, see Currie, *Frege*, Chapters 1 and 2; Textor, *Routledge Philosophy Guidebook to Frege*, Chapter 2.

3 Note, though, that Frege himself would not have appealed to modal notions – such as necessity and possibility – in explaining validity. This is because of views he held about the nature of logic, which we needn't concern ourselves with here. Note also that Frege subscribed to a view – *logicism* – according to which arithmetic can ultimately be reduced to logic. See Frege, *Foundations of Arithmetic*. For commentary, see Noonan, *Frege*, Chapter 3; Kenny, *Frege*, Chapters 4 and 5.

4 Note that Frege himself was interested in the language of mathematics rather than with natural language as such, though it has since been commonplace in the philosophy of language for Frege's insights to be extended to natural languages like English.

5 We would actually have to complicate matters just a little more by including brackets in the vocabulary, and rules governing them among the syntactical rules.

6 The fact that the semantic value of any expression is a matter of its contribution to the truth-values of sentences in which it appears is no doubt related to what Frege calls one of his fundamental principles, the principle that one should "never ask for the meaning of a word in isolation, but only in the context of a sentence" (Frege, *Foundations of Arithmetic*, p.x). See Baldwin, "Philosophy of Language", p.62.

7 This is often translated as "reference" or "meaning". I have chosen the technical term "semantic value" in place of these, since Frege's notion has a precise definition which applies to types of expression – e.g. predicates and sentences – which we would not normally take as referring to anything in the ordinary sense of reference. I have altered passages from Frege quoted in the text accordingly, and have signaled this by the use of square brackets.

8 For ease of reference, all of Frege's theses are listed in an appendix at the end of Chapter 2.

9 This probably explains in part why Frege's term *Bedeutung* is often translated as "reference". But, as noted above, this is best avoided, since it makes the idea that the notion of *Bedeutung* can be applied to, e.g., predicates and sentences seem quite odd, when in fact it is no odder than its application to names.
10 Frege, "On *Sinn* and *Bedeutung*", p.158.
11 Ibid., p.158.
12 We need to distinguish between the notion of an ordered pair and the notion of a *set*, represented by the use of curly brackets { , }. Roughly, a set is a collection of objects, where the order of the objects is irrelevant to the identity of the set. Thus, {1, 2} is the *same* set as {2, 1}. In the case of an ordered pair (represented by normal brackets), the order does matter. So, for example, (1, 2) is a different ordered pair from (2, 1). *Mutatis mutandis* for ordered triples, etc.
13 Frege, "Function and Concept", p.133.
14 Note that we have distinguished between *functional expressions*, which are linguistic items (such as predicates), and *functions*, which according to Frege are extra-linguistic, abstract entities. Thus, the semantic value of a functional expression is a function. See §1.6.
15 Rather than long-windedly speaking of the extension of the function which a functional expression stands for, in the text I have simply spoken of the extension of a functional expression. Two functional expressions have the same extension, if the functions which they stand for have the same extension.
16 Note that Frege himself – for reasons which needn't concern us here – would have demurred from speaking of functions f and g as identical; see Dummett, *Frege*, p.209. I gloss over this here in order to keep things simple (see n.1 above).
17 This displays how Frege gives a systematic account of meaning in both of the senses distinguished in the preface: theses 1–7 provide part of a systematic theory of meaning in the "informal" sense, and we now see how they can be used as a basis for a systematic theory of meaning in the "formal" sense.
18 Note that thesis 6 allows us to simplify things: we can specify the semantic values of incomplete expressions by specifying their extensions (i.e. the extensions of the functions which they stand for). Note also that we don't need a separate axiom giving the extension of the universal quantifier, since this is taken care of by the second half of compositional axiom 3.

Chapter 2

Frege and Russell

Sense and definite descriptions

2.1 The introduction of sense

We have been looking at Frege's attempt to give a systematic account of meaning. We started out with the idea that the validity of arguments depends upon the semantic properties possessed by the expressions out of which their constituent sentences are constructed. So, one way to find out what semantic properties a systematic treatment of meaning should employ would be ask which properties of expressions are relevant to the validity of arguments in which they appear. We saw that an initially plausible answer to this question, in the case of whole sentences, was the property of truth. So we defined the semantic value of an expression as that feature of it which determines whether sentences in which it appears are true or false. This led us to identify the semantic values of proper names as their bearers, of sentences as their truth-values, and of functional expressions as functions. We saw that we were able to do this in a way which respects the principle of compositionality, so that the semantic value of a complex expression is

systematically determined by the semantic values of its parts and the way they are combined. Thus far, then, we have been attempting to give a systematic account of the intuitive notion of meaning by constructing a semantic theory which trades in just *one* semantic property, semantic value (although, as noted, expressions from different syntactic categories will be assigned different sorts of semantic value). But is one semantic property a sufficiently rich basis on which to construct a philosophical account of a phenomenon as complex as that of human language? It would be odd if it were: we don't expect physics to use just one physical property in its explanations, or biology to use just one biological property in its explanations, so it would be strange if a theory of meaning could get by with just the one meaning-relevant property of semantic value. We begin this chapter by looking at Frege's reasons for thinking that we have to appeal to some other semantic property in addition to semantic value in our account of the intuitive notion of meaning: the property of having a *sense*.[1]

2.1.1 The problem of bearerless names

Let's take the case of names as an example. We are trying to give a systematic account of the meanings of names, and in the theory of semantic value described in Chapter 1 we attempt to do this in terms of the assignment of the property of having a semantic value to the names, where the semantic value of a name is the object which it refers to. But if having a reference were the only semantic property in terms of which we could explain the functioning of names, we would be in trouble with respect to names which simply have no bearer. Consider a sentence such as "Odysseus was set ashore at Ithaca while sound asleep". The name "Odysseus" has no bearer, since the character is entirely fictional. Since the name has no reference, and the semantic value of a sentence is determined by the semantic values of its parts, it follows that the sentence "Odysseus was set ashore at Ithaca while sound asleep" has no semantic value either. So if having a semantic value were the only semantic property, we would have to say that the sentence is meaningless. But we can certainly understand the sentence: it is certainly not just meaningless

gibberish. So it seems that we will have to attribute some other semantic property to the name "Odysseus" in addition to its reference. We could sum this up as follows: *names without a reference (semantic value) are not meaningless; so there must be some other semantic property possessed by names in addition to having a reference (semantic value).*

2.1.2 The problem of substitution into belief contexts

According to the principle of compositionality, the semantic value of a complex expression is determined by the semantic values of its parts. It followed from this that (thesis 3) substitution of a part of a sentence with another expression which has the same semantic value will leave the truth-value of the whole sentence unchanged. This means, in particular, that substitution of one name in a sentence by another having the same reference should leave the truth-value of that sentence unchanged. But this appears to be false. Consider the following sentence, where John is a person with absolutely no knowledge about Mark Twain (except perhaps that he is the author of *Huckleberry Finn*):

(1) John believes that Mark Twain is Mark Twain.

This will be true, unless John has some very bizarre views on identity. But "Mark Twain" and "Samuel Clemens" are co-referential: they are different names for the same person. So the following, which results by substituting one of the occurrences of "Mark Twain" by "Samuel Clemens", should also be true:

(2) John believes that Mark Twain is Samuel Clemens.

But, of course, this is actually false, since John knows nothing about Mark Twain except that he is the author of *Huckleberry Finn*. This suggests that we are either going to have to give up thesis 4, that the semantic value of a name is its bearer, or thesis 3, and thereby thesis 2, the principle of compositionality. But these are both central and indispensable planks of Frege's theory of semantic value. *We shall see that Frege's attempt to solve this problem without giving up either of theses 2, 3, or 4, requires the*

introduction of another semantic property of names in addition to their having a reference or semantic value.[2]

This point is perfectly general, and applies to expressions in the other syntactic categories as well. In fact, it is perhaps clearest when made with respect to sentences. Thesis 3 in that case amounts to: in a complex sentence, the substitution of a component sentence by another sentence having the same truth-value should not change the truth-value of the complex sentence. But belief contexts threaten this principle as well. Consider the following sentence, where John is a person with a working knowledge of British geography, but absolutely no knowledge of particle physics:

(3) John believes that London is south of Glasgow.

Here, the overall sentence is true, as is the "embedded" sentence "London is south of Glasgow". Now the sentence "Electrons are negatively charged sub-atomic particles" is also true, so substituting it for the embedded sentence in the above sentence should not result in a change of truth-value in the overall sentence. But it does result in such a change, since

(4) John believes that electrons are negatively charged particles

is false. Again this suggests that Frege will have to give up either thesis 1, that the semantic value of a sentence is its truth-value, or thesis 3, and thereby thesis 2. We shall see that Frege's attempt to solve this problem without giving up either of theses 1, 2, or 3 requires the introduction of another semantic property of sentences in addition to their having a semantic value (truth-value). Likewise, Frege's attempt to solve the analogous problem in the case of functional expressions, without giving up theses 2, 3, 5, or 6, requires the introduction of another semantic property of functional expressions in addition to their having a semantic value.[3]

2.1.3 The problem of informativeness

When someone understands an expression, we say that he knows its meaning: meaning is that semantic property of an expression which someone with an understanding of that expression grasps.

Now suppose that the meaning of an expression, in this intuitive sense, was identified with its semantic value. What would follow? Consider Frege's famous example concerning the planet Venus. It took an empirical discovery in astronomy to discover that this planet was both the celestial object known as the Evening Star and also the celestial object known as the Morning Star. Consider the state of a competent language user before this empirical discovery (or of a competent language user after the discovery who is unaware of it). Such a person understands the identity statement "The Morning Star is the Evening Star", even though they do not know its truth-value. Frege's point is that if meaning were identified with semantic value, this would be impossible. We can set out his reasoning here as follows:

(i) Suppose (for reductio) that meaning is to be identified with semantic value.
(ii) Understanding a sentence requires understanding its constituents. In other words, knowing the meaning of a sentence requires knowing the meanings of its constituents. So,
(iii) Understanding "The Morning Star is the Evening Star" requires knowing the meanings of, inter alia, "The Morning Star" and "The Evening Star". So,
(iv) Understanding "The Morning Star is the Evening Star" requires knowing the semantic values (references) of "The Morning Star" and "The Evening Star". But,
(v) The semantic value (reference) of "The Morning Star" is the same as that of "The Evening Star": the planet Venus. So,
(vi) Understanding "The Morning Star is the Evening Star" requires knowing that the semantic values (references) of "The Morning Star" and "The Evening Star" are the same: in other words, requires knowing that "The Morning Star is the Evening Star" is true. But
(vii) It is possible to understand "The Morning Star is the Evening Star" without knowing its truth-value. So,
(viii) The meaning of an expression cannot be identified with its semantic value.[4]

So, in giving an account of meaning we are going to have to introduce some semantic property in addition to semantic value, grasp of which constitutes understanding. Again, Frege introduces the property of having a sense to play the role of this semantic property. But what exactly is sense, and how exactly does the introduction of sense enable Frege to solve the three problems we have just outlined? We deal with these questions in the following section.

2.2 The nature of sense

One very important characteristic of sense is spelled out in the following thesis:

Thesis 8: The sense of an expression is that ingredient of its meaning which determines its semantic value.

In addition to having semantic values, expressions also have semantic properties which determine what those semantic values are. The property which determines semantic value is the property of having a certain *sense*. Thus, a name has a reference – stands for a particular object – and also has a sense, some means of determining which particular object this is. Take the case above, of the name "The Evening Star". The sense of this is some condition which an object has to satisfy in order to count as the reference of the name. Perhaps the simplest way of spelling out such a condition would be to specify some *descriptive condition*, like "that object which appears in such and such a place in the sky at such and such times in the evening". If an object satisfies this condition, then it is the reference of "The Evening Star". It turns out, on empirical investigation, that Venus satisfies this condition, so it follows that the name "The Evening Star" refers to the planet Venus. Now, someone who knows which descriptive condition an object has to satisfy in order to count as the reference of "The Evening Star" understands the name; but it does not follow that he knows what the reference of the name actually is. I can know that whatever

object it is that appears at such and such a place in the sky at such and such a time in the evening is referred to by "The Evening Star" without knowing which object that is: I may not have done any astronomy. Thus we also have the following thesis:

Thesis 9: It is possible to know the sense of an expression without knowing its semantic value.

It is this thesis which allows us to solve the problem of informativeness. We first of all set out the additional theses:

Thesis 10: The sense of an expression is what someone who understands the expression grasps.

Thesis 11: The sense of a complex expression is determined by the senses of its constituents.

Consider Frege's Evening Star–Morning Star example. Understanding a sentence requires understanding its constituents. Together with thesis 10, this entails that knowing the sense of a sentence requires knowing the senses of its constituents. So, understanding "The Morning Star is the Evening Star" requires grasping the senses of, inter alia, "The Morning Star" and "The Evening Star". But, from thesis 9, it is possible to grasp the senses of "The Morning Star" and "The Evening Star" without knowing their references. So, from theses 11 and 2, it is possible to grasp the sense of the sentence "The Evening Star is the Morning Star" without knowing its truth-value. So this explains how it is possible to understand a sentence without knowing its truth-value, which is just to explain how sentences can be informative. The introduction of sense thus enables Frege to solve the problem of informativeness.

It might be useful to pause for a moment to see how this account also enables us to explain the *uninformativeness* of "The Evening Star is the Evening Star". In order to do so, we add a further thesis, the thesis of the *transparency of sense:*

> **Thesis 12:** If someone grasps the senses of two expressions, and the two expressions actually have the same sense, then she must know that the two expressions have the same sense.

Suppose someone understands "The Evening Star is the Evening Star". Then, they must grasp the senses of the two occurrences of "The Evening Star". Since each occurrence has the same sense, it follows from thesis 12 that the speaker must know that they have the same sense. But sense determines semantic value (expressions with the same sense must have the same semantic value, from thesis 8), so since the speaker knows they have the same sense, he must also know that they have the same semantic value. From thesis 2 it follows that he must know the truth-value of "The Evening Star is the Evening Star". This explains why the sentence is uninformative.

The introduction of sense also enables Frege to solve the problem of bearerless names. Consider again the example of "Odysseus". Suppose, for the sake of argument, that the sense of this is given by some descriptive condition such as "The hero of Homer's *Odyssey* and the son of Laertes and Antikleia". Clearly, someone can grasp such a condition even if there actually is no object which satisfies it: someone can know what it would be for a person to be referred to by the name "Odysseus" even if there is in fact no such person. This is even clearer in the case of a term like "The twelve-headed student in my class". I can certainly understand this term: this in part explains why I am able to understand the sentence "The twelve-headed student in my class has more than two heads". That is, I know what would have to be the case for someone to be referred to by the term, and I can possess this knowledge even given the fact that (thankfully!) there is no twelve-headed student in my class. Thus,

Thesis 13: An expression can have a sense even if it lacks a semantic value.

Or as Frege puts it himself:

> The words "the celestial body most distant from the Earth" have a sense, but it is very doubtful if they also have a [semantic value]. The expression "the least rapidly convergent series" has a sense but demonstrably there is no [semantic value], since for every given convergent series, another convergent, but less rapidly convergent, series can be found. In grasping a sense, one is certainly not assured of a [semantic value].[5]

Frege thus solves the problem of bearerless names: *names with no semantic value (reference) are not necessarily meaningless, because they can nevertheless possess a sense.*

What follows about sentences containing bearerless names? For example, consider "Odysseus was set ashore at Ithaca while sound asleep". Since one of the expressions, "Odysseus", lacks a reference, and since the semantic value of a complex expression is determined by the semantic values of its parts (thesis 2), it follows that this sentence itself does not have a semantic value. In other words, it lacks a truth-value: it is neither true nor false. This gives us another of Frege's theses:

Thesis 14: A sentence which contains an expression which lacks a semantic value is neither true nor false.

We'll return to the question of the status of sentences containing bearerless names later in this chapter. We must now consider how Frege uses the introduction of sense to solve the problem of substitution into belief contexts.

Recall that substitution into belief contexts appeared to threaten thesis 3, that substitution of co-referential parts of a sentence should leave the truth-value of the whole sentence unchanged, plus the generalisation of that thesis to the case where the parts of the complex sentence are themselves sentences, that the substitution of a component sentence by another sentence having the same truth-value should leave the truth-value of the complex sentence unchanged. For example,

(1) John believes that Mark Twain is Mark Twain

is true, whereas

(2) John believes that Mark Twain is Samuel Clemens

is false, even though the latter was obtained from the former via the substitution of apparently co-referential names.

Frege's response to this is to save thesis 3 by denying that "Mark Twain" and "Samuel Clemens" are indeed co-referential in the relevant sort of belief context. Customarily, outside of belief contexts, they refer to the man who authored *Huckleberry Finn* and so on, as in

(5) Mark Twain was an American.

But within belief contexts they refer to the *senses* they ordinarily possess outwith belief contexts (as in e.g. (5)). Frege expresses this by saying that in "Mark Twain was an American", "Mark Twain" has its *customary reference* and *customary sense*. However, in belief contexts "Mark Twain" refers, not to the man, but rather to the customary sense of the name. *Frege expresses this by saying that in belief contexts a name refers to its customary sense*, and he calls this its *indirect reference*. Since "Mark Twain" and "Samuel Clemens" have different customary senses and therefore different indirect references, it follows that in moving from (1) to (2) we have not actually substituted one co-referential expression for another, so that we do not after all have a counterexample to thesis 3. The identification of indirect reference with customary

sense thus allows us to avoid the problem of substitution into belief contexts. We'll have more to say about Frege's solution to this problem in due course, but for the moment we can sum things up in the thesis:

Thesis 15: In a belief context, the (indirect) reference of a proper name is its customary sense.

It is worth noting that just as thesis 2 led to thesis 3, so thesis 11 (the compositionality of sense) leads to:

Thesis 16: Substitution of one expression in a sentence with another which has the same sense will leave the sense of the sentence unchanged.

This can be used to give us some clue as to what the senses of sentences are. Consider again the move from

(3) John believes that London is south of Glasgow

to

(4) John believes that electrons are negatively charged particles.

Recall that since the semantic value of a sentence is a truth-value, and since the embedded sentences in (3) and (4) have the same truth-values (T), the fact that (3) is true and (4) is false appears to generate a problem for thesis 3. Again, Frege's way round this problem is to apply (an analogue of) thesis 15. In (3) and (4) the embedded sentences do not have their customary semantic values (truth-values): within belief contexts their semantic values are their *customary senses*. But what is the customary sense of a

sentence? Well, what are we referring to in using "London is south of Glasgow" within a belief context such as (3)? Intuitively, we are referring to John's thought *that London is south of Glasgow*. That is to say, we are using the sentence to specify the *content* of a *thought*. Now, if we substitute the embedded sentence in (3) with one which expresses the *same thought* (as opposed to merely having the same truth-value), is it possible for (3) to change truth-value? If not, the identification of the sense of a sentence with the thought it expresses will respect thesis 16. The question as to the identity of thoughts is a thorny one (when does one sentence express the same thought as another sentence?), but intuitively it looks as if the substitution of "London is South of Glasgow" by another sentence expressing the same thought will leave the truth-value of (3) unchanged. For example, the sentence "London is south of Glasgow" intuitively expresses the same thought as "Glasgow is north of London", and the substitution of the latter for the former results in the true sentence

(6) John believes that Glasgow is north of London.

These considerations suggest the next of our Fregean theses:

Thesis 17: The sense of a sentence is a thought.

It is important to note that for Frege, a thought is not something *psychological* or *subjective*. Rather, it is *objective* in the sense that it specifies some condition in the world the obtaining of which is necessary and sufficient for the truth of the sentence which expresses it. This is a theme to which we shall return, but for the moment we note that, in the terminology introduced in §1.6, we can re-express thesis 17 as the view that *the sense of a sentence is its truth-condition*.[6]

Before leaving this section we should pause to note one problem that will have been staring the attentive reader in the face. According to thesis 8, the sense of an expression determines its semantic value. According to thesis 2, the semantic value (truth-value)

of a sentence is determined by the semantic values of its parts. If we put these together we get the result that the sense of an expression determines the truth-values of sentences in which it appears (since the sense of an expression determines its semantic value, which in turn determines the truth-values of sentences in which it appears). This is nicely put in Michael Dummett's characterisation of the notion of sense:

> The sense of an expression is that part of its meaning which is relevant to the determination of the truth-value of sentences in which the expression occurs.[7]

This is problematic because it is uncomfortably close to our definition of the notion of semantic value (§1.3): the semantic value of an expression is that feature of it which determines whether sentences in which it occurs are true or false. Doesn't it follow that we have to *identify* sense and semantic value, so that there is after all no distinction to be drawn between them, so that Frege's theory of meaning, which rests on the distinction, is thrown into chaos?

Dummett points out that there is actually no problem here. Recall the thesis in which we introduced the notion of sense:

Thesis 8: The sense of an expression is that ingredient of its meaning which determines its semantic value.

That is to say, the sense of an expression is that ingredient of its meaning relevant to the determination of the truth-values of the sentences in which it occurs. What does it mean to say that the sense of an expression is "an ingredient of its meaning"? Dummett spells this notion out as follows:

> What we are going to understand as a possible ingredient in meaning will be something which it is plausible to say constitutes part of what someone who understands the word or expression implicitly grasps, and in his grasp of which his understanding in part consists.[8]

In other words, the claim that sense is, in this technical manner of speaking, an ingredient in meaning is more or less a restatement of:

> **Thesis 10:** The sense of an expression is what someone who understands the expression grasps.

Now the key to seeing why there is no tension here with the characterisation of semantic value is to note that semantic value is not, in this technical manner of speaking, an ingredient in meaning. In other words,

> **Thesis 18:** The semantic value of an expression is no part of what someone who understands the expression grasps.

This, indeed, is the upshot of Frege's solution to the problem of informativeness. The argument is that if the semantic value of an expression was part of what was grasped by someone who understands it, there would be no possibility of, for example, understanding a sentence without knowing its truth-value. Thesis 10 and 18 thus ensure that the characterisation of semantic value, together with the characterisation of sense given in thesis 8, do not force the identification of sense with semantic value.

It is worth noting that although semantic value is not in the special way introduced above an ingredient in meaning, it is still part of the intuitive notion of meaning, and something which has to be dealt with in a systematic way by a philosophical theory of meaning. Dummett writes:

> To say that reference [semantic value] is not an ingredient in meaning is not to deny that reference [semantic value] is a consequence of meaning, or that the notion of reference [semantic value] has a vital role to play in the general theory of meaning: it is only to say that the understanding which a speaker of a language has of a word in that language . . . can never consist merely in his associating a certain thing with it as its referent

> [semantic value]; there must be some particular *means* by which this association is effected, the knowledge of which constitutes his grasp of its sense.[9]

We started out attempting to say something systematic about the intuitive notion of meaning. We have now reached the point where we have distinguished between two levels in meaning: sense, that which is grasped by someone who knows the meaning of an expression, and semantic value, that which is determined by sense. This distinction, together with theses 1–18, takes us a long way in our task of saying something systematic about the intuitive notion of meaning. We continue this task in the next section.

2.3 The objectivity of sense: Frege's critique of Locke

We noted above that the sense of a sentence is a thought, and that according to Frege thoughts are in some sense objective, as opposed to subjective or psychological. This is an extremely important part of Frege's position. Indeed, in the introduction to *The Foundations of Arithmetic*, he states the following as the first of his three "fundamental principles":

> Always to separate sharply the psychological from the logical, the subjective from the objective.[10]

This applies not only to the senses of sentences, but to the senses of expressions generally. But what exactly does it mean to say that sense is objective and not subjective? One thing that it means is that grasping a sense – understanding an expression – is not a matter of associating that expression with some subjective item like a *mental image*, *picture*, or *idea*. Frege is quite explicit about the need to distinguish senses, which are objective, from ideas, which are subjective:

> The *Bedeutung* [semantic value] and sense of a sign are to be distinguished from the associated idea . . .
>
> The *Bedeutung* of a proper name is the object itself which we designate by using it; the idea which we have in that case is wholly subjective; in between lies the sense, which is indeed no longer subjective like the idea, but is yet not the object itself.[11]

The view that understanding an expression consists in the possession of some associated idea or image is one that has a long list of adherents in the history of philosophy. In distinguishing the sense of an expression from any associated idea, Frege was directly attacking this tradition, and setting the scene for similar attacks that were later to be mounted by Wittgenstein (see also §5.1). The classic example of this view of sense can be found in Book III of *An Essay Concerning Human Understanding* by the English empiricist philosopher John Locke (1632–1704).

Some creatures who utter, for example, the word "cube" *understand* that word, and some don't. A parrot, for example, can say the word, but unlike a normal human speaker of English, the parrot possesses no understanding of what is said. In Fregean terminology, the human speaker grasps the sense of "cube", whereas the parrot does not. But what does this difference consist in? Locke's suggestion is that the word "cube" is, in the case of the competent human speaker, associated with an *idea of a cube* in that speaker's mind, while in the case of the parrot there is no such idea and so no such association. Locke is thus led to the view that understanding an expression consists in associating it with some idea:

> Words, in their primary or immediate signification, stand for nothing but the ideas in the mind of him that uses them.[12]

Locke takes ideas to be mental images or pictures: an idea of a cube is taken to be a mental image or inner picture of a cube.[13] This is clear from the way Locke speaks throughout the *Essay*. For example, in his account of memory the talk of ideas is explicitly cashed out in terms of picturing and imagery:

> The ideas, as well as children of our youth, often die before us. And our minds represent to us those tombs to which we are approaching; where though the brass and marble remain, yet the inscriptions are effaced by time, and the imagery moulders away. The pictures drawn in our minds are laid in fading colours.[14]

We could thus sum up Locke's view of sense as follows (where the sense of "cube" determines that it refers to, precisely, *cubes*):

> A speaker grasps the sense of "cube" if and only if he is disposed to have a mental image of a cube whenever he hears or utters the word.

Why does Frege object to this account of sense? Locke's account leads to a tension, which the reader will probably have noticed already, between the *public* nature of meaningful language, and the *private* nature of ideas and mental images. On the one hand, language is public in that different speakers can attach the *same* sense to their words, and one speaker can *know* what another speaker means by his words. Different speakers can *communicate* with each other in virtue of the common senses that they have attached to their words. On the other hand, ideas are private. As Locke himself puts it, a man's ideas are "all within his own breast, invisible, and hidden from others, nor can of themselves be made to appear".[15] Also, my ideas, my "internal conceptions", are visible only to my consciousness, and likewise your ideas, your "internal conceptions", are visible only to your consciousness. But we are attempting to give an account of sense, an account that should help explain how we are able to *communicate with each other* via the use of language: and how can a theory which construes grasp of sense in terms of the possession of private inner items help explain our ability to use language in successful public communication?

In fact, Frege develops this rhetorical question into a powerful argument against Locke's account of sense.[16] The argument can be set out as follows:

> (i) One crucial role of sense is to explain how linguistic communication is possible: the success of language in facilitating communication between two speakers is to be accounted for in terms of their grasping the same senses.
> (ii) Private, inner items have no role to play in explaining the practical success of language in facilitating communication between two speakers.

So,

> (iii) Understanding an expression – grasping a sense – cannot consist in the possession of some inner, private item.

Frege is explicit about (i): that Frege thinks sense can play this role is evident in a passage in which he again distinguishes between sense and idea, and refers to the role of sense in communication, or as he puts it, the transmission of thought:

> A painter, a horseman, and a zoologist will probably connect different ideas with the name "Bucephalus". This constitutes an essential distinction between the idea and the sign's sense, which may be the common property of many people, and so is not a part or a mode of the individual mind. For one can hardly deny that mankind has a common store of thoughts which is transmitted from one generation to another.[17]

And it seems perfectly reasonable to demand that any notion of sense be able to play this role. What, though, is the argument for premise (ii)? One way to argue for this is to think about a possibility which Locke himself muses on in the *Essay*, that:

> The same object might produce in several men's minds different ideas at the same time; e.g. the idea that a violet produced in one man's mind by his eyes were the same that a marigold produced in another man's, and vice versa.[18]

Locke realises that this possibility poses a problem for his account of what constitutes understanding. The possibility is one of *systematic inversion* between the ideas of the colours blue and yellow. If we consider two speakers, say Smith and Jones, the possibility is one in which the idea produced in Jones's mind by bananas, egg-yolks, and the skin of someone with jaundice, is of the same type as the idea produced in Smith's mind by the sea on a sunny day, Glasgow Rangers shirts, and city buses in Birmingham; and vice versa, the idea produced in Smith's mind by bananas,

egg-yolks, and jaundiced skin, is of the same type as the idea produced in Jones's mind by the sea, Glasgow Rangers shirts, and Birmingham buses.

The crucial point is this: *since the inversion is systematic in this way, if the possibility were actually to obtain, it would not manifest itself in any of the linguistic behaviour displayed by Smith and Jones*. For example: both would learn to use the word "yellow" by being shown things like bananas, egg-yolks, etc., and both would learn "blue" by being shown the sea on a sunny day, Glasgow Rangers shirts, etc.; other things being equal, both would call the same things blue and yellow in the same circumstances; and other things being equal, both would respond in the same way to orders framed using the words "yellow" and "blue".[19]

Now, if the possibility were to obtain, someone advancing an account of sense along the lines of Locke's would have to say that Smith and Jones were not communicating successfully: after all, they would be attaching different senses to the words "blue" and "yellow", because of the difference in the ideas which these words produce. So, one would expect Locke to provide some argument to the effect that we could always detect, somehow or other, whether the inversion possibility did or did not obtain, perhaps by showing that there is actually some way in which it eventually manifests itself in speakers' behaviour. But, surprisingly, Locke does not attempt to do this, and instead goes on to make the following comment:

> Since this [the systematic inversion] could never be known, because one man's mind could not pass into another man's body . . . neither the ideas hereby, nor the names, would be at all confounded, or any falsehood be in either . . . whatever [the] appearances were in his mind, he would be able regularly to distinguish things for his use by those appearances, and understand and signify those distinctions marked by the names "blue" and "yellow", as if the appearances or ideas in his mind were exactly the same with the ideas in other men's minds . . . [and if the inversion hypothesis were true] it would be of little use, either for the improvement of our knowledge, or conveniency of life; and so we need not trouble to examine it.[20]

But this is an amazing thing for Locke to say, and it effectively amounts to an admission that his account of sense is unable to play any role in the task of explaining linguistic communication. We can see the problems this causes by putting the admission alongside a consequence of his account of sense:

Locke's admission: If the inversion hypothesis were true, it wouldn't really matter, because everything in our linguistic lives would be as before, and we could do everything we previously did with language.

A consequence of Locke's account of sense: If the inversion hypothesis were true, Smith and Jones wouldn't really be communicating, because they would have different ideas annexed to their words, and so would associate different senses with those words.

Putting the admission together with the consequence of the account of sense thus gives us:

It is possible that we could do everything we currently do with language – use language for all of the purposes for which we currently use it – and yet not *really* be communicating, and not *really* attach the same senses to our words.

But this is absurd: there is absolutely nothing in our everyday conceptions of communication and grasp of sense which allows for such a possibility. In particular, the admission of such a possibility effectively renders the notion of sense empty. The conclusion is that Locke's account of sense is unacceptable.[21] We can sum up the results of this section in a further thesis of Frege's:

Thesis 19: Sense is objective: grasping a sense is not a matter of having ideas, mental images, or private psychological items.

Before moving on to look at some problems surrounding the notion of sense, we should pause to spell out a further important thesis of Frege's regarding sense. This thesis will reappear in later chapters of this book. In his late essay "Thought", Frege distinguishes between two different sorts of law: *normative* (or *prescriptive*) and *descriptive*. Scientific laws fall into the latter category: they tell us what will actually happen, given such and such initial conditions. For example, Newton's First Law of Motion tells us that a body will remain at rest or travel with constant velocity in a straight line unless it is acted upon by a net unbalanced set of forces. The law *describes* how the body will behave given these conditions. This contrasts with a moral law, such as "Do not kill". This does not describe how people will actually behave: rather, it lays down a prescription as to how they *ought* to behave. The law constitutes a norm with respect to which people ought to regulate their behaviour: for this reason it is called a *normative* law. Frege makes the distinction between normative and descriptive laws because he wants to stress the point that the laws of logic are laws of thought in the normative sense: logic lays down prescriptions on how we *ought* to reason, on what inferences it is *proper* to make. It does not describe how we *actually* reason, or describe the inferences we will *in fact* make. The important point for our purposes is that sense is also normative: this is hardly surprising, given that the laws of thought are normative, and that the senses of sentences are thoughts. And it is independently plausible. Just as a normative law of conduct tells us how we ought to behave, the sense of an expression must in some sense tell us how we ought to use that expression. The sense of an expression lays down a normative constraint which determines whether a particular use of that expression is *correct* or *incorrect*: someone who uses an expression in a manner which is out of accord with its sense will be deemed to have made a *mistake*. We'll return to the issue of the normativity of sense in due course. For the moment, we can sum things up thus:

Thesis 20: The sense of an expression is normative: it constitutes a normative constraint that determines which uses of that expression are correct and which are incorrect.[22]

2.4 Four problems with Frege's notion of sense

2.4.1 Problems about objectivity

Thesis 19 really only articulates a way in which sense is not subjective: grasping a sense is not constituted by the possession of mental images, and so on. But can we say anything more positive about what the objectivity of sense consists in? Frege tries to say something more in the following passage:

> Somebody observes the Moon through a telescope. I compare the Moon itself to the reference [semantic value]; it is the object of the observation, mediated by the real image projected by the object glass in the interior of the telescope, and by the retinal image of the observer. The former I compare to the sense, the latter is like the idea or intuition [experience]. The optical image in the telescope is indeed one-sided and dependent upon the standpoint of observation; but it is still objective, inasmuch as it can be used by several observers. At any rate it could be arranged for several to use it simultaneously. But each one would have his own retinal image.[23]

The objectivity of sense over images seems to consist in the fact that whereas I cannot literally have your mental image, we can nevertheless literally grasp the same sense. And recall that our grasp of a common stock of senses is supposed to help explain the occurrence of linguistic communication. So Frege needs to characterise sense in such a way that (i) two different individuals can literally grasp the same sense, and (ii) their grasp of this sense helps explain how they can communicate with each other. But does he manage to do this? Let's consider the case of a proper name, such as "Aristotle". So far, the only characterisation of the

sense of a name that we have considered is that given by some *associated descriptive condition*. For example, we could take the sense of "Aristotle" to be given by the description "the pupil of Plato and teacher of Alexander the Great". Frege himself suggests this as one possible sense that might be attached to the name "Aristotle". But in the passage in which he makes this suggestion he makes a concession which is the exact parallel of the concession we saw Locke making in the preceding section. In fact, it turns out that, with the senses of names construed in terms of descriptions, we can aim an argument against Frege which is similar to his own argument against Locke.

Suppose that Smith and Jones associate different descriptions with the name "Aristotle". Smith, say, associates "Aristotle" with "the pupil of Plato and teacher of Alexander", while Jones associates it with "the teacher of Alexander who was born in Stagira". Frege clearly thinks that this is possible:

> In the case of an actual proper name such as "Aristotle" opinions as to the sense may differ. It might, for instance, be taken to be the following: the pupil of Plato and teacher of Alexander the Great. Anybody who does this will attach another sense to the sentence "Aristotle was born in Stagira" than will a man who takes as the sense of the name: the teacher of Alexander the Great who was born in Stagira.[24]

Now what would we expect Frege to say about such a situation, given what we know of his views on sense? Recall that the sense of a sentence is a thought (thesis 17). Since the sense of a complex expression is determined by the senses of its constituents (thesis 11), the thought expressed by "Aristotle was born in Stagira", as uttered by Smith, will differ from the thought it expresses when uttered by Jones. If the sentence is uttered in an exchange between Smith and Jones there will be no transmission of a common thought: in short, there will be no communication. So, given his views on sense, Frege is committed to saying that in the situation envisaged Smith and Jones do not really communicate with each other. But what he actually says is somewhat different. Here is the rest of the passage quoted above:

> So long as the thing referred to remains the same, such variations of sense may be tolerated, although they are to be avoided in the theoretical structure of a demonstrative science and ought not to occur in a perfect language.[25]

Frege seems to be conceding that such variations in sense will make no practical difference as far as the use of language in communication is concerned. But, as with Locke's concession over the systematic inversion possibility, this concession leaves Frege in a precarious position. In fact, it generates the same absurdity:

> It is possible that we could do everything we currently do with language – use language for all of the purposes for which we currently use it – and yet not really be communicating, and not really attach the same senses to our words.

Again, this really only renders the notion of sense empty. But the problem is that it is a consequence of Frege's account of sense together with his concession in the passage quoted above. So it looks as though Frege, like Locke, has not succeeded in capturing a plausible objective notion of sense.

This, though, really only has the status of a challenge to Frege, since we have been working with a crude conception of what the sense of a proper name might be, namely, some associated description. The challenge to Frege is: *explain what sense is in a way that renders it objective and apt to play a part in explaining linguistic communication.*

2.4.2 Problems about sense and analysis

Frege's solution to the problem of informativeness has an odd consequence. Anglo-American philosophy in the 20th and 21st centuries has been dominated by what is known as *analytic philosophy*. One of the main characteristics of analytic philosophy is its view of philosophy as consisting essentially in *analysis*. Analytic philosophers

attempt to tackle the traditional problems of philosophy via the analysis of language: a proper conception of language will enable us either to solve or dissolve these problems.[26] The advances Frege made in logic greatly facilitate the task of solving philosophical problems in this manner:[27] so much so that Frege is often hailed as "the founder of analytic philosophy". But what exactly is analysis? Intuitively, we would want to say that this consists in discovering relationships among *senses*: the analysis of "X knows that P" as "X has a justified, true belief that P" claims to exhibit an identity in sense between the two expressions. Alternatively, we could view identity in sense as corresponding to the intuitive notion of *synonymy*: two expressions have the same sense if and only if they are synonymous. The project of analysis would then be construed as the project of discovering philosophically interesting relations of synonymy, or identity in sense, and the employment of these in the solution or dissolution of philosophical problems.

The difficulty is that Frege's solution to the problem of informativeness appears to rule out this intuitive description of the analytic project. Recall that the problem was to account for the fact that whereas

(7) The Morning Star is the Evening Star

is potentially informative,

(8) The Evening Star is the Evening Star

is not. For Frege, (7) is potentially informative because although "The Morning Star" and "The Evening Star" both refer to the same object, they have different senses. This contrasts with (8), whose uninformativeness is explained by the fact that the two signs flanking the "is" of identity have the same sense. Thus, the informativeness of some identity statements is explained in terms of the signs on either side of the identity sign having the same reference but expressing different senses; while the uninformativeness of other identity statements is explained in terms of the signs on either side of the identity sign having the same sense as well as the

same reference. *The problem with this is that given the assumption that analysis concerns relations between senses, it also entails that there can be no such thing as an informative analysis.*[28] To see this, suppose that analytic philosophers have succeeded in showing by analysis that knowledge is justified true belief. Then consider

(9) Knowledge is justified true belief.

Suppose that analysis really is concerned with discovering relations in the realm of sense, and that (9) is a good analysis. Then consider a philosopher who understands (9). Since he grasps the sense of (9) he must also grasp the senses of its constituents. Since "knowledge" and "justified true belief" ex hypothesi have the same sense, it follows from thesis 12 that anyone who grasps (9) must know that they have the same sense. But sense determines reference (thesis 8), so since the philosopher knows that they have same sense, he must also know that they have the same reference. From thesis 2, it follows that he must know the truth-value of (9). So, any philosopher who understands (9) must know its truth-value. But this is just to say that there can be no such thing as an informative and interesting analysis. And if there can be no such thing as an interesting and informative analysis, what becomes of the project of analytic philosophy?[29] The dilemma for Frege here is clear. Either he gives up his explanation of why (7) and (8) differ in their potential informativeness, or he renders himself unable to account for the possibility of informative analyses, like that encapsulated in (9). The only way of avoiding this dilemma would seem to be to reject the intuitive conception of analysis as a project concerning relations of sense. This may be possible, but it would further deepen the mystery as to what sense actually is.[30]

2.4.3 Problems about indirect reference and belief contexts

Recall that Frege's solution to the problem of substitution into belief contexts is summed up in:

Thesis 15: In a belief context, the reference of a proper name is its customary sense.

Thus, in

(2) John believes that Mark Twain is Samuel Clemens

"Mark Twain" refers, not to the celebrated American author, but rather to the sense of "Mark Twain", as it appears outside of belief contexts. As we saw, Frege expresses this point in the following way: in belief contexts expressions refer to their customary senses, or, the indirect reference of an expression is its customary sense.

Now consider the following example of a perfectly well-understood sentence:

(10) James believes that John believes that Mark Twain is Samuel Clemens.

It is natural to suppose that just as the occurrence of "Mark Twain" in an indirect context such as (2) refers to the sense it possesses in a direct context such as (5), the occurrence of "Mark Twain" in a *doubly* indirect context such as (10) will refer to the sense which it possesses in a *singly* indirect context such as (2). So, on Frege's account, the name "Mark Twain", as it appears in (10) refers to the *sense* that it possesses in (2), what Frege calls its *indirect sense*.

But what is this indirect sense? All we know is the reference of "Mark Twain" as it appears in (2), namely, its customary sense. But this won't allow us to work out the sense of "Mark Twain" as it appears in (2), since the sense–reference relation is many–one: the same referent can be associated with indefinitely many senses (as Russell put it, "there is no route back from the reference to the sense"[31]). The upshot of this is that we seem to be unable to say what the sense of "Mark Twain" in (2) is, and therefore unable to say what the reference of "Mark Twain" is in (10). Given that the reference of an expression is what determines the contribution it makes to the truth-values of sentences in which it appears, we should be unable to appraise sentences involving doubly indirect

contexts such as (10) for truth or falsity. But, of course, we are perfectly capable of making such appraisals: so Frege's account of indirect sense and indirect reference is thrown into doubt.

Is there any way Frege can get round this problem? Why does he need to distinguish between the sense of "Mark Twain" in an indirect context such as (2) and a direct context such as (5) in the first place? In other words, why can't he just *identify* the indirect sense of "Mark Twain", the sense it possesses in (2), with its customary sense, the sense it possesses in (5)? Given that we know the sense of "Mark Twain" as it appears in (5) (or at least we are supposing that we do, for the sake of argument), this would solve the problem about the reference of "Mark Twain" as it appears in the doubly indirect sentence (10). As Dummett points out, however, this option is not straightforwardly available to Frege because of his adherence to thesis 8, that sense determines semantic value. It follows from thesis 8 that if an expression has a certain reference in one context, but a different reference in another, then that expression must have different senses in each of the two contexts. Now, given that "Mark Twain" as it appears in (5) refers to the celebrated American author, and as it appears in (2) refers to its customary sense, it follows that "Mark Twain" has a different reference in each of these, and hence must express a different sense in each. In other words, the sense of "Mark Twain" in (2) is different from the sense it possesses in (5). The identification of indirect sense with customary sense is frustrated, so that we are left in the dark as to what the indirect sense, and hence the doubly indirect reference of "Mark Twain", actually is.

Dummett suggests that the whole difficulty arises from interpreting thesis 8 as meaning that the semantic value of an expression must be determined by its sense *alone*. Dummett claims that this is a mistake, based on a misleading tendency to speak about the semantic values of expressions in isolation from the sentential contexts in which they occur. Indeed, Frege explicitly counsels against this tendency in the second of the three fundamental principles set out in the introduction to *The Foundations of Arithmetic*:

> Never to ask for the reference of a word in isolation, but only in the context of a proposition.[32]

Dummett interprets this as a claim that only a particular occurrence of an expression in a sentence has a semantic value, and that this semantic value is determined *jointly* by the sense of the expression *together* with the kind of *context* in which it occurs. Then, as Dummett puts it,

> The sense of a word may thus be such as to determine it to stand for one thing in one kind of context, and for a different thing in some other kind of context. We may therefore regard an expression occurring in an indirect context as having the same sense as in a direct context, though a different reference.[33]

Frege can thus after all equate the indirect and customary senses of "Mark Twain", putting the differences in reference in (2) and (5) down to the fact that the expression is appearing in different types of context in each.

How plausible is this solution to the difficulty? There are at least two points that one can make here. Firstly, Dummett's suggestion that the difficulty can be solved by equating the indirect and customary senses of "Mark Twain" appears to be *ad hoc*. All Dummett has shown is that there is *nothing to prevent* Frege from identifying indirect and customary sense. But in order to convincingly get Frege out of trouble here, we surely require some *positive* grounds for identifying customary and indirect sense, grounds that go beyond the observation that we avoid the relevant difficulty if we make the identification. Secondly, this point is made all the more pressing by the fact that there are actually some reasons *against* making the identification. Suppose that we do equate customary and indirect sense, and consider two expressions which arguably have the same customary sense, say, "chiropodist" and "foot-doctor". Then, it follows that "chiropodist" and "foot-doctor" have the same indirect sense, and therefore the same reference in doubly indirect contexts. Now it follows, from thesis 3, that the move from

(11) John believes that Frank believes that all chiropodists are chiropodists

to

(12) John believes that Frank believes that all chiropodists are foot-doctors

should not result in a change in truth-value. But of course it is easy enough to think of cases in which (11) is true yet (12) is false. So the identification of customary and indirect sense threatens a clash with one of Frege's fundamental theses about semantic value. The upshot is that Frege's solution to the problem of substitution into belief contexts is not entirely satisfactory.

2.4.4 Problems about bearerless names

We suggested in §1.8 and §1.9 that one of the reasons for the introduction of the notion of sense was that it gave us a way of accounting for the fact that sentences containing bearerless names are nevertheless meaningful. Names without a bearer are not meaningless, since they can nevertheless possess a sense. The key thesis is therefore:

Thesis 13: An expression can have a sense even if it lacks a semantic value.

Some philosophers have questioned whether this thesis is so much as coherent, given the technical notion of semantic value which Frege is working with. Recall that semantic value was defined in the following way:

Definition: The semantic value of an expression is that feature of it which determines whether sentences in which it occurs are true or false.

But can we really make anything of the idea that an expression can have sense even though no feature of it determines whether

sentences in which it occurs are true or false? The difficulty is perhaps clearest when we consider the case in which the expressions dealt with in thesis 13 are whole sentences. Recall that a sentence expresses a thought as its sense and has a truth-value as its semantic value. So thesis 13, in the case of sentences, amounts to the claim that a sentence can express a thought even though it has no truth-value. The intuitive implausibility of this claim is well brought out by Gareth Evans:

> What can it mean on Frege's, or on anyone's principles, for there to be a perfectly determinate thought which simply has no truth-value? . . . If someone understands and accepts a sentence containing an empty name, then, according to Frege, he thereby forms a belief; not a belief about language, but a belief about the world. But what sense can be made of a belief which literally has no truth-value?[34]

Doesn't this rob the technical notion of semantic value of any importance we might have thought it possessed for an account of the meaningfulness of language? How can we introduce a technical notion like semantic value, the possession of which by expressions is supposed to contribute to an explanation of their meaningfulness, and then allow that expressions can be meaningful even if they fail to possess it? As Evans puts it

> The semanticist seeks to account [for the fact that a certain body of discourse is significant]. Following Frege, as part of this procedure he decides to construct a theory of semantic value, the main aim of which is to help to explain how the significance of sentences depends upon the significance of their parts. But it is just not open to the semanticist to say "There is a gap in my theory; here is a group of viable sentences which might be used to express and transmit thoughts, but to which my theory does not apply".[35]

This is a problem which Frege will have to face head on, if he accepts that the semantic value of a name is the object it stands for, and that sentences containing bearerless names may nevertheless express a thought.[36]

In fact, Evans argues for a "Russellian" interpretation of Frege on which he is not advancing thesis 13.[37] Evans's interpretation emphasises a comment Frege makes about sense that we have not so far mentioned:

> It is natural, now, to think of there being connected with a sign (name, combination of words, written mark), besides that which the sign designates, which may be called the [semantic value] of the sign, also what I should like to call the sense of the sign, wherein the mode of presentation is contained.[38]

This suggests a supplement to thesis 8:

Thesis 8a: The sense of an expression is the mode of presentation of its referent.

Evans cashes out the notion of a mode of presentation in terms of the notion of a "way of thinking" of an object, and this forces us in the direction of rejecting thesis 13: if there is no object you're thinking of, so no semantic value that is being presented, then there is no sense. However, it can be argued that there is an alternative construal of "mode of presentation" on which thesis 13 remains intact.[39] The issues here are complicated and, needless to say, extremely controversial.

2.5 Kripke on naming and necessity

In his highly influential book *Naming and Necessity*, Saul Kripke raised a number of objections against what he construed as Frege's views on the sense and semantic value of proper names.[40] Thus far, whenever we have spoken of the sense of a name, we have taken it to be some descriptive condition: an object is the referent of the name if and only if the description is true of it. Thus, we can take the sense of "Aristotle" to be given by the description "the pupil of Plato and teacher of Alexander the Great". Kripke argues that if

we accept this as an account of the sense of "Aristotle", we end up saying some very implausible things about Aristotle. In order to spell out Kripke's argument we need to introduce the distinction between *necessary truth* and *contingent truth*.

Consider a true sentence such as "2 + 2 = 4". Could things have turned out in such a way that this would have been false? It seems not: there are simply no possible situations in which the sum of 2 and 2 is not 4.[41] Likewise for "All bachelors are unmarried": there is no way there could have been a married bachelor. Philosophers attempt to capture this feature of "2 + 2 = 4" and "All bachelors are married" by saying that they are *necessary truths*. Sometimes this notion is glossed as follows: a necessary truth is one which is true, not only in this, the actual world, but also in all logically possible worlds. "2 + 2 = 4" and "All bachelors are unmarried" contrast with "Trump was the President in 2017" or "Germany won the 2014 football World Cup". Although these are actually true, we can conceive of logically possible situations in which they are false: there are logically possible worlds in which Trump lost the 2016 election, and so would not have been President in 2017, and there are logically possible worlds in which some team other than Germany won the 2014 World Cup (e.g. Scotland). "Trump was President in 2017" and "Germany won the 2014 World Cup" are *contingent truths*. They are true in the actual world, but not in all possible worlds.

There is another distinction which is related to the distinction between necessary and contingent truth, but which must also be distinguished from it. An *analytically true* sentence is one which is true purely in virtue of the senses of its constituents. "All bachelors are unmarried" would be an example of an analytic truth: its truth is settled by the senses of "bachelor", "unmarried", and so on. An example of a *synthetic* truth would be "Obama was President of the USA in 2016". This is true, but not *purely* in virtue of the senses of the constituent expressions. The relationship between the necessary–contingent and analytic–synthetic distinctions is a matter of some controversy. Philosophers are generally agreed that all analytic truths are necessary, but not all philosophers hold that all necessary truths are analytic. Kant (1724–1804), for example, held that arithmetical truths such as "2 + 2 = 4" are necessary but synthetic.[42]

Kripke's objection to taking the sense of "Aristotle" to be given by a description like "the pupil of Plato and teacher of Alexander the Great" is quite simple: if we take the sense of the name to be given by this description, then certain sentences that are only contingently true turn out to be necessarily true. Suppose that the sense of "Aristotle" is given by "the pupil of Plato and teacher of Alexander the Great". Then consider the sentence

(13) Aristotle was the pupil of Plato and the teacher of Alexander the Great.

Since the sense of "Aristotle" is *given* by "the pupil of Plato and the teacher of Alexander the Great" this sentence is true, and, moreover, true purely in virtue of facts about sense. In other words, it is analytically true. But if it is analytically true, it is also necessarily true, true in all logically possible worlds. But surely it is at most a contingent truth: we have no trouble in conceiving of situations in which Aristotle was taught by someone other than Plato and did not himself teach Alexander the Great.

So, if the sense of "Aristotle" is given by the description, (13) would be a necessary truth. But (13) is not a necessary truth, it is at most a contingent truth. So the sense of "Aristotle" cannot be given by the description.[43]

Kripke argues that this is a consequence of the fact that ordinary proper names, unlike definite descriptions, are what he calls *rigid designators*: where a rigid designator is an expression which refers to the same thing in every possible world in which that thing exists. Definite descriptions are *non-rigid designators* since their reference can change from possible world to possible world: "the US president in 2016" in one world refers to Barack Obama, in another to Joe Biden (imagine that Obama got run over by a bus in late 2015), and in yet another to Benny Hill. So it is unlikely that the sense of an ordinary proper name could be given by some complex definite description.

Kripke introduces an important distinction which may allow Frege to deflect this objection. He distinguishes between the idea that a definite description *gives the meaning* of a proper name and the idea that a definite description *fixes the reference* of a

proper name. Suppose the view is not that "Aristotle" is *synonymous* with "the pupil of Plato and the teacher of Alexander the Great" (so that the latter would give the meaning of the former), but instead that the reference of "Aristotle" is fixed via the property of being the pupil of Plato and the teacher of Aristotle. The name "Aristotle" refers, in the actual world, to whoever possesses that property; but the property is a *contingent* property of whoever possesses it. Since we are not taking the definite description to give the meaning of the name, it doesn't follow that the individual concerned has to possess that property in all (or indeed any) of the various non-actual possible worlds. So if we view Frege as claiming only that definite descriptions fix the reference of ordinary proper names rather than give their meanings, Kripke's initial objection appears to be neutralised.

Although Kripke objects to the idea that *in general* the reference of an ordinary proper name is fixed by some associated definite description, he does think that there are cases where it is plausible to view reference as fixed by a definite description. His view of these cases has some important consequences. We introduced above the concepts of necessity and contingency. Another important distinction is that between *a priori* and *a posteriori* truths. An *a priori* truth is one that can be known independently of sense-experience, an *a posteriori* truth is one that cannot be known independently of sense-experience. Kripke points out that

> In contemporary discussion very few people, if any, distinguish between the concepts of statements being *a priori* and their being necessary.[44]

Against this identification Kripke argues that the concepts of necessity and contingency are *metaphysical* concepts, whereas the concepts of the *a priori* and *a posteriori* are *epistemological*: the latter are concerned with our *knowledge*, whereas the former are not. In fact, Kripke argues that there may be statements which, although only contingently true, may nevertheless be *a priori*. Take the standard metre-rod, S, in Paris. Suppose we use the description "the length of S at t" to fix the reference of "one metre". Since there are possible worlds in which the length of S was different at time t, the statement "S is one metre at t" expresses only a

contingent truth: but because we have used the length of S at t to fix the reference of "one metre" at t, the sentence is *a priori* true. So "S is one metre at t" expresses a *contingent a priori truth*.

In general, however, Kripke thinks that we cannot view definite descriptions as fixing the reference of ordinary proper names. According to this sort of view a speaker would be viewed as referring to an individual in virtue of that individual's uniquely satisfying some complex definite description, although there is now no suggestion that the definite description gives the meaning of the name. Kripke uses two types of example to show that, in general, this is not a correct account of how the reference of names gets fixed. First of all, the fact that an individual satisfies a complex definite description associated with an ordinary proper name is not *sufficient* to ensure that someone using the name refers to that individual. Take the name "Gödel", and suppose that its reference is fixed by the definite description "the man who proved the incompleteness of arithmetic". Kripke invites us to consider the following (purely fictional) sort of situation:

> Suppose that Gödel was not in fact the author of [the proof that arithmetic is incomplete]. A man named "Schmidt", whose body was found in Vienna under mysterious circumstances many years ago, actually did the work in question. His friend Gödel somehow got hold of the manuscript and it was thereafter attributed to Gödel. On the view in question, then, when our ordinary man uses the name "Gödel", he really means to refer to Schmidt, because Schmidt is the unique person satisfying the description, "the man who discovered the incompleteness of arithmetic".[45]

If the reference of "Gödel" were fixed by the definite description, then in this scenario the ordinary man would actually be referring to Schmidt and not Gödel, because it is Schmidt and not Gödel who uniquely satisfies the description. Kripke finds this consequence implausible:

> [S]ince the man who discovered the incompleteness of arithmetic is in fact Schmidt, we, when we talk about "Gödel", are in fact always referring to Schmidt. But it seems to me that we are not. We simply are not.[46]

In addition, Kripke argues that it is not *necessary* for an individual to uniquely satisfy a definite description associated with a name in order for the individual to be referred to by uses of that name. Take the name "Richard Feynman". The ordinary speaker, it seems, can certainly use this name to refer to Richard Feynman, as when I say to my father "Richard Feynman will be on television this evening". But there seems to be no description, associated by me with the name, which Feynman uniquely satisfies. The same thing goes for the name "Cicero":

> In fact, most people, when they think of Cicero, just think of a famous Roman orator, without any pretension to think either that there was only one famous Roman orator or that one must know something else about Cicero to have a referent for the name. Consider Richard Feynman, to whom many of us are able to refer. He is a leading contemporary theoretical physicist. Everyone *here* [Kripke is speaking in a seminar at Princeton University in 1970] (I'm sure!) can state the contents of one of Feynman's theories so as to differentiate him from Gel-Mann. However, the man in the street, not possessing these abilities, may still use the name "Feynman". When asked he will say: well he's a physicist or something. He may not think that this picks out anyone uniquely. I still think he uses the name "Feynman" as a name for Feynman.[47]

So according to Kripke it is neither sufficient nor necessary for an ordinary proper name to refer to an individual that the individual in question satisfy some definite description associated with the name: in general, the reference of a proper name is not fixed by some associated definite description.

Just as Kripke argues that there may be contingent *a priori* truths, he argues also that there may be necessary *a posteriori* truths. This follows from his claim that many ordinary proper names are rigid designators, in other words stand for the same individual in every possible world in which that individual exists. Then, for example, "Hesperus is Phosphorus" if it expresses a truth, expresses a necessary truth, even though there is no way, other than empirical investigation, of establishing that it is actually true.[48]

The problems raised in this and the previous section give an indication of the sorts of difficulties that a defender of the notion

of sense must attempt to overcome. It is worth noting that some of them turn on taking the senses of proper names to be given by descriptions. Some philosophers have pointed out, however, that although Frege sometimes speaks as if the sense of a name is given by a description, there is nothing in the notion of sense which forces him to do so. Dummett, for example, writes:

> In trying to say what the senses of different names may be, Frege is naturally driven to citing such definite descriptions: but there is nothing in what he says to warrant the conclusion that the sense of a proper name is always the sense of some complex description. All that is necessary, in order that the senses of two names which have the same referent should differ, is that we should have a different way of recognizing an object as the referent of each of the names: there is no reason to suppose that the means by which we effect such a recognition should be expressible by means of a definite description.[49]

But what *is* sense, if it is not given by definite descriptions? A clue is perhaps provided by thesis 20, the claim that sense is normative. A norm is another term for a *rule*: rules are normative in that they lay down constraints on what counts as correct or incorrect behaviour. So perhaps we can take senses to be simply rules governing the use of expressions, which determine their semantic values and which are grasped by those who understand the expression in question. We can then leave it open whether these rules, in the case of, for example, names, would be spelled out in terms of definite descriptions. But what are these rules? Can they be spelled out in a way which respects all of the various theses about sense that we have attributed to Frege? And can Frege respond to the problems we have raised in this section? We cannot attempt to deal with all of these questions in this book: this is perhaps the starting point for further study of Frege's notion of sense. But we'll return to at least some of them in due course.[50]

2.6 A theory of sense?

In §1.7 we gave an example of a semantic theory for a simple language: a theory which shows how the *truth-values* of complex

sentences are systematically determined by the semantic values of their constituent names and predicates. We also suggested that this theory could be used to show how statements of the *truth-conditions* of complex sentences can be derived on the basis of assignments of references and extensions to their parts. Recall that thesis 17 – the sense of a sentence is a thought – can be re-expressed as the claim that the sense of a sentence is its truth-condition. What this suggests is that a semantic theory of the sort described in §1.7 – which in the first instance shows how the semantic values of complex expressions are determined by the semantic values of their parts – might also be used as a *systematic theory of sense*. There would be no need for a separate systematic theory of sense over and above the systematic theory of semantic value. Or, alternatively, we could provide an account of how the senses of complex expressions are determined without having to directly ascribe senses to their parts. This idea has seemed attractive to a number of philosophers. For example, Gareth Evans writes

> Frege nowhere appears to have envisaged a theory which would entail, for any sentence of the language, S, a theorem of the form,
>
> The sense of S is . . .,
>
> derived from axioms which would state the sense of the primitive words of the language. Frege had no more idea of how to complete a clause like
>
> The sense of "and" is . . .
>
> than we do.[51]

We shall see later when we discuss Davidson's ideas on theories of meaning (Chapter 8) that it is just as well that Frege did not envisage this type of theory of sense, since it appears to be hopeless. The important point for present purposes is that on the basis of assignments of references and extensions to the names and predicates of our simple language, we can systematically generate statements of the truth-conditions, or senses, of the complex sentences in which they appear.

Doesn't this show that a systematic theory of sense – of truth-conditions – doesn't require us to ascribe senses to the subsentential expressions of the language with which it deals, since the senses of sentences are generated purely on the basis of axioms spelling out the *semantic values* of the subsentential expressions? This would be a mistake. The axioms specify the semantic values of the subsentential expressions, but they must specify them in a way which reflects how they are determined by the senses the competent speakers of the language associate with them. The idea is that the axioms of the semantic theory display the senses of the subsentential expressions, even though they do not explicitly state them. Dummett puts the point as follows:

> In the case in which we are concerned to convey, or stipulate, the sense of an expression, we shall choose the means of stating what the referent [semantic value] is which displays the sense; we might here borrow a famous pair of terms from the *Tractatus*, and say that, for Frege, we say what the referent [semantic value] of a word is, and thereby show what its sense is.[52]

2.7 Force and tone

Thus far we have been attempting to tell a story about meaning in terms of two semantic properties, reference and sense. In addition to these, Frege introduces two other semantic properties, *force* and *tone*.

We'll take tone first. Consider the two sentences

(14) Beckham is English and he is a good loser
(15) Beckham is English but he is a good loser.

Intuitively, (14) and (15) have same truth-value (let's take it to be T, for the sake of argument). Substituting "but" for "and" does not lead to a change in truth-value. But equally intuitively, there is also some sort of difference between (14) and (15). In order to capture this difference, Frege introduced the notion of tone (he sometimes also refers to this as "colouring" or "illumination").

Frege held that "and" and "but" have the same sense, but different tones. Like sense, tone is an ingredient in meaning: the tone of an expression is part of what is grasped by someone who is competent with it. As Dummett puts it:

> A mistake about the . . . tone . . . intended to be understood as attached to a sentence or expression would ordinarily be accounted a misunderstanding of its meaning.[53]

But unlike the sense of an expression, its tone is not relevant to determining the truth-values of sentences in which it appears. This is shown by the fact that the truth-tables for (14) and (15) are identical:

P	Q	P and Q	P but Q
T	T	T	T
T	F	F	F
F	T	F	F
F	F	F	F

Other examples of expressions which have the same sense but which differ in tone are: "dog" and "cur; "sweat" and "perspiration"; "walk" and "perambulate".[54]

The notion of sense, as developed by Frege, corresponds to a notion of *sentence-meaning*. The sense of an expression is intended to capture what a sentence strictly and literally means: in other words, the sense of an expression gives its *literal meaning* (or what it means, when stripped of its tone). Take a sentence such as "Jones makes an excellent cup of tea". The strict and literal meaning of this sentence is *that Jones makes an excellent cup of tea*. That is, the truth-condition of this sentence is given by

(16) "Jones makes an excellent cup of tea" is true if and only if Jones makes an excellent cup of tea.

But what about the countless sentences of language that do not have truth-conditions, but which are nevertheless perfectly

meaningful? For example, neither of "Does Jones make an excellent cup of tea?" or "Jones, make an excellent cup of tea!" are the sorts of sentences that *have* truth-conditions. So far, in our account of sense and semantic value, we have considered only sentences which are in the *indicative* grammatical mood, and we have suggested that the sense of these sentences is given by their truth-conditions. But there are other grammatical moods: the *interrogative* mood, exemplified by "Does Jones make an excellent cup of tea?", and the *imperatival* mood, exemplified by "Jones, make an excellent cup of tea!". So what can we say about the literal meanings of sentences in these other moods?

Frege's idea is to represent the meaning of a sentence by an ordered pair, consisting of a sense, together with an indication of force. Since we already have a story about the senses of sentences in the indicative mood, we shall use this as a basis for giving an account of the meanings of sentences in other, non-indicative moods. This is what Frege thinks ought to be done (he never got so far as actually doing it in any detail). Consider the three sentences

(17) Jones makes an excellent cup of tea
(18) Does Jones make an excellent cup of tea?
(19) Jones, make an excellent cup of tea!

The sense of (17) is the thought *that Jones makes an excellent cup of tea*. Now we can do a number of things with the thought that Jones makes an excellent cup of tea. We can assert it, we can ask whether it is true, and we can command that it be made true. Each of these corresponds to a different force that might be attached to the thought: the force of an assertion, the force of a question, the force of a command. We can represent the meanings of (17)–(19) as

(Jones makes an excellent cup of tea, force of an assertion)
(Jones makes an excellent cup of tea, force of a question)
(Jones makes an excellent cup of tea, force of a command).

Thus, we can give an account of the literal meanings of many types of sentences in terms of the notion of sense developed for indicative sentences together with the additional notion of force.[55] We can now formulate our final Fregean thesis:

Thesis 21: In addition to sense and semantic value, we must also introduce the notions of *force* and *tone*.

Note that force and tone are invoked as part of an account of sentence-meaning.[56] It is usual to distinguish between sentence-meaning and *speaker's-meaning*. Speaker's-meaning concerns what information the utterer actually intends to convey. It is clear that the speaker's-meaning and the literal meaning of a sentence can come apart. Consider our sentence (17). As we noted, its sentence-meaning is *that Jones makes an excellent cup of tea*. But imagine I am having a conversation with a visitor to my department who asks, of my colleague, Jones, whether he is an interesting philosopher. If I respond by uttering the sentence (17), what information do I intend to convey to the visitor? Roughly, I intend to convey the information that Jones is an uninteresting philosopher. Thus, I can use a sentence which literally means that Jones makes an excellent cup of tea to assert that Jones is an uninteresting philosopher. Another way of putting this would be to say that I can utter sentence (17) to perform the *speech-act* of asserting that Jones is an uninteresting philosopher.[57] Obviously, there are many different types of speech-act: giving commands, asking questions, and so on. Note that I can even perform the speech-act of asserting that Jones is an uninteresting philosopher by uttering a sentence which is non-indicative in mood: for example, I could reply to my visitor's question with either "Are you kidding?" or "Pull the other one!".

Given that a sentence has a particular literal meaning, what determines its speaker's-meaning, or the particular speech-acts its user performs on a given occasion of utterance? Answering this question is the province of *pragmatics*. Very roughly, we can say that speaker's-meaning is determined in part by the *context of utterance*. Whether we can say anything more systematic than this is an extremely difficult question. For the moment, we'll sum things up by saying that pragmatics concerns the determination of speaker's-meaning, while semantics concerns the determination of sentence-meaning or literal-meaning.[58]

We can sum up our account of Frege as follows. For Frege, whether or not a sentence is grammatically well-formed is determined by syntactical rules. These are the province of *syntax* (§1.2). Whether or not it is true is determined by the references and extensions of its constituents (§§1.3–1.7). The references and extensions of its constituents are determined by their senses, which in turn determine the sense (truth-condition) of the sentence (§§2.1–2.2). In addition to sense, we must also discern force and tone as ingredients in meaning (§2.6). Sense, force, and tone together determine *sentence-meaning*, what a sentence literally and strictly means. All this belongs to the province of *semantics*. Given the sentence-meaning which belongs to the sentence, the context of utterance determines the speaker's-meaning, or speech-acts actually performed by the utterance of the sentence. The story as to how sentence-meaning and context of utterance jointly determine speaker's meaning is the province of *pragmatics*.[59]

2.8 Russell on names and descriptions

Bertrand Russell (1872–1970), in a series of famous works, challenged some of the views of Frege's that we have been outlining. In fact, Russell raised two distinct sorts of criticism of Frege. On the one hand, he accepted, in broad outline, the account of semantic value that we outlined in Chapter 1, but argued that Frege had gone wrong on some important points of detail. On the other, he attacked Frege's introduction of the notion of sense over and above that of semantic value. We'll consider criticisms of the first sort in this section and the next, and criticisms of the second sort in §2.10. This will allow us to outline Russell's famous "theory of definite descriptions".

In §1.4, we saw that Frege held the following thesis:

Thesis 4: The semantic value of a proper name is the object which it stands for.

We could also view this as a *definition* of "proper name": proper names are precisely those expressions which have objects as their semantic values. It would then be an open question whether names, as they appear in natural languages such as English, are proper names as thus defined. Frege clearly held that ordinary names, such as "Aristotle" and "Odysseus", are proper names in this sense. He also held that *definite descriptions* are proper names in this sense. A definite description is a phrase of the form "the so and so". So examples of definite descriptions would be "the King of France", "the man in the iron mask", "the celestial body most distant from the earth", and "the least rapidly convergent series". Frege thus held that the semantic value of a definite description is the object which it stands for: the contribution which a definite description makes to the truth-values of complex sentences in which it appears is determined by the fact that it stands for a certain object.

Recall that one of the problems we raised for Frege in §2.4 was the problem of bearerless names. The easiest way into Russell's views is to see him as starting from essentially the same problem. Take a definite description such as "the King of France". There is simply no object for which this stands, so if we view it as a proper name, it follows that it has no semantic value. But then there is no property in virtue of which it makes a contribution to the truth-values of complex sentences in which it appears. Thus, on Frege's view, the sentence

(20) The King of France is bald

has no truth-value: it expresses a thought which is neither true nor false. We saw that this notion is only dubiously coherent. One way of avoiding the problem was suggested by the Austrian psychologist and philosopher Alexius Meinong (1853–1920): the expression "the King of France" *does* refer to an object, but a non-existent one. Russell attempts to avoid the problem in a way which does not involve the highly implausible postulation of non-existent objects. *Rather, Russell attempts to avoid the problem by denying that definite descriptions are proper names*. If definite descriptions are not proper names, their semantic values are not given by the objects (if any) which they stand for, so the fact that there are

many definite descriptions which do not stand for any object does not lead to the problem faced by Frege. But if the semantic behaviour of definite descriptions is not to be explained in terms of their standing for objects, how is it to be explained? What are definite descriptions if they are not proper names?

In order to see Russell's answer to these questions, we should think about the correct translation of (20) into Frege's logical language. (20) has the grammatical form of a subject-predicate sentence. So if we were taking grammatical form as a guide to the correct translation of (20) into logical symbolism, it would come out as

(21) Fa.

A property, being bald, represented by "F" is predicated of the King of France, represented by "a". Now we can say that in translating a sentence into Frege's logical symbolism we are attempting to capture its *logical form*: we exhibit the form of the sentence in such a way that the contributions its constituents make to its truth-value are thereby exhibited. Thus, in translating (20) by (21) we are suggesting that "the King of France" contributes to determining the truth-value of (20) in the manner of a proper name, and "bald" contributes in the manner of a predicate. *Russell suggests that in this instance we should not take the grammatical form of the sentence as a guide to its logical form*. In fact, the logical form of (20) is given by something quite different from (21). Russell analyses (20) by first breaking it up into three different parts:

(i) There is at least one King of France.
(ii) There is at most one King of France.
(iii) Anything which is a King of France is also bald.

The conjunction of (i), (ii), and (iii) amounts to the claim that the King of France is bald. Thus, the logical form of (20) can be represented by the translation of the conjunction of (i), (ii), and (iii) into Frege's logical symbolism. Taking "F" to abbreviate the predicate ". . . is a King of France"[60] and "G" to abbreviate the predicate ". . . is bald", this comes out as

(22) $(\exists x)((Fx \,\&\, Gx) \,\&\, (\forall y)(Fy \rightarrow x = y))$.

Representing the logical form of (20) by (22) does not involve representing "the King of France" as a proper name, as in (21). We can represent the logical form of sentences in which definite descriptions appear without viewing those descriptions as having objects as their semantic values. But if definite descriptions do not have objects as their semantic values, what are their semantic values? *The answer to this is that definite descriptions have second-level functions as their semantic values*. In (20), the definite description "the King of France" is translated by

(23) $(\exists x)((Fx \,\&\, \ldots x) \,\&\, (\forall y)(Fy \rightarrow x = y))$,

which stands for a function from concepts (first-level functions) to truth-values. Russell's criticism of Frege's view of definite descriptions can thus be summed up as follows. *Frege assigns definite descriptions the wrong sort of semantic values: Frege views definite descriptions as proper names, as having objects as their semantic values; but in fact definite descriptions have second-level functions as their semantic values*.[61]

Note how this solves the problem of bearerless definite descriptions. Since definite descriptions are not proper names, but rather functional expressions which do not have objects as their semantic values, the failure of a definite description to stand for an object does not imply that it has no semantic value: we still have an account of how it contributes to the truth-values of complex sentences in which it appears, an account which runs along the lines of that given for functional expressions in Chapter 1. What, then, is the truth-value of (20)? Given the analysis into (22) we can see that it is false, since $(\exists x)Fx$ is false (since there is no King of France).

In fact, Russell attempts to avoid the problem of bearerless names in exactly the same way, *by treating ordinary names as disguised definite descriptions*. That is, Russell claims that even ordinary names are not proper names in the sense defined by thesis 4. Consider the sentence

(24) Odysseus was set ashore at Ithaca while sound asleep.

Frege would analyse this as having the logical form given by

(25) Mb

where "M" translates the predicate ". . . was set ashore at Ithaca while sound asleep" and "b" translates "Odysseus". Again, Russell suggests that we should not take the grammatical form of (24) as an infallible guide to its logical form. "Odysseus" is really a disguised definite description: for simplicity, suppose that it is the definite description "the hero of Homer's *Odyssey*" (24) thus really amounts to

(26) The hero of Homer's *Odyssey* was set ashore at Ithaca while sound asleep.

We now analyse this in the same way that we analysed (20), using "N" to translate the predicate ". . . is a hero of Homer's *Odyssey*":

(27) $(\exists x)((Nx \,\&\, Mx) \,\&\, (\forall y)(Ny \rightarrow x = y))$.

This analysis shows that "Odysseus" is not really a proper name. Rather, the semantic value of "Odysseus" is the second-level function denoted by:

(28) $(\exists x)((Nx \,\&\, \ldots x) \,\&\, (\forall y)(Ny \rightarrow x = y))$.

So the fact that there actually is no Odysseus does not cause problems: we can still account for the contribution "Odysseus" makes to the truth-values of sentences containing it, by giving a story along the lines of that we gave for functional expressions in Chapter 1. The problem of bearerless names has disappeared.

If neither definite descriptions nor ordinary names can be viewed as genuine proper names, what can? Isn't Russell committed to the view that there are *no* genuine proper names? In fact, Russell thinks that the only genuine proper names might turn out to be demonstrative expressions, such as "this" or "that":

> We may even go so far as to say that, in all such knowledge as can be expressed in words – with the exception of "this" and

> "that" and a few other words of which the meaning varies on different occasions – no [genuine proper] names occur, but what seem like [genuine proper] names are really descriptions.[62]

In fact, Russell's views here are even more drastic than they sound. Not only are "this" and "that" the only possible genuine proper names, but even they cannot be construed as proper names if they are taken as referring to physical objects. They can only be viewed as proper names if they are taken to refer to sense-data or "objects of sense":

> We say "This is white". If you agree that "this is white", meaning the "this" that you see, you are using "this" as a proper name. But if you try to apprehend the proposition that I am expressing when I say "This is white", you cannot do it. If you mean this piece of chalk as a physical object, then you are not using a proper name. It is only when you use "this" quite strictly, to stand for an actual object of sense, that it is really a proper name.[63]

It is easy to see the rough form of Russell's reasoning here. The only genuine proper names are those for which the problem of bearerless names cannot possibly arise. Sense-data seem like good candidates for objects whose existence cannot possibly be doubted: if it seems to me that there is a red sense-datum in my visual field, then there is. So the only genuine proper names are names of sense-data: demonstratives applied to currently existing experiences. Whether Russell's restricted conception of what count as genuine proper names is plausible is a question which we cannot enter into here, as it would take us far afield into Russell's epistemology. In §2.10 we consider further Russell's critique of Frege.

2.9 Scope distinctions

As noted above, on Russell's analysis, the sentence

(20) The King of France is bald

comes out *false*. But what about the sentence

(29) The King of France is not bald.

Is this true or false? Russell worries that on his analysis this sentence will *also* come out as false. This would be a problem, because it would seem to lead to a counterexample to the logical law known as the law of excluded middle. This law states that for any given sentence, either it or its negation must be true. But if both (20) and (29) are false, we seem to have a case where *neither* a sentence nor its negation are true. As Russell himself puts it:

> By the law of excluded middle, either "A is B" or "A is not B" must be true. Hence either "the present King of France is bald" or "the present King of France is not bald" must be true. Yet if we enumerated the things that are bald and then the things that are not bald, we should not find the present King of France in either list. Hegelians, who love a synthesis, will probably conclude that he wears a wig.[64]

Russell responds to this worry by distinguishing between two different ways of reading (29). On one reading, it is false, but is not really the negation of (20), so that no counterexample to the law of excluded middle is generated. On the other reading, it is the negation of (20), but is actually true, so that again there is no threat to the law of excluded middle. In order to distinguish between the two readings, Russell uses the notion of *scope*.

Consider

(30) All philosophers are not stupid.

Clearly, this is not consistent with the existence of a stupid philosopher: if a stupid philosopher exists, then (30) is false. Taking "F" as ". . . is a philosopher" and "G" as ". . . is stupid", (30) formalises as

(31) $(\forall x)(Fx \rightarrow - Gx)$.

Notice that the negation operator "–" occurs inside the part of (31) which is governed by the universal quantifier ($\forall$x). We say that the negation operator has *narrow* scope with respect to the universal quantifier (or equivalently, that the universal quantifier has *wide* scope with respect to the negation operator). Contrast this with

(32) It is not the case that all philosophers are stupid.

This is consistent with the existence of a stupid philosopher: so long as there is at least one other philosopher who is *not* stupid (32) still comes out true. Using the same letters as above, this formalises as

(33) $-(\forall x)(Fx \rightarrow Gx)$.

Notice that in this case the negation operator occurs outside the part of (33) which is governed by the universal quantifier. We say that the negation operator has *wide* scope with respect to the universal quantifier (or equivalently, that the universal quantifier has *narrow* scope with respect to the negation operator).

Now just as we can distinguish between the negation operator's having wide or narrow scope with respect to a quantifier, we can also distinguish between their having wide or narrow scopes with respect to a definite description. Consider the formalisation of (29), which clearly displays its falsity:

(34) $(\exists x)((Fx \,\&\, -Gx) \,\&\, (\forall y)(Fy \rightarrow x = y))$.

In this the negation operator has narrow scope with respect to the definite description (i.e. to the functional expression (23)). Russell points out that (34) is *not* the negation of (29): in order to get the negation of (29) we need to find something in which the negation operator has *widest* scope. The only candidate for this is

(35) $-(\exists x)((Fx \,\&\, Gx) \,\&\, (\forall y)(Fy \rightarrow x = y))$.

In this, the negation operator has wide scope with respect to the definite description. This shows that (35), and not (34) is the

negation of (29). Now (35) is actually *true*, since it is the result of negating $(\exists x)((Fx \ \& \ Gx) \ \& \ (\forall y)(Fy \rightarrow x = y))$ which on Russell's analysis is false. So, the negation of (22) is true, and there is no problem for the law of excluded middle.[65] As Russell puts it "we escape the conclusion that the King of France has a wig".[66]

2.10 Russell's attack on sense

In addition to making criticisms of detail about Frege's assignments of semantic value, Russell also criticised Frege's introduction of the notion of sense. There are two main criticisms of the notion of sense. The first is that the distinction between sense and semantic value is actually *incoherent*: Russell describes Frege's attempt at drawing the distinction as leading to an "inextricable tangle" and writes that "the whole distinction of [sense] and [semantic value] has been wrongly conceived".[67] Russell provides the argument that the notion of sense is incoherent on pp.48–51 of "On Denoting". But we shall not consider this argument here: it is truly one of the most mysterious passages in twentieth-century philosophy.[68] Instead we shall concentrate on Russell's second main line of criticism of the notion of sense. This does not involve the claim that the notion of sense is incoherent: rather, Russell tries to show that the invocation of the notion of sense is simply *superfluous*. Frege introduced the notion of sense in an attempt to solve a number of puzzles. Russell argues that these puzzles can be solved in a way which does not involve any appeal to the notion of sense, so that Frege's rationale for introducing the notion is simply undercut.

Recall that Frege's introduction of sense was motivated by the desire to solve three main problems: the problem of bearerless names, the problem of substitution into belief contexts, and the problem of informativeness. We have already seen how Russell attempted to solve the first of these: empty definite descriptions and ordinary names still have a semantic value because they are not actually genuine proper names, but rather have second-level functions as their semantic values. We now look at how Russell deals with the other two problems.

Recall that the problem of substitution into belief contexts was a problem because it threatened thesis 3: substituting a constituent of a sentence with another which has the same semantic value will leave the truth-value of the sentence unchanged. Consider

(36) Smith believes that the composer of *Fidelio* had cirrhosis of the liver.

Now suppose that Smith doesn't realise that the composer of *Fidelio* is in fact the same person as the composer of the Moonlight Sonata. That is, Smith doesn't know that

(37) The composer of *Fidelio* is the composer of the Moonlight Sonata.

Suppose that "The composer of *Fidelio*" and "The composer of the Moonlight Sonata" are proper names. Then, they have the same semantic value, since they pick out the same person, Beethoven. So we should be able to substitute "The composer of the Moonlight Sonata" for "The composer of *Fidelio*" in (36) without changing (36)'s truth-value. But in fact the substitution results in the false

(38) Smith believes that the composer of the Moonlight Sonata had cirrhosis of the liver.

So it looks as if we have a counterexample to thesis 3.

Russell wants to hold on to thesis 3, so he tries to explain away the apparent counterexample. In order to show how he attempted to do so, it will be best to first restate the problem in slightly more formal terms. On the assumption that definite descriptions are genuine proper names, we can (partially) translate (36) into logical symbolism as follows, where "a" translates "the composer of *Fidelio*" and "F" translates the predicate ". . . has cirrhosis of the liver":

(39) Smith believes that Fa.

Taking "b" to abbreviate "the composer of the Moonlight Sonata", (37) gets translated as:

(40) $a = b$.

And (38) gets translated as

(41) Smith believes that Fb.

The counterexample is now clear: we substitute into the true (39) on the basis of the true (40) and get the false (41) as a result.

Russell's response to the problem is basically this: (39), (40), and (41) misrepresent the logical form of (36), (37), and (38), and when we see their true logical form, we'll see that there is simply no room for the sort of substitution which takes us from (39) and (40) to (41). For Russell, "the composer of *Fidelio*" is not a proper name. Rather, it gets treated in the same way as "the King of France", so that the logical form of (36) is more accurately captured by

(42) Smith believes that $(\exists x)((Gx \,\&\, Fx) \,\&\, (\forall y)(Gy \rightarrow x = y))$

where "G" stands for the predicate ". . . is a composer of *Fidelio*". The logical form of (37) is given by

(43) $(\exists x)(Gx \,\&\, Hx \,\&\, (\forall y)(Gy \rightarrow x = y) \,\&\, (\forall z)(Hz \rightarrow x = z))$

where "H" abbreviates the predicate ". . . is a composer of the Moonlight Sonata".

Now whereas (40) allowed us to substitute into the true (39) to obtain the false (41), (43) simply does not allow us to substitute into (42): (42) and (43) are simply not of the right logical form to allow a substitution, so *a fortiori*, there is no possibility of a substitution which takes us from a true sentence to a falsehood. Alternatively, the logical form of (38) is given by

(44) Smith believes that $(\exists x)((Hx \,\&\, Fx) \,\&\, (\forall y)(Hy \rightarrow x = y))$.

This is false, but it cannot be reached by substituting into (42) on the basis of (43). Thesis 3 is safe. Note that since Russell views natural names as disguised definite descriptions, he can apply

this solution of the problem to cases which involve natural names rather than explicit definite descriptions. This allows him to solve the problem as presented in §2.1.[69]

Russell is thus able to solve the problem of substitution into belief contexts without invoking the notion of sense. But what about the problem of informativeness? Recall from §2.1 that Frege argued for the introduction of sense on the basis of a reductio argument: if the only semantic property that could be ascribed to expressions was possessing a semantic value, then it would not be possible to understand a sentence and yet fail to know its truth-value. Now, a crucial step in the reductio, as presented in §2.1, was premise (v): the semantic value of "The Morning Star" is the same as that of "The Evening Star", namely, the planet Venus. Russell will reply that this presupposes that "The Evening Star" and "The Morning Star" are proper names, expressions which have the objects they stand for as their semantic values. But this is false. Natural names are disguised definite descriptions, and have second-level functions, rather than the objects they pick out, as their semantic values. The fact that "The Evening Star" and "The Morning Star" both pick out the same object thus does not entail that they have the same semantic value, and Frege's argument is blocked: the assumption that the only semantic property of an expression is its possession of a semantic value does not imply that one can never understand a sentence without knowing its truth-value.

How plausible are Russell's solutions to these problems? The solution to the problem of substitution into belief contexts is supposed to save the following thesis:

> **Thesis 3:** Substituting a constituent of a sentence with another which has the same semantic value will leave the semantic value (i.e. truth-value) of the sentence unchanged.

In the case where the constituents are proper names, this amounts to: substituting one name in a sentence with another which picks out the same object should leave the truth-value of the sentence

unchanged. Now Russell is quite right to say that if his theory of descriptions is accepted, the examples we considered do not constitute counterexamples to this version of thesis 3. Since the expressions which are substituted are not proper names, this version of thesis 3 simply fails to have any application. But, of course, the version involving names is just one version of thesis 3. Since expressions other than names are also taken to have semantic values, there will be other versions of thesis 3. In particular, where the expressions concerned are definite descriptions, which on Russell's theory have second-level functions as their semantic values, thesis 3 will amount to: substituting one definite description in a sentence with another which has the same second-level function as its semantic value should leave the truth-value of the sentence unchanged. Does Russell's solution to the problem preserve this version of thesis 3? It does not, given the assumption that functions have extensional identity conditions (thesis 6). Consider the second-level function which is the semantic value of "the composer of *Fidelio*", denoted by:

(45) $(\exists x)((Gx \,\&\, \ldots x) \,\&\, (\forall y)(Gy \rightarrow x = y))$.

What is the extension of this function? This is a second-level function, a function from concepts (first-level functions) to truth-values, so we can represent its extension as a set of ordered pairs of concepts and truth-values. Consider a few predicates: "J" for ". . . was an uncle of Karl Beethoven", "K" for ". . . enjoyed drinking wine", "L" for ". . . went deaf", "M" for ". . . was English". Then we can represent the extension of the function (45) as:

(46) $\{(J, T), (K, T), (L, T), (M, F), \ldots\}$.

Consider the function which is the semantic value of "the composer of the Moonlight Sonata", denoted by:

(47) $(\exists x)((Hx \,\&\, \ldots x) \,\&\, (\forall y)(Hy \rightarrow x = y))$.

What is the extension of this second-level function? It can be represented as:

(48) {(J, T), (K, T), (L, T), (M, F), . . .}.

This is identical to (46), and the problem for Russell is now apparent. "The composer of *Fidelio*" and "The composer of the Moonlight Sonata" have the functions denoted by (45) and (47) as their respective semantic values. But these have the same extension, so given that functions have extensional identity conditions, "the composer of *Fidelio*" and "the composer of the Moonlight Sonata" have the same semantic value. Now go back to the sentence which led to the apparent counterexample to thesis 3:

(36) Smith believes that the composer of *Fidelio* had cirrhosis of the liver.

Since, *even given Russell's theory of definite descriptions*, "the composer of *Fidelio*" has the same semantic value as "the composer of the Moonlight Sonata", the following should be true given that (36) is:

(38) Smith believes that the composer of the Moonlight Sonata has cirrhosis of the liver.

But (38) is false. So the counterexample to thesis 3 still remains.

Note that these reflections also threaten Russell's solution to the problem of informativeness. Given that "The Morning Star" and "The Evening Star" are to be treated as disguised definite descriptions, a line of argument similar to that just given above will show that they have the same second-level functions as their semantic values. Russell's attempt to block Frege's reductio argument will thus fail.

Is there any way Russell can respond to these objections?[70] If no response can be made on behalf of Russell the conclusion will be that he fails in his attempts to solve the problem of substitution into belief contexts and the problem of informativeness without invoking the notion of sense, and that accordingly some notion of sense is after all required.[71]

2.11 Russell on communication

Recall from §2.3 that Frege invoked the notion of sense in an attempt to explain linguistic communication. Communication, or the "transmission of thought", is to be explained in virtue of the common senses which speakers attach to their words. We saw in §2.3 that Frege's attempt to use the notion of sense in an account of communication was not entirely successful, but leave this worry on one side for the moment. Does Russell fare any better? That is to say, can Russell give an account of communication in a manner which does not involve the invocation of a notion of sense?

Suppose that Smith says to Jones: "Bismarck was an astute diplomat". According to Frege, a condition on Smith's successfully communicating by means of this utterance is that Smith and Jones both attach the same sense to the name "Bismarck". Russell, however, wishes to reject the idea that natural names such as "Bismarck" even have a sense. So what account can he give of the conditions on successful linguistic communication? This question is all the more pressing for Russell given his view that natural names are really disguised definite descriptions and his explicit admission that different definite descriptions may be abbreviated by the same natural name for different people:

> Common words, even proper names, are usually really descriptions. That is to say, the thought in the mind of a person using a proper name correctly can generally only be expressed explicitly if we replace the proper name by a description. Moreover, the description required to express the thought will vary for different people, or for the same person at different times.[72]

So suppose that for Smith, the name "Bismarck" is an abbreviation of the description "the first Chancellor of Germany", and for Jones an abbreviation of the description "the most powerful man in Europe". Doesn't it follow from this that Smith's saying "Bismarck was an astute diplomat" to Jones will not result in a successful episode of communication? Russell suggests not: the fact that Smith and Jones abbreviate different descriptions by the

name "Bismarck" need not frustrate their attempt at communication. Mark Sainsbury suggests that the key sentence in making sense of Russell's suggestion here is the following:

> The only thing constant (so long as the name is rightly used) is the object to which the name applies.[73]

Sainsbury fleshes this out as follows:

> Suppose I utter the words "Bismarck was an astute diplomat". As I use the name "Bismarck", the thought in my mind may be best described as *that the first Chancellor of Germany* is an astute diplomat. However, I realize that you may associate the name with a different description, perhaps "the most powerful man in Europe". So when I utter my sentence, I am not trying to get you to share my thought; rather, I am trying to get you to have a thought, concerning Bismarck (however you think of him), that he is an astute diplomat. My intentions would be satisfied if you realized that I had tried to say, of the most powerful man in Europe, that he was an astute diplomat.[74]

Also

> It is not necessary that [Smith] and [Jones] should share a thought. All that is required is that the possibly various descriptions they associate with the name "Bismarck" should stand for the same thing; for this is what it means to say that the name "Bismarck" has a (public) reference in the language community to which [Smith] and [Jones] belong.[75]

How successful is Sainsbury's defence of a Russellian account of linguistic communication? Even if Sainsbury's defence of Russell's account of linguistic communication is successful, we saw that Russell runs into serious problems in his attempts to solve the problems of informativeness and substitution into belief contexts. What this suggests is that some notion of sense is required, even if not that invoked by Frege. We cannot pursue this matter further here. In the following two sections we look at important papers on reference by Strawson and Donnellan, followed by a brief account

of the causal-historical theory of reference that Kripke proposes as an alternative to those of Frege and Russell.

2.12 Strawson and Donnellan on referring and definite descriptions

In his classic 1950 paper "On Referring", P. F. Strawson outlines a critique of Russell's theory of descriptions. In this section, we give a brief outline of Strawson's main criticism of Russell followed by a brief account of some criticisms of both Russell and Strawson voiced by Keith Donnellan in his famous 1966 paper "Reference and Definite Descriptions".

Strawson notes:

> Russell's Theory of Descriptions . . . is still widely accepted among logicians as giving a correct account of the use of such expressions [e.g. "the King of France is wise"] in ordinary language. I want show . . . that this theory, so regarded, embodies some fundamental mistakes.[76]

His main objection turns on a distinction he draws between *sentence-types*, on the one hand, and *statements which tokens of that sentence-type are used to make* (by particular speakers in particular contexts of utterance).[77] Strawson argues that the former can be regarded as significant or insignificant, while only the latter can be regarded as true or false. Likewise for subsentential expressions: expression-types can be regarded as significant or insignificant, whereas only particular tokens of expression-types can be regarded as referring or failing to refer. He writes:

> "Mentioning", or "referring", is not something an expression does; it is something a speaker does; it is something that someone can use an expression to do. Mentioning, or referring to, something is a characteristic of *a use* of an expression, just as "being about" something, and truth-or-falsity, are characteristics of *a use* of a sentence.[78]

Consider the sentence

(49) The King of France is wise.

According to Russell's analysis, this sentence is false, whereas for Strawson, the sentence itself is not something that can legitimately be described as true or false: "we cannot talk of the sentence being true or false, but only of its being used to make a true or false assertion, or (if this is preferred) to express a true or false proposition".[79] So what about a particular utterance of a token of (39)? According to Russell, such an utterance would be false. Strawson disagrees:

> [S]uppose someone were in fact to say to you with a perfectly serious air: "The king of France is wise". Would you say," That's untrue"? I think it's quite certain that you wouldn't. But suppose he went on to ask you whether you thought that what he had just said was true, or was false; whether you agreed or disagreed with what he had just said. I think you would be inclined, with some hesitation, to say that you didn't do either; that the question of whether his statement was true or false simply didn't arise, because there was no such person as the king of France.[80]

In order to explain why this is so, Strawson uses the distinction drawn above between sentence-types and statements made by tokens of that type:

> The sentence [(49)] is certainly significant; but this does not mean that any particular use of it is true or false. We use it truly or falsely when we use it to talk about someone; when, in using the expression, "The king of France", we are in fact mentioning someone. The fact that the sentence and the expression, respectively, are significant just is the fact that the sentence *could* be used, in certain circumstances, to say something true or false, that the expression *could* be used, in certain circumstances to mention a particular person.[81]

Although an utterance of (49) could be used to make an assertion which would be true or false, in circumstances in which France is a republic and not a monarchy a presupposition of its making such a statement – that there is such a person as the King of France – is not satisfied, so that the utterance of (49) fails to make an assertion, in other words fails to express a statement with a truth-value:

> [T]his comes out from the fact that when, in response to his statement, we say (as we should) "There is no king of France", we should certainly *not* say that we were *contradicting* the statement that the king of France is wise. We are certainly not saying that it's false. We are, rather, giving a reason for saying that the question of whether it's true or false simply doesn't arise.[82]

Thus, according to Russell, an utterance of (49) by me now would be false, whereas according to Strawson, an utterance of (49) by me now would not succeed in expressing a statement that was either true or false. Who is right? We'll return to this question at the end of §7.6.

In his famous 1966 paper "Reference and Definite Descriptions" Keith Donnellan takes both Russell and Strawson to task for neglecting a distinction between two distinct uses of definite descriptions.

Contrast two contexts in which the sentence "The murderer of Smith is insane" may be uttered. In the first, it is uttered by a detective, who, inspecting Smith's mutilated corpse, comes to the conclusion that whoever did this to Smith must be insane. Suppose it turns out that in fact Smith committed suicide. Then, there is no-one who satisfies the description "the murderer of Smith", and, as with (49) above, Russell and Strawson disagree about the impact this has on truth-value of the utterance. According to Russell the utterance is false, according to Strawson, the utterance fails to say anything either true or false. Donnellan calls this kind of use of a definite description an *attributive* use.

The main problem with Russell's account, according to Donnellan, is that it entirely neglects a radically different kind of use to which definite descriptions may be put. Consider again an

utterance of "Smith's murderer is insane", this time uttered in court by a journalist watching the odd behaviour of Green, who is in the dock as the accused undergoing cross-examination for the alleged murder of Smith. According to Donnellan, in this sort of case the journalist succeeds in referring to Green even if in fact Green is not Smith's murderer. This sort of case, in which a definite description can be used to say something true or false even if in fact nothing satisfies it, Donnellan calls a *referential* use. Donnellan's main criticism of Russell is that his account of the semantics of definite descriptions fails to deal with this sort of case.

Among the criticisms Donnellan levels against Strawson is the following. According to Strawson, the utterance of "The murderer of Smith is insane" made by the detective on finding Smith's body is neither true nor false. Donnellan concedes that this may well be the case, but his problem concerns the *explanation* which Strawson suggests: according to Strawson "The murderer of Smith is insane" fails to have a truth-value since, because no-one actually did murder Smith, the phrase "the murderer of Smith" fails to refer. That this is the kind of explanation that Strawson offers is apparent from his remarks on (49):

> When we utter the sentence without in fact mentioning anybody by the use of the phrase, "The king of France", the sentence doesn't cease to be significant; we simply fail to say anything true or false because we simply fail to mention anybody by this particular use of that perfectly significant phrase. It is, if you like, a spurious use of the expression; though we may (or may not) mistakenly think it is a genuine use.[83]

Donnellan objects that this presupposes that "the G" in "the G is H" fails to refer if there is in fact no G: a presupposition that is shown to be false by the possibility of referential uses of definite descriptions. So Strawson's explanation fails.[84]

2.13 Kripke's causal-historical theory of reference

In §2.5 we saw that Kripke rejects Frege's account of names, and he takes his criticisms to apply also to Russell's view that names are disguised definite descriptions. How, then, is the reference of

a proper name, in general, fixed? Kripke proposes to replace the Fregean and Russellian view with the following alternative picture:

> Someone, let's say, a baby, is born; his parents call him by a certain name. They talk about him to their friends. Other people meet him. Through various sorts of talk the name is spread from link to link as if by a chain. A speaker who is on the far end of this chain, who has heard about, say, Richard Feynman, in the market place or elsewhere, may be referring to Richard Feynman even though he can't remember from whom he first heard of Feynman or from whom he ever heard of Feynman. He knows that Feynman is a famous physicist. A certain passage of communication reaching ultimately to the man himself does reach the speaker. He then is referring to Feynman even though he can't identify him uniquely . . . a chain of communication going back to Feynman himself has been established, by virtue of his membership in a community which passed the name on from link to link.[85]

Note that the cases where it is plausible to view definite descriptions as playing a reference-fixing role are those in which the object is originally christened with the name. In some cases this will be via ostention ("I call this baby 'Rosa'"), but in others it may be by description (the Chief Inspector announces that "'Jack the Ripper' refers to whoever committed the last three Whitechapel murders"). This only applies to the first link in the chain: after the initial link, causal considerations take prominence.

Kripke admits that this picture is a rough one, and he expresses doubt as to whether it can be refined into a set of necessary and sufficient conditions for reference. Indeed, he explicitly disavows the ambition to construct an alternative "theory" of reference at all. But the picture is already sharp enough to admit of potential counterexamples. The name "Edward Elgar" was given, we may suppose, to the distinguished English composer by his parents in some initial "dubbing" or "baptismal" ceremony in Worcestershire in the nineteenth century. The name has finally reached me via a long causal chain, reaching back through time to this initial baptismal act. Does it follow that when impressed by my dog's noble countenance I start to call him "Edward Elgar" I refer to the distinguished composer and not to my Springer Spaniel? Of course

not: and Kripke adds a simple clause to his account designed to deal with this sort of example:

> An initial "baptism" takes place. Here the object may be named by ostension, or the reference of the name may be fixed by a description. When the name is "passed from link to link" the receiver of the name must, I think, intend when he learns it to use it with the same reference as the man from whom he heard it.[86]

When, in throwing the ball over the dog's head, I shout "Go fetch it Elgar", my intention is not to use it with the same reference as André Previn, from whom I originally picked up the name. So there is no real counterexample.

Even with the addition of the clause that the receiver intends to use the name with the same reference as the person from whom he learned it, though, other counterexamples can be found. Gareth Evans cites the following example from Isaac Taylor's 1898 book *Names and their History*:

> In the case of "Madagascar" a hearsay report of Malay or Arab sailors misunderstood by Marco Polo . . . has had the effect of transferring a corrupt form of the name of a portion of the African mainland to the great African island.[87]

Suppose that Marco Polo picked up the name "Madagascar" from Solly, a native. And suppose that the use we moderns make of "Madagascar" can be traced back to Marco Polo. Then, since we intend to use the name with the same reference as Marco Polo, and since Marco Polo presumably intended to use it with the same reference as Solly, it follows that in our contemporary use of "Madagascar" we refer not to the island, but to the relevant portion of the African mainland. This example clearly calls for some more sophisticated revisions in Kripke's causal picture. However, we cannot pursue the causal-historical theory further here. In the next chapter we move on to look at a school of philosophers whose imposition of constraints on what can and cannot possess sense provided the central plank of their philosophical outlook, and who provide an essential backdrop for understanding the philosophy of language of the second half of the twentieth century: the logical positivists.[88]

2.14 Appendix: Frege's theses on sense and semantic value

Thesis 1: The semantic value of a sentence is its truth-value (true or false).

Thesis 2: The semantic value of a complex expression is determined by the semantic values of its parts (*compositionality of semantic value*).

Thesis 3: Substitution of a constituent of a sentence with another which has the same semantic value will leave the semantic value (i.e. truth-value) of the sentence unchanged (*Leibniz's Law*).

Thesis 4: The semantic value of a proper name is the object which it refers to or stands for.

Thesis 5: The semantic value of a predicate is a function.

Thesis 6: Functions are extensional: if function f and function g have the same extension, then f = g.

Thesis 7: The semantic value of a predicate is a first-level function from objects to truth-values; the semantic value of a sentential connective is a first-level function from truth-values to truth-values; the semantic value of a quantifier is a second-level function from concepts (first-level functions) to truth-values.

Thesis 8: The sense of an expression is that ingredient of its meaning which determines its semantic value.

Thesis 8a: The sense of an expression is the mode of presentation of its semantic value.

Thesis 9: It is possible to know the sense of an expression without knowing its semantic value.

Thesis 10: The sense of an expression is what someone who understands the expression grasps.

Thesis 11: The sense of a complex expression is determined by the senses of its constituents (*compositionality of sense*).

Thesis 12: If someone grasps the senses of two expressions, and the two expressions actually have the same sense, then she must know that the two expressions have the same sense (*transparency of sense*).

Thesis 13: An expression can have a sense even if it lacks a semantic value.

Thesis 14: A sentence which contains an expression which lacks a semantic value is neither true nor false.

Thesis 15: In a belief context, the (indirect) reference of a proper name is its customary sense.

Thesis 16: Substitution of one expression in a sentence with another which has the same sense will leave the sense of the sentence unchanged.

Thesis 17: The sense of a sentence is a thought (its truth-condition).

Thesis 18: The semantic value of an expression is no part of what someone who understands the expression grasps.

Thesis 19: Sense is objective: grasping a sense is not a matter of having ideas, mental images, or private psychological items.

Thesis 20: The sense of an expression is normative: it constitutes a normative constraint which determines which uses of that expression are correct and which are incorrect.

Thesis 21: In addition to sense and semantic value, we must also introduce the notions of *force* and *tone*.

Chapter summary

In this chapter we continued our exposition and assessment of Frege's systematic account of meaning, mainly as set out in his famous paper "On *Sinn* and *Bedeutung*". We looked at Frege's reasons for introducing a notion of *sense* in addition to the notion of *semantic value*. According to Frege, we need the notion of sense in order to deal with three problems, concerning bearerless names, informative identity statements, and substitution of co-referring expressions in belief contexts. We then outlined Frege's views on the nature of sense, including his view that sense is objective and normative. We then looked at a number of challenges to Frege's theory, including those developed by Kripke in his seminal book *Naming and Necessity*. Following a brief exposition of Frege's notions of *force* and *tone*, we outlined the critique of Frege's theory contained in Russell's theory of definite descriptions: according to Russell, Frege is wrong to view definite descriptions such as "the King of France" as having objects as semantic values and instead suggests that they have second-level functions as their semantic values. He also argues that we don't need the notion of sense to deal with the problems about bearerless names, etc. We outlined some potential problems for Russell's view before proceeding to look at two classic commentaries, Strawson's "On Referring" and Donnellan's "Reference and Definite Descriptions", and Sainsbury's defence of Russell's account of communication. We finished with a brief look at Kripke's alternative to the Fregean and Russellian views, the causal-historical theory of reference.

Study questions

1. Construct an example involving functional expressions to illustrate the problem of substitution into belief contexts outlined in §2.1.2.
2. How plausible is the hidden premise in Frege's argument concerning the problem of informativeness (§2.1.3)? Can you think of a counterexample?

3. How does Frege's argument concerning the problem of informativeness generalise to the cases of sentences and predicates?
4. How might a defender of Frege's argument against Locke widen the initial inversion hypothesis (§2.3) to take care of "Yellow is more like white than is blue"? Are there any cases that the inversion hypothesis cannot be extended to deal with?
5. Does Frege's own view of the senses of names leave him open to a version of his own objection to Locke's imagist theory of meaning?
6. Explain the difference in meaning between

   (i) Soames is clever and he voted for Trump

   and

   (ii) Soames is clever but he voted for Trump.

   Give some examples of your own of expressions which have the same sense but which differ in tone. Is it possible to give a systematic theory of tone? Justify your answer.
7. How would Russell analyse "the golden mountain is golden"? Does this analysis give an accurate account of its truth-value?
8. Is there any way Russell can respond to the objections raised in §2.10 against his proposed solutions to the problem of substitution into belief contexts and the problem of informativeness?
9. How plausible is Sainsbury's defence of Russell's account of linguistic communication?
10. What is the truth-value of the sentence "the King of France is bald"? What would Strawson say? Is he right?
11. Make up an example of your own to illustrate Donnellan's distinction between attributive and referential uses of definite descriptions.

Further reading

The student should begin by looking at the following four essays by Frege: "Function and Concept", "On Concept and Object", "On *Sinn* and *Bedeutung*", and "Thought". All of these are reprinted in Beaney (ed.), *The Frege Reader*. The first three of these can

also be found in Geach and Black (eds), *Translations from the Philosophical Works of Gottlob Frege*. The last can also be found in Strawson (ed.), *Philosophical Logic*. Frege's "Letter to Jourdain" (reprinted in Beaney (ed.), *The Frege Reader*) is also important. After that, it would be advisable to look at some of the recent introductory textbooks, for example, Kenny, *Frege*; Noonan, *Frege*; or Weiner, *Frege*. See also Currie, *Frege: An Introduction to His Philosophy* and Textor, *The Routledge Philosophy Guidebook to Frege on Sense and Reference*. After those, it might be useful to look at McCulloch, *The Game of the Name* (especially Chapters 1 and 5). McCulloch's *The Mind and Its World* also contains much useful material on Frege, and is especially good on the relationship between themes in Frege's philosophy of language and current issues in philosophy of mind. Chapter 3 is devoted to Frege, but this book is probably best read in its entirety. The doyen of Frege scholarship is Michael Dummett. His books *Frege: Philosophy of Language* and *The Interpretation of Frege's Philosophy* are classics, and set the scene for all subsequent discussion of Frege. But they are very long and, in places, very demanding. A crash course in Dummett's Frege might consist of the Introduction, and Chapters 1, 5, 6, and 19 of *Frege: Philosophy of Language,* though Dummett's books require and deserve close and careful study. For a brief and clear discussion of Frege by Dummett, see Chapters 8–10 of his *The Nature and Future of Philosophy*. An alternative, "Russellian", interpretation of Frege's views on sense – one which interprets Frege as denying thesis 13 – can be found in Evans, *The Varieties of Reference*. See also McDowell, "On the Sense and Reference of a Proper Name" and the two books by McCulloch mentioned above. For a powerful critique of Evans's interpretation, see Bell, "How 'Russellian' Was Frege?". A spirited defence of a Lockean view of sense can be found in Chapter 7 of Lowe, *Locke on Human Understanding*. For a critical reply, see Miller, "Review Article on Lowe". Lowe replies in Chapter 4 of his later (2005) book *Locke*. A useful discussion of Kripke's *Naming and Necessity* can be found in Chapter 4 of McCulloch's *The Game of the Name*. See also the essays by Stalnaker and Stanley in Hale, Wright, and Miller (eds.), *A Companion to the Philosophy of Language*, and Devitt, *Designation*. Dummett defends Frege against Kripke's objections

in the appendix to Chapter 5 of *Frege: Philosophy of Language*. For a book-length survey of Kripke on these issues, see Noonan, *Routledge Guidebook to Kripke on Naming and Necessity*. For an introduction to philosophical issues surrounding the notions of necessity and contingency, see Divers, *Possible Worlds* and Melia, *Modality*.

A useful introduction to Russell's philosophy as a whole is Grayling, *Russell: A Very Short Introduction*. For Russell's theory of descriptions, the relevant papers are "On Denoting" and "Descriptions and Incomplete Symbols" (Chapter 6 of "The Philosophy of Logical Atomism"). Both of these are reprinted in Marsh (ed.), *Logic and Knowledge*. Other relevant works are "Descriptions" (Chapter 16 of Russell's *Introduction to Mathematical Philosophy*), and "Knowledge by Acquaintance and Knowledge by Description" (Chapter 5 of Russell's *The Problems of Philosophy*). There is a good discussion of Russell's theory, and a very useful account of how it relates to his more general epistemological doctrines, in Chapters 2 and 3 of McCulloch's *The Game of the Name*. Sainsbury's account of Russell on communication can be found in the section on Russell in his "Philosophical Logic", in Grayling (ed.), *Philosophy: A Guide Through the Subject*. See also Sainsbury's entry on "Frege and Russell" in *The Blackwell Companion to Philosophy*, and Chapter 9, "Reference", of Blackburn, *Spreading the Word*. Other works worth consulting are: Makin, *The Metaphysicians of Meaning*; Lycan, "Names"; and the various essays in Part 4, "The Nature of Reference", of LePore and Smith, *The Oxford Handbook of Philosophy of Language*.

For very interesting background to the issues dealt with in this chapter, see Ray Monk's excellent two-volume biography of Russell, *Bertrand Russell: The Spirit of Solitude* and *Bertrand Russell: The Ghost of Madness*.

Notes

1 "Sense" is how we translate Frege's German word *Sinn*.
2 Without giving up these theses, that is, as applied to cases where belief contexts and the like are not involved. To anticipate, Frege introduces the notion of sense, and then formulates thesis 15 to deal with the special case of sentences involving belief contexts.

3 Contexts such as "believes that . . ." which appear to threaten thesis 3 are known as intensional contexts. Other examples of intensional contexts are "wishes that . . ." and "It is necessarily the case that . . .".
4 Note that there is a hidden premise here: if S knows the reference of "a" and the reference of "b", and if the reference of "a" is in fact identical to the reference of "b", then S knows that the reference of "a" is identical to the reference of "b". Is this plausible?
5 Frege, "On *Sinn* and *Bedeutung*", p.153.
6 There are some problems here, which we cannot pause to discuss. For example, all logically true sentences have the same truth-condition, so it would follow that they have the same sense.
7 Dummett, *Frege*, p.91.
8 Ibid., p.93.
9 Ibid.
10 Frege, *Foundations of Arithmetic*, p.x.
11 Frege, "On *Sinn* and *Bedeutung*", pp.154–155.
12 Locke, *An Essay Concerning Human Understanding*, III, ii, 2.
13 Some Locke scholars have denied that ideas are mental images for Locke, but the case against them seems to me to be pretty conclusive. See McCulloch *The Mind and Its World*, Ch. 2. For the opposing point of view, see J. Yolton, *Perceptual Acquaintance from Descartes to Reid*.
14 Locke, *An Essay Concerning Human Understanding*, II, x, 5.
15 Ibid., III, ii, 1.
16 The argument is cryptically stated in §26 of Frege, *The Foundations of Arithmetic*. It is developed by Greg McCulloch in Chapter 3 of *The Mind and Its World*, and I follow McCulloch's excellent exposition in what follows. I should point out, however, that I try to squeeze a stronger conclusion out of the argument than McCulloch. McCulloch uses the argument to discredit Locke's theory of communication, but he thinks that it leaves open the possibility of a view which "takes over the bulk of Frege's theories of sense and reference, [which uses the theory of reference to explain communication], but which uses Lockean ideas to do the jobs earmarked for sense" (pp.74–75). McCulloch goes on to develop arguments along Wittgensteinian lines against this sort of position. The argument I now give in the text suggests that the original Fregean argument actually provides us with the resources to mount an attack on Locke's view, construed as an account of sense, and so can be used to supplement the Wittgensteinian arguments which McCulloch goes on to develop.
17 Frege, "On *Sinn* and *Bedeutung*", p.154.

18 Locke, *An Essay Concerning Human Understanding*, II, xxxii, 15.
19 Note that we would have to widen the initial inversion hypothesis to take care of things like "Yellow is more like white than is blue", and so on. But there seems to be no good reason why such a widening cannot in principle be carried out.
20 Locke, *An Essay Concerning Human Understanding*, II, xxx, 15.
21 For further attacks on the Lockean account of sense, see §5.1.
22 The issues about objectivity and normativity are not independent. Both are essential facets of Frege's opposition to *psychologism*, the doctrine that logic is a branch of empirical psychology. Viewing sense as a matter of the possession of mental images is obviously a slide in the direction of psychologism, since sense determines semantic value, which in turn determines whether or not logical inferences are valid. Frege's point is that this sort of view of sense, in addition to failing to accommodate its objectivity, would also fail to accommodate its essential normativity: the laws of logic would become laws of a descriptive science, psychology, rather than normative laws of thought.
23 Frege, "On *Sinn* and *Bedeutung*", p.155.
24 Ibid., p.153.
25 Ibid.
26 Of course, I'm oversimplifying here. For a more nuanced discussion, see Beaney, "What is Analytic Philosophy?".
27 For example, Russell's theory of definite descriptions (see §2.8), a paradigm example of analytic philosophy, would not have been possible if Frege's symbolic logic had not been invented.
28 This is sometimes known as the "Paradox of Analysis".
29 It is especially ironic that this should be an apparent consequence of views held by someone dubbed "the founder of analytic philosophy"!
30 The issues concerning sense and analysis are obviously considerably more complicated than I imply here. For a full discussion see Beaney, *Frege*. For an indication of one way Frege might attempt to deal with this worry, see Carruthers, *Tractarian Semantics*, p.16–17.
31 Russell, "On Denoting", p.50.
32 Frege, *Foundations of Arithmetic*, p.x.
33 Dummett, *Frege*, p.268.
34 Evans, *Varieties of Reference*, p.24.
35 Ibid., p.23.
36 Note that Frege may try to get round this problem by assigning bearerless names an arbitrary object – the number zero, say – as their semantic value. But this is highly problematic. As McCulloch puts it

"from the point of view of the meaning-theorist's enterprise, there is an evident problem in reaching for the logician's technical solution of supplying some arbitrary object as semantic value. For a semantic representation which has it that 'The present King of France' is not about a man, or someone who rules France, but instead concerns some arbitrary object like zero, has surely lost all contact with linguistic reality" (McCulloch, *Game of the Name*, p.46).

37 For Evans's interpretation, see especially *Varieties of Reference*, §1.6. The denial of thesis 13 is "Russellian", since for Russell e.g. genuine proper names are meaningful only if they have referents. See p.65 below.

38 Frege, "On *Sinn* and *Bedeutung*", p.152.

39 See Bell, "How 'Russellian' Was Frege?", §5.

40 Kripke talks of the "Frege–Russell" view of names. As we'll see, this masks important differences between Frege and Russell. But the terminology is understandable, given that Russell takes ordinary proper names to be abbreviations of definite descriptions and Frege uses definite descriptions as examples of the senses of names.

41 Of course, in asking this we mean: could things have turned out in such a way that this sentence – *given the meaning that it has* – would have been false? If the sentence had meant something other than it actually means, say, *that Clinton won the 2017 election*, it might have been false given *that* meaning, but this is irrelevant to the point made in the text.

42 This contrasts with Frege himself, who held that arithmetic is analytic.

43 Some philosophers have tried to avoid this problem by widening the sense-giving description, so that it becomes a large disjunction of all of the properties commonly attributed to the putative bearer. See Searle, "Proper Names". Kripke responds that this suggestion falls to essentially the same objection: it will still be at best a contingent truth that the bearer has any of the properties mentioned in the disjunction. See Kripke, *Naming and Necessity*, p.62. For Searle's reply, see *Intentionality*, pp.242–261. As an alternative to Frege's theory, Kripke proposes a "causal theory of reference". See §2.13 below.

44 Kripke, *Naming and Necessity*, p.34.

45 Ibid., pp.83–84.

46 Ibid., p.84.

47 Ibid., p.81.

48 Kripke explores the implications of this result for views in the philosophy of mind, in particular the materialist "identity theory", in Lecture 3 of *Naming and Necessity*.

49 Dummett, *Frege*, pp.98–99; see also Evans, *Varieties of Reference*, p.18.
50 It should be noted that, in addition to the problems outlined, Frege faces further problems stemming from the fact that he sees senses as abstract objects: they are neither physical nor mental, and subsist in what he calls "the third realm". Given that senses are abstract, there is no possibility of our coming into causal contact with them, so how can we ever know anything about them?
51 Evans, *Varieties of Reference*, pp.25–26.
52 Dummett, *Frege*, p.227. See also Evans, *Varieties of Reference*, pp.25–27.
53 Dummett, *Frege*, p.84.
54 So, the following all give correct accounts of the truth-conditions of the sentences quoted:

"Fido is a dog" is true iff Fido is a dog
"Fido is a dog" is true iff Fido is a cur
"Fido is a cur" is true iff Fido is a dog
"Fido is a cur" is true iff Fido is a cur.

55 This shows how (17), (18), and 19) are connected: it tells us why e.g. (17) might be the answer to (18), or a report that (19) has been obeyed.
56 See e.g. Dummett, *Frege*, p.95, where he explicitly speaks of force as an ingredient in meaning, alongside sense and tone. For criticism of Dummett's conception of the role of force, see McGinn, "Semantics for Non-Indicative Sentences".
57 Note, though, that when I utter (17) in the example just outlined, I also perform the speech-act of asserting that Jones makes an excellent cup of tea, because of the sentence-meaning of (17). This shows that it is possible to perform more than one speech-act (and even more than one type of speech-act) by means of a single utterance.
58 Note that the distinction between semantics and pragmatics may be drawn in more than one way. See Travis, "Pragmatics"; Longworth, "Semantics and Pragmatics".
59 Note the direction of explanation here: an account of sentence-meaning is given, and this is then used as input in an explanation of speaker's-meaning. Many philosophers think that the order of explanation should actually be the reverse of this: the notion of speaker's-meaning should be elucidated first, and the notion of sentence-meaning then elucidated in terms of it. This issue will be discussed in Chapter 7.

60 Note that it is crucially important that we take the predicate to be ". . . is *a* king of France". If we took it as ". . . is *the* King of France" we would be using the very notion we are attempting to analyse.
61 Note that Russell is therefore misdescribing the conclusion of his theory of descriptions when he claims that it entails that a description does not possess "any significance on its own account". On Russell's theory, descriptions do possess significance on their own account, by having second-level functions, rather than objects, as their semantic values. See McCulloch, *Game of the Name*, p.65, fn. 8. This will be important in what follows.
62 Russell, *Introduction to Mathematical Philosophy*, p.178.
63 Russell, "Philosophy of Logical Atomism", p.201.
64 Russell, "On Denoting", p.48.
65 Note that Russell uses slightly different terminology. The case where the description has wide scope with respect to the negation operator is described by Russell as one in which the description has a "primary occurrence"; the case in which the description has narrow scope with respect to the negation operator is one in which it has a "secondary occurrence".
66 Russell, "On Denoting", p.53.
67 Ibid., p.50. Note that Russell uses "meaning" for "sense" and "denotation" for "semantic value".
68 For discussion of the mysterious passage, see the appendices to Makin, *Metaphysicians of Meaning*, and the works referred to therein.
69 Note that scope issues, of the sort discussed in §2.8, also arise when definite descriptions appear in sentences which contain propositional attitude operators, such as "believes that . . .". For example, the following would be a case in which the definite description has wide scope with respect to "believes that . . .":

> (*) One and only one man wrote *Fidelio* and Smith believes that that man had cirrhosis of the liver.

Clearly, in this sort of case, the substitution of the definite description by another which in fact picks out the same individual will not result in a change in truth-value. If (*) is true, so is

> (**) One and only one man wrote the Moonlight Sonata and Smith believes that that man had cirrhosis of the liver.

So substitutions of descriptions which pick out the same individual are fine, so long as the descriptions have wide scope (or primary

occurrences) in the relevant sentences. Russell describes this by saying that his theory does not interfere with "verbal substitutions" of one description for another. Matters are different when a description has narrow scope, as in

> (***) Smith believes that one and only one man wrote *Fidelio* and that that man had cirrhosis of the liver.

See the discussion in Russell, "On Denoting", p.52.

70 Russell could say that functions have *intensional* identity conditions, but this would look like introducing a notion of sense through the back door. See Dummett, *Frege*, pp.208–209.

71 One way forward here might be to follow Frege in retaining a notion of sense – thereby allowing us to deal with the problems of substitution into belief contexts and informativeness – but to neutralise the problem of bearerless names by rejecting Frege's thesis 13 (which says that an expression can have a sense even if it lacks a semantic value). See Evans, *Varieties of Reference* for a similar strategy.

72 Russell, *Problems of Philosophy*, pp.29–30.

73 Ibid., p.30.

74 Sainsbury, "Philosophical Logic", p.71.

75 Sainsbury, "Frege and Russell", p.670.

76 Strawson, "On Referring", p.321.

77 Strawson doesn't put it in quite these terms, but I think this accurately captures the distinction he relies on.

78 Strawson, "On Referring", p.326

79 Ibid.

80 Ibid., p.330.

81 Ibid.

82 Ibid.

83 Ibid., p.331.

84 Evaluation of Donnellan's objections to Russell and Strawson is outwith the scope of the present chapter, but we should note that a Russellian may attempt to respond by drawing on the distinction between semantics and pragmatics (see §2.6 above). Just as a sentence with the strict and literal meaning that Jones makes an excellent cup of tea ("Jones makes an excellent cup of tea") can be used in a particular context by a particular speaker with particular communicative intentions to say that Jones is an uninteresting philosopher, a sentence with the strict and literal meaning that one and only one person killed Smith and that person is insane

("The murderer of Smith is insane" on a Russellian account of its semantics) can be used in a particular context by a particular speaker to say that Green is insane. For discussion and a guide to further reading on this and other issues concerning descriptions, see Ludlow and Neale, "Descriptions".

85 Kripke, *Naming and Necessity*, p.91.

86 Ibid., p.96.

87 Evans, "Causal Theory of Names", p.11. See the Appendix to *Naming and Necessity* for Kripke's comments on Evans.

88 Note that in these first two chapters our strategy has been as follows. We take a stretch of natural language, and our aim is to apply Frege's theory of semantic value to it. But in order to do this, we have to be able to translate it into the language of Frege's logical symbolism. But even though simple arguments can be translated into the formal language, what assurance do we have that all of the parts of language that we would intuitively take to be significant can be so formulated? It should be pointed out that Frege thought this an impossible task in the case of natural language: his main interest was in the language of science and mathematics. But other philosophers have not shared Frege's pessimism over the prospects of applying his ideas to natural language; this is a point we'll return to in Chapter 8.

Chapter 3

Sense and verificationism

Logical positivism

3.1 From the *Tractatus* to the verification principle

In Chapters 1 and 2 we looked at some aspects of Frege's attempt to systematise our intuitive notion of meaning, and how that attempt was modified in various ways by Russell. In this chapter we will look at another attempt: that carried out by the *logical positivists*. Logical positivism was a school of philosophy, centred in Vienna, which grew up in the 1920s and 30s, and which was informally institutionalised in the "Vienna Circle". The leading figure in the Circle was Moritz Schlick (1882–1936), and it counted amongst its supporters philosophers such as Neurath, Waismann, Feigl, Gödel, Ayer, Carnap, and Hahn. The main philosophical influences on the Circle stemmed from the Scottish empiricist David Hume (1711–1776), the Irish empiricist George Berkeley (1685-1753), and, less distantly, Frege and Russell. The influence of Frege and Russell on the logical positivists was largely transmitted through the *Tractatus Logicus-Philosophicus*, the famous early work of Ludwig Wittgenstein (1889–1951) that was first published in 1921.

In the preface to the *Tractatus*, Wittgenstein had written, "I am indebted to Frege's great works and to the writings of my friend Mr Bertrand Russell for much of the stimulation of my thoughts".[1] In the *Tractatus* Wittgenstein developed a sophisticated theory of the relationship between language and the world that synthesised and modified the views of Frege and Russell outlined in the previous two chapters. For example, like Frege, Wittgenstein held that the meaning of a sentence is its truth-condition:

> 4.024 To understand a proposition means to know what is the case if it is true.[2]

However, whereas Frege held that sentences and names have both sense and semantic value, Wittgenstein held that sentences have sense but not semantic value and that names have semantic value but not sense:

> 3.3 Only propositions have sense; only in the nexus of a proposition does a name have meaning.
>
> 3.203 A name means an object. The object is its meaning.
>
> 3.144 Situations can be described but not *given names*. (Names are like points; propositions are like arrows – they have sense.)

Whereas Frege held that the logical connectives had various truth-functions as semantic values, Wittgenstein denied this:

> 4.0312 My fundamental idea is that the "logical constants" are not representatives.
>
> 4.441 There are no "logical objects".

Wittgenstein admired Russell's theory of descriptions and believed that Russell's distinction between grammatical form and logical form constituted a tremendous advance in philosophy:

> 4.002 Language disguises thought. So much so, that from the outward form of the clothing it is impossible to infer the form of the thought beneath it, because the outward form of the clothing is not designed to reveal the form of the body, but for entirely different purposes.

> 4.0031 It was Russell who performed the service of showing that the apparent logical form of a proposition need not be its real one.

The doctrine of the *Tractatus* that most influenced the logical positivists concerned the nature and extent of substantive propositions or meaningful statements. According to Wittgenstein, all substantive propositions ("pictures") are at most *a posteriori* true:

> 2.224 It is impossible to tell from a picture alone whether it is true or false.

All substantive propositions are truth-functional compounds of elementary propositions, where elementary propositions are logically independent of each other and consist of names standing for simple objects. Further, an elementary proposition is true if and only if it corresponds to a state of affairs. It corresponds to a state of affairs if and only if the objects denoted by the names in the proposition are combined, in the state of affairs, in the same way that the names are combined in the elementary proposition.

On the other hand, there are no substantive propositions that are *a priori* true:

> 2.225 There are no pictures that are true *a priori*.

A priori truths, such as the truths of logic, do not express substantive propositions but only tautologies:

> 6.1 The propositions of logic are tautologies.
>
> 6.11 Therefore the propositions of logic say nothing. (They are the analytic propositions.)

Thus, propositions are either *a priori* but analytic ("say nothing"), or do say something that is at most knowable empirically or *a posteriori*. If an apparent species of statement, therefore, is neither analytic nor *a posteriori*, its instances express no proposition whatsoever. For example, ethical judgements are neither analytic nor *a posteriori*.[3] Hence:

> 6.42 So too it is impossible for there to be propositions of ethics.
>
> 6.421 It is clear that ethics cannot be put into words.

The *Tractatus* can thus be viewed as attempting to formulate a test for the meaningfulness of an apparent species of statement, and under the influence of the *Tractatus*, the theory of meaning was central to the concerns of the Vienna Circle. One important motivation was the thought that clarity about what it is for a statement to be meaningful would help in getting clear on the proper scope and extent of philosophical enquiry itself. This would then help remedy the fact that philosophy, unlike science, appeared to make little or no progress either in the prosecution of its central concerns, or on the question as to how those concerns are properly to be conceived. The logical positivists attempted to give a rigorous account of what constitutes meaningfulness via the provision of a *criterion of significance*. The criterion aimed to give a systematic account of what counts, and what does not count, as literally meaningful or significant, or, alternatively, as possessing sense: if a sentence satisfies the criterion it counts as possessing sense, whereas if it fails to satisfy the criterion it counts as literally meaningless. The criterion would greatly aid the progress of philosophy by ensuring that philosophers do not get embroiled in literally meaningless metaphysical speculation: if the criterion of significance entails that sentences apparently about God, the nature of moral and aesthetic facts, or a reality which transcends the world of experience, are, appearances to the contrary, actually senseless, then we will have a clear explanation of why the metaphysical speculation about the *truth* of those sentences is utterly fruitless. If a sentence is literally meaningless, it cannot be true (or false), so speculation about its truth-value is simply misguided. As Ayer and Hahn put it:

> The originality of the logical positivists lay in their making the impossibility of metaphysics depend not upon the nature of what could be known but upon the nature of what could be said. Their charge against the metaphysician was that he breaks the rules which any utterance must satisfy if it is to be literally significant.[4]

> There is no possibility of piercing through the sensible world disclosed by observation to a "world of true being": any metaphysics is impossible! Impossible, not because the task is too difficult for our human thinking, but because it is meaningless, because every attempt to do metaphysics is an attempt to speak in a way that contravenes the agreement as to how we wish to speak, comparable to the attempt to capture the queen (in a game of chess) by means of an orthogonal move of the bishop.[5]

According to the criterion suggested by the logical positivists, there are only two ways in which a sentence can be literally significant: by expressing an *a posteriori thought*, a thought which is empirically verifiable, or by expressing an *a priori thought*, a thought which is true purely in virtue of facts about meaning (sense). The first type of statement is dealt with by the positivists' *verification principle*, which attempts to spell out in detail what conditions a statement has to satisfy in order to qualify as literally meaningful in virtue of being susceptible to empirical verification.[6] The second type of statement is dealt with by the positivists' account of *a priori* truth: *a priori* truths are *analytic*, in the sense that they are true purely in virtue of meaning (sense). In §3.2 and §3.3 we look at the attempt to spell out the verification principle in *Language, Truth, and Logic*, the classic popularisation of the logical positivist doctrine by A. J. Ayer (1910–1989); and in §3.4 we investigate the positivist account of the *a priori*, via a discussion of Chapter 4 of that work, and we'll see how Ayer attempts to find a place for philosophy itself as an *a priori* activity. The dichotomy between statements which are meaningful in virtue of being empirically verifiable, and statements which are true purely in virtue of meaning, receives further development in the famous distinction between *internal and external questions* developed by Rudolf Carnap (1891–1970): that distinction is the subject of §3.5. In §3.6 we briefly consider the impact of the logical positivist theory of meaning on ethical language. Finally, in §3.7 we look at some further aspects of Ayer's views on empirically verifiable statements: this sets the scene for Quine's assault on the logical positivists' theory of meaning, which is the subject of our next chapter.

3.2 The formulation of the verification principle

The verification principle can be summed up in Schlick's famous slogan, "the meaning of a statement consists in its method of verification". Statements which qualify as literally meaningful in virtue of being associated with a method of empirical verification are deemed to possess *factual meaning*: statements can thus be literally meaningful without possessing factual meaning if they are *a priori* and analytic (tautologies, in the terminology favoured by Ayer). This latter type of statement will be dealt with in §3.4: our concern in this section is with statements that are literally meaningful in virtue of possessing factual meaning, in virtue of trading in matters of fact.

Ayer first introduces the criterion of factual significance in the following passage:

> The criterion which we use to test the genuineness of apparent statements of fact is the criterion of verifiability. We say that a sentence is factually significant to any given person if, and only if, he knows how to verify the proposition which it purports to express – that is, if he knows what observations would lead him, under certain conditions, to accept the proposition as being true, or reject it as being false. If, on the other hand, the putative proposition is of such a character that the assumption of its truth, or falsehood, is consistent with any assumption whatsoever concerning the nature of his future experience, then, as far as he is concerned, it is, if not a tautology, a mere pseudo-proposition.[7]

Interestingly enough, almost immediately after this passage, Ayer goes on to suggest that in addition to its application to putative propositions, the verification principle can also be applied to putative questions:

> And with regard to questions, the procedure is the same. We enquire in every case what observations would lead us to answer the question, one way or the other; and, if none can be

> discovered, we must conclude that the sentence under consideration does not, as far as we are concerned, express a genuine question, however strongly its grammatical appearance may suggest that it does.[8]

This is interesting because it shows that Ayer's intention, in formulating the verification principle, is not just to provide a criterion for drawing a distinction within the class of *declarative* sentences, between those which possess factual meaning, with genuine factual content, apt to be true or false, and those which, grammatical appearances to the contrary, possess no such content. The verification principle can also be applied within the class of *interrogative* sentences, to distinguish those interrogatives which express genuine questions, from those which merely appear to do so. And there seems to be every reason to expect that the application of the principle can be further widened to include, for example, *imperatival* sentences, expressions of *intention* and expressions of *desire*. With respect to imperativals, the principle would distinguish between those sentences which have the grammatical appearance of commands, and which express genuine commands, and those with the same grammatical appearances which do not: a sentence in the imperatival mood would express a genuine command if there is a series of observable actions, which, when carried out, would constitute obeying the command. Likewise for expressions of intention: a sentence with all the grammatical appearances of an expression of intention would express a genuine intention only if there is a series of observable actions which, when carried out, would constitute the implementation of the intention. And similarly for expressions of desire: a putative expression of desire would only count as a genuine expression of desire if there were some observable events whose occurrence constitutes satisfaction of the desire.[9] This suggests that Ayer's application of the verification principle is perhaps much wider than has been generally appreciated: the principle is intended not only to distinguish between sentences which have factual content and those which do not, but also to draw a distinction in point of genuineness within classes of sentences that are admitted to have no fact-stating function to begin with. The point of the verification principle is not so much

to distinguish between sentences which possess literal significance and sentences which do not; rather it is to distinguish between sentences which possess *some sort of significance* (be it literal significance, imperatival significance, conative significance, and so on) and sentences which possess *no* sort of significance whatsoever.[10] We will return to this issue in §3.6.

Having made the more general import of the verification principle clear, though, we shall concentrate on declarative sentences, and on what the principle has to tell us about their possession of factual significance. After the passages quoted above, Ayer proceeds to draw two distinctions: between *practical* verifiability and verifiability *in principle*, and between what he terms the "strong" and "weak" senses of verifiability. In short, a statement is practically verifiable for me if I possess the practical means of placing myself in a situation in which I would make observations sufficient for verifying that statement; a statement is verifiable in principle, if, though I do not possess the practical means of placing myself in such a situation, I can nevertheless theoretically conceive of what sort of position I would need to be in, and what observations I would need to make, in order to verify the statement. Thus, the statement "I have five pennies in my pocket" is verifiable in practice, whereas "There are horses on a planet in Alpha Centauri" is merely verifiable in principle. A statement is *strongly* verifiable if its truth could be *conclusively* established in experience, while a statement is said to be *weakly* verifiable if it is possible for experience to render its truth *probable*. Ayer in each case chooses the more liberal of the two notions – i.e. verifiability in principle, and weak verifiability – in spelling out what is involved in the verification principle. We require only verifiability in principle, in order to conserve the factual meaning of statements such as "There are horses on a planet in Alpha Centauri"; and we require only weak verifiability, in order to avoid ruling out scientific generalisations, such as "Iron tends to expand when heated".

Given this, the question we must ask in order to determine whether a given indicative sentence possesses genuine factual meaning is not: "Would any possible observations make its truth or falsehood logically certain?", but rather: "Would any possible observations be relevant to the determination of its truth or falsehood?".

But what does it mean to say that a statement is such that there are possible observations which are relevant to determining whether it is true or false, or that there are observations which would count as evidence for it? Ayer attempts to spell this out as follows:

> Let us call a proposition which records an actual or possible observation an experiential proposition. Then we may say that it is the mark of a genuine factual proposition, not that it should be equivalent to an experiential proposition, or any finite number of experiential propositions, but simply that some experiential propositions can be deduced from it in conjunction with certain other premises without being deducible from those other premises alone.[11]

This implies that statements about practically inaccessible parts of space and time and scientific generalisations and statements of law do have genuine factual content, as we would hope and expect. But it also implies, given the assumption that they are not analytic, that statements like "the world of sense-experience is unreal" are devoid of factual meaning: "no conceivable observation . . . could have any tendency to show that the world revealed to us by sense-experience was unreal".[12] And it implies, when generalised, that a question such as "Is reality ultimately composed of one substance or many?" (the traditional dispute in metaphysics between monism and pluralism) is not a genuine question: no possible observation could provide an adequate basis for answering it one way rather than the other.[13]

However, as Ayer admits in the long introduction to the second edition of *Language, Truth, and Logic*, this formulation of the verification principle fails. Take some experiential proposition (or "observation statement", in his new terminology) "O", and any "nonsensical" statement "N". Then, since the observation statement "O" can be deduced from "N", together with "if N then O", but not from "if N then O" on its own, "N" counts as factually meaningful according to this formulation of the verification principle. Thus, if we take "N" to be "the Nothing nothings" and "O" to be "that pillar box is red", we can establish that "the Nothing nothings" is factually significant. This shows that Ayer's initial formulation of the verification principle implies that any indicative

statement whatsoever possesses factual significance: in other words, that it lays down no constraints whatsoever on the ascription of factual meaning.

In response to this problem, Ayer, in the introduction to the second edition, suggests a more sophisticated version of the principle of verification. He does this by defining verifiability in a recursive fashion: he begins with the notion of an observation statement, defines a notion of *direct verifiability* in terms of it, and then proceeds to define a notion of *indirect verifiability* in terms of direct verifiability. The verification principle is then reformulated as follows: in order for a statement to possess factual meaning, it has to be either directly or indirectly verifiable.

So how are direct and indirect verifiability defined? Ayer defines direct verifiability as follows:

> [A] statement is directly verifiable if it is either itself an observation statement, or is such that in conjunction with one or more observation statements it entails at least one observation statement which is not deducible from these further premises alone.[14]

Note that "N", in the counterexample raised above against the initial formulation of the principle, does not qualify as directly verifiable: it is not an observation statement, and the fact that it entails "O" when conjoined with "if N then O" does not qualify it as directly verifiable, since "if N then O" is not an observation statement, as required by the definition of direct verifiability. Indirect verifiability is defined as follows:

> a statement is indirectly verifiable if it satisfies the following conditions: first, that in conjunction with certain other premises it entails one or more directly verifiable statements which are not deducible from these other premises alone; and secondly, that these other premises do not include any statement that is not either analytic, or directly verifiable, or capable of being independently established as indirectly verifiable.[15]

On the face of it, it looks as if "N" in the original counterexample fails to qualify as indirectly verifiable too, so that it is not after all certified as factually significant by the verification principle.

Unfortunately, however, even this new and more sophisticated formulation of the verification principle turned out to be susceptible to counterexamples. Alonzo Church constructed an example to show that, like the original formulation in the first edition of *Language, Truth, and Logic*, the revised formulation too admits any non-analytic statement (and hence any arbitrary "nonsensical" statement) as factually significant.[16]

Take S = (not-O1 and O2) or (not-N and O3), where O1, O2, and O3 can be any trio of observation statements that are all logically independent of one another, and N is the "nonsensical" statement of your choice (e.g. "the Nothing nothings"). Then S qualifies as directly verifiable on Ayer's definition. To see this, note that in conjunction with O1, S entails O3 (this can be demonstrated quite easily, using elementary logic of the sort taught in first year university courses, so this is left as an exercise for the reader). Thus S entails an observation statement (O3) when some other statements (O1) are assumed. But, by hypothesis, O1 is logically independent of O3, and so does not entail it. So S entails an observation statement (O3) in conjunction with some other observation statements (O1), though that observation statement (O3) does not follow from those other observation statements (O1) alone. S is thus directly verifiable.

Given that S is directly verifiable, it is now easy to establish that our "nonsensical" N is indirectly verifiable. Note that N, when conjoined with S, entails O2 (again, this is simple to prove, and is left as an exercise). O2 is an observation statement, and thus qualifies as directly verifiable. Thus, N entails a directly verifiable statement (O2) when conjoined with S, though O2 does not follow from S on its own. (To see why O2 does not follow from S on its own, note that there is a possible assignment of truth-values to the constituents of S on which S comes out true even when O2 is false: take N false, O3 true, O1 true, and O2 false.) N thus counts as indirectly verifiable on Ayer's definition, and hence as factually significant even on the revised version of the verification principle. Once again, it seems that the verification principle places no real constraints on the ascription of factual meaning.

Of course, the defender of the verification principle may now attempt to further modify the principle so that it excludes the

offending counterexample, and the literature on the topic is replete with attempts at such modifications. But we shall not here enter further into questions about the precise formulation of the verification principle: details of the relevant literature are contained in the guide to further reading at the end of this chapter. Rather, we shall consider a more general question about the nature of the principle.

3.3 Foster on the nature of the verification principle[17]

John Foster has discerned two distinct versions of the principle of verification in *Language, Truth, and Logic*: the *evidence principle* and the *content principle*, respectively. The evidence principle corresponds to Ayer's first-edition formulation, which attempted to demarcate the class of factually significant statements as those on which observational evidence could bear: the factually significant statements are those for which there are possible observations which would be relevant to the determination of the statements' truth or falsehood. We saw above that the evidence principle, as formulated by Ayer in the first edition of *Language, Truth, and Logic*, allowed any statement to count as factually significant. But as Foster points out, one way in which Ayer could deal with this problem would be to return to the original formulation and then "try to impose on the notion of evidence-for-or-against whatever restrictions are needed for a suitable criterion of significance".[18] Ayer could thus attempt to provide a modified version of the evidence principle. But, as we saw in the previous section, this is not in fact how Ayer proceeds in the introduction to the second edition. Ayer there attempts to formulate a version of what Foster terms the content principle, according to which "the factual significance of a statement lies in its observational content – that is, in its contribution to the deduction of observation-statements" and "a statement has factual significance if and only if its content is purely observational, i.e. if and only if the statement falls within the scope of an observational language".[19] The content principle thus encapsulates a stronger form of verificationism than the evidence principle. This can be seen by considering scientific hypotheses

which postulate the existence of objects which are not directly observable (e.g. electrons and protons). As Foster puts it:

> "Realistically" construed, these hypotheses do not fall within the scope of an observational language, but they are still held to be subject to observational tests, according to how well or badly they explain our observational data. So under the content-principle they count as nonsensical, but under the evidence principle they qualify as factually significant – at least they do so unless, on the lines of the radical sceptic, we reject the epistemological methods of science altogether. And quite generally, we can see that given ordinary (non-sceptical) standards of what counts as evidence, the content principle yields a much more stringent criterion of factual significance than its rival.[20]

The content principle actually endangers the factual significance of many everyday statements about physical objects, other minds, and so on, which we (and Ayer) would intuitively expect to be factually significant. For to count as factually significant by the lights of the content principle, statements such as *The table is square* or *Jones has toothache* will have to be shown to be translatable into statements which do have explicitly observational content. And such attempts at reductive analysis have been singularly unsuccessful, with the failure of Ayer's phenomenalistic reductionism being a case in point:

> Any attempt to translate statements about the physical world into statements about sense-contents, or statements about other minds into statements about behaviour, seems to involve a radical distortion of what the original statements mean.[21]

Given that the content principle requires the tenability of unpromising reductionism about many of our prima facie factually significant statements, wouldn't it be more advisable, in attempting to reformulate the verification principle so as to avoid the various counterexamples, to go back to the more liberal evidence principle, and base the attempted reformulation on it? Foster argues, however, that this more promising line of development is closed off to Ayer:

> [I]f there are any grounds for accepting the verification principle, they are grounds for accepting it in its content-restricting form. The less stringent evidence principle may give more plausible results, but, by its very liberality, it becomes devoid of rationale.[22]

If Foster is right in this claim, it is bad news indeed for anyone sympathetic to the logical positivist project, but sceptical about the possibility of plausible reductive analyses of our everyday talk about physical objects and other minds. But as we shall now see, the argument which Foster relies on to derive the conclusion that the verification principle must be formulated in the strong content-restricting manner is far from conclusive.

Foster argues as follows. First of all, "let us suppose that contrary to the content-principle, we take physical statements to be factually significant and construe them, 'realistically', as irreducibly about some mind-independent reality".[23] Although this view of the semantics of physical statements is contrary to the content principle, it is allowed by the evidence principle: even though physical objects are not directly observable, "physical statements are open to observational evidence through their explanatory role with respect to sense-experience". That is to say, certain physical statements – concerning some mind-independent reality – may be verified by the fact that they provide the best explanation of the available experiential data. They are thus secured factual significance under the evidence principle. But now comes the crunch, according to Foster: "given our realistic construal of physical statements, we could hardly regard their openness to this kind of observational evidence as a prerequisite of their having the significance we attach to them". Why not? Well, imagine a *radical sceptic*, someone who denies that physical statements are open to observational evidence in the way described. Then, according to Foster,

> We would not see this as a challenge to our semantic position. We would not see it as something which, if we accepted it, would oblige us to abandon the claim that physical statements are both factually significant and concerned with a mind independent reality. Rather, we would see it as raising a separate and purely epistemological issue . . . We may reject the scepticism;

> but both we and the sceptic would accept that the epistemological issue only arises because the significance of physical statements, construed realistically, is already secure.[24]

So, we would view the radical sceptic as raising a purely epistemological worry, and even in allowing that there *is* such an epistemological issue, we would be conceding the significance of the physical object statements: more precisely, we would be granting their significance *before* the question about their openness to observational evidence has been resolved. So "Unless we accept the content-principle, there is no rationale for making openness to observational evidence a requirement of factual significance".[25] If we are going to accept a version of the verification principle, we thus have no option but to accept it in its more stringent – and hence far more unpromising – form as the content principle.

How plausible is this argument of Foster's? There are at least two reasons for doubting whether it does actually establish that the verification principle can only ever be formulated as the content principle.

First of all, Foster faces the following dilemma. The "radical sceptic" can be construed as a sceptic who questions the bearing of *a certain piece (or set of pieces) of observational evidence* on the physical statements in question. When the sceptic is thus construed, Foster's claim that we would not treat this scepticism as even posing a threat to the factual significance of the statements in question seems plausible enough. But this sort of radical sceptic is not the sort of sceptic we should focus on in a discussion of the evidence principle. What we require is a much more radical sceptic, a sceptic who goes further and questions whether *there is even any possible* observational evidence which could bear on the physical statement in question. Is it so obvious that we would not treat this sort of sceptical claim as posing a threat to the significance of the relevant physical statements? The verificationist intent on defending the evidence principle will likely – and quite rightly – see this as a completely unmotivated claim. In fact, it clearly amounts only to the bald assertion that the evidence principle is false, and as such can hardly provide any independent reasons for thinking that that principle has no rationale.

Secondly, Foster's argument appears to make the following underlying assumption: you cannot raise an epistemological doubt about a statement P without already granting that P is meaningful. On the face of it, this may seem plausible enough. But when the truly radical nature of the scepticism involved is appreciated, it begins to look less plausible. One way of seeing this is to note that certain well-known forms of constitutive scepticism can be advanced by epistemological routes. We shall look at some such versions of constitutive scepticism in the next two chapters. In Chapter 5 we shall see that Kripke's Wittgenstein's radical sceptic attempts to argue for constitutive scepticism about meaning ("that there is no such thing as meaning") via an epistemological route. Kripke's Wittgenstein questions whether there is any fact of the matter corresponding to claims about meaning, such as "I currently mean the addition function by the '+' sign", and does so in the following way. He first of all allows us idealised epistemological access to all of the facts about our qualitative mental lives, our behavioural histories and dispositions, and so on. He then claims that even when such idealised access is granted, we could never be in a position to justify any particular claims about what we mean. The conclusion he draws is that there is, contrary to our original intuitions, no such thing as a fact about what we mean. Now, this sceptical argument, and the resulting constitutive scepticism, can indeed be criticised on a number of scores, as we shall shortly see. But the important point for our present purposes is that the mere fact that such an argumentative strategy is possible – constitutive scepticism advanced via an epistemological route – shows that Foster's assumption is wrong. When the scepticism in question is sufficiently radical, we can raise an epistemological doubt about a statement without conceding its meaningfulness, because we can use the radical epistemological doubt as a *route* to the conclusion that the statement is not after all factually meaningful.[26]

The conclusion to be drawn from this is as follows. Ayer's original attempts at formulating the evidence principle and content principle both fail, because neither sets down any genuine constraints on the ascription of factual significance. There is thus a pressing need for a revised formulation of the verification principle. But in seeking such a revised formulation, we are not constrained, in the

manner argued by Foster, to seek for a version of the stronger and more improbable content principle. For all that Foster has shown, a revised version of the evidence principle may indeed be the best bet for the defender of the verification principle. However, we cannot discuss the verification principle further at this point. In the next section we move on the look at the logical positivist account of the other class of literally significant statements, the *a priori*.

3.4 The *a priori* and the linguistic theory of necessity

According to logical positivism, there are only two ways in which a statement can qualify as literally possessing sense: by possessing factual significance, in the sense of factual significance defined by the verification principle, or by being *analytic*. Ayer, in Chapter 4 of *Language, Truth, and Logic*, defines the notions of analytic and synthetic as follows: "a proposition is analytic when its validity depends solely on the definitions of the symbols it contains, and synthetic when its validity is determined by the facts of experience".[27] This is repeated in the introduction to the second edition: "a proposition is analytic if it is true solely in virtue of the meaning of its constituent symbols".[28] The class of literally significant statements is thus exhausted by, on the one hand, the class of statements which in principle admit of empirical verification, and, on the other, the class of statements which are true (or false) purely in virtue of the meanings or senses of their constituent expressions.

What, then, of the putative truths of logic and mathematics? The logical positivists wish to secure literal significance for logical and mathematical statements, such as the law of non-contradiction, the law of excluded middle, and the statement that 2 + 2 = 4. But the necessity of these statements prevents them from qualifying as literally significant in virtue of possessing factual significance:

> As Hume conclusively showed, no general proposition whose validity is subject to the test of actual experience can ever be logically certain. No matter how often it is verified in practice, there still remains the [logical] possibility that it will be confuted on some future occasion. The fact that a law has been

> substantiated in *n–1* cases affords no *logical* guarantee that it will be substantiated in the *n*th case also, no matter how large we take n to be. And this means that no general proposition referring to a matter of fact can ever be shown to be necessarily and universally true. It can at best be a probable hypothesis.[29]

Or, as Hans Hahn put it slightly earlier:

> Observation discloses to me only the transient, it does not reach beyond the observed; there is no bond that would lead from one observed fact to another, that would compel future observations to have the same result as those already made . . . Whatever I know by observation could be otherwise.[30]

Given this, there appear to be three choices available: (i) secure factual significance for the statements of logic and mathematics by denying that they are necessarily true; (ii) argue that they are analytic; or (iii) reject them as literally senseless, as failing to possess literal significance.

As noted above, Ayer has no wish to embrace option (iii), and finds that option (i), the option pursued by John Stuart Mill (1806–1873), "to be discrepant with relevant logical facts".[31] If we look at cases in which, for example, a mathematical truth such as the statement that $2 \times 5 = 10$ appears to be confuted by experience, we will find that giving up the relevant mathematical statement is never a genuine option. Suppose I look into the cupboard and observe five pairs of shoes. I then count the individual shoes, and find that they are only nine in total. Ayer argues that there are no possible circumstances in which we would take this sort of episode as establishing the falsity of the statement that $2 \times 5 = 10$: we would perhaps say that my initial observation of five pairs of shoes was mistaken, that one of the shoes vaporised during the process of counting, that two of shoes had coalesced into one, or that I had made a slip while counting. But "the one explanation which would in no circumstances be adopted is that ten is not always the product of two and five".[32]

Now we might well wish to question this argument of Ayer's. The relevant consideration is surely not what explanations we would or would not *in fact* invoke in the envisioned circumstances,

but rather what explanations would be *rationally optional*. Ayer needs to argue not only that giving up the mathematical statement that $2 \times 5 = 10$ is something we would never opt for *in practice*, but that it is an option we can choose only on pain of irrationality. If mathematical statements are, as he claims, analytic, then this may indeed follow: in rejecting the statement we would be violating one of the conventions that governs our use of language, and in that sense, we would be violating a norm of rationality. But he cannot *assume* that mathematical statements are analytic at this stage without simply begging the question against Mill and the defenders of option (i). Be that as it may, Mill's view that the statements of logic and mathematics are *a posteriori* and contingent is widely rejected, so we will not discuss it any further here. Instead, we shall concentrate on assessing the plausibility of Ayer's defence of option (ii), the claim that the statements of logic and mathematics are analytic.

Ayer's defence of the claim that mathematics and logic are analytic takes the following form. He argues that it enables us to explain the central – and historically problematic – features of both of those disciplines. The first feature has already been mentioned: the *necessity* of mathematics and logic. How can the claim that mathematics and logic cannot possibly be confuted in experience be squared with the widespread belief that we possess knowledge of mathematical and logic truths? Kant, in order to answer this question, claimed that in the case of geometry, for example, "space is the form of intuition of our outer sense, a form imposed on us by the matter of sensation".[33] Likewise, in order to account for the necessity of arithmetic, Kant postulated that it was "concerned with our pure intuition of time, the form of our inner sense".[34] Ayer's thought is that we can explain the necessity of geometry, arithmetic, and logic, without getting embroiled in the metaphysical and epistemological extravagances of Kant's transcendental idealism. The truths of mathematics are analytic in the sense that they "simply record our determination to use words in a certain fashion".[35] Given this, it is easy to see why they are necessary and logically certain:

> We cannot deny them without infringing the conventions which are presupposed by our very denial, and so falling into

> self-contradiction. And this is the sole ground of their necessity . . . There is nothing mysterious about the apodeictic certainty of logic and mathematics. Our knowledge that no observation can ever confute the proposition "7 + 5 = 12" depends simply on the fact that the symbolic expression "7 + 5" is synonymous with "12", just as our knowledge that every oculist is an eye-doctor depends on the fact that the symbol "eye-doctor" is synonymous with "oculist". And the same explanation holds good for every other *a priori* truth.[36]

According to this view – the "linguistic theory of necessity" – mathematical and logical truths arise from conventional connections between the symbolic expressions of our language. In denying such truths, then, we deny the very conventions which make the meaningful use of language possible. The denial of an analytic truth is in this sense "self-stultifying", and this accounts for the necessity which analytic statements possess. As Hahn puts it:

> Logic does not by any means treat of the totality of things, it does not treat of objects at all but only of our way of speaking about objects: logic is generated by language. The certainty and universal validity, or better, the irrefutability of a proposition of logic derives just from the fact that it says nothing about objects of any kind.[37]

The fact that it is impossible to find an object which is both red all over and not-red all over, for example, stems not from the way in which the mind and the noumenal world co-operate in the transcendental constitution of the world of experience, but from the conventions underlying our use of language. The law of non-contradiction is not a deep truth about reality, but just a prescription of a method of speaking about things: it expresses our conventional stipulation that "the designation 'red' is to be applied to some objects and the designation 'not-red' to all other objects".[38]

If Ayer is right, the necessity of logical and mathematical statements can thus be explained. We now move on to the second feature that has troubled philosophers of logic and mathematics: their *utility*. If mathematics and logic "say nothing about objects

of any kind", or have no factual significance, does it follow that it is a mistake to think that there can be genuine *discoveries* in mathematics and logic? And what of the usefulness of these *a priori* disciplines? If mathematical statements merely illustrate the way in which we use certain symbols, how could we ever make any interesting discoveries in mathematics, and how could such "discoveries" have any utility? Ayer and Hahn answer this worry as follows:

> The power of logic and mathematics to surprise us depends, like their usefulness, on the limitations of our reason. A being whose intellect was infinitely powerful would take no interest in logic and mathematics. For he would be able to see at a glance everything that his definitions implied, and, accordingly, could never learn anything from logical inference which he was not fully conscious of already.[39]
>
> Logical propositions, though being purely tautologous, and logical deductions, though being nothing but tautological transformations, have significance for us because we are not omniscient. Our language is so constituted that in asserting such and such propositions we implicitly assert such and such other propositions – but we do not see immediately all that we have implicitly asserted in this manner. It is only deduction that makes us conscious of it.[40]

To take the simplest sort of case: if I know that Jones is not wearing a rose in his buttonhole, and that either Jones is wearing a rose in his buttonhole or is wearing a carnation in his buttonhole, I may not be consciously aware that it follows from these pieces of knowledge that Jones is wearing a carnation in his buttonhole. I may fail to pick up one of the implications of the factual knowledge I currently possess. But my knowledge of the logical truth "if (not-P and (P v Q)) then Q" allows me to work out the *consequences* of what I already know: in this case, that Jones is wearing a carnation in his buttonhole. In this way logic allows me – as a mortal, finite being, not aware of all of the logical implications of my current factual knowledge – to work out those implications. It thus extends my knowledge, and can serve as an instrument of

discovery: and such discoveries can be genuinely interesting and surprising when the process of deriving those consequences is very intricate and complicated, as it often is in logical and mathematical practice. That said, however, it should be noted that there is a restriction on the way in which logic and mathematics can extend our knowledge. As Ayer puts it, "there is a sense in which they may be said to add nothing to our knowledge. For they tell us only what we may be said to know already".[41] Thus, in the simple case above, if I know that Jones is not wearing a rose in his buttonhole, and that Jones is wearing either a rose or a carnation in his buttonhole, I know *already*, or know *implicitly*, that Jones is wearing a carnation in his buttonhole. What logical deduction does is enable me to make this implicit knowledge *explicit*. Logic thus never widens the extent of my factual knowledge: it serves merely to make explicit that factual knowledge which previously I possessed only implicitly. But this process of rendering implicit knowledge explicit, Ayer and Hahn will claim, is enough to secure the utility of mathematics for finite beings like ourselves, and enough to show how there can be a sort of discovery in those disciplines.

The logical positivist account of the *a priori* thus seems attractive: it accounts for the necessity and utility of logic and mathematics without incurring the metaphysical and epistemological debts of rival conceptions, such as Kant's. But is it plausible? We shall see in a moment that there are strong reasons for thinking that though attractive, the logical positivist account of *a priori* truth is deeply flawed. Before that, however, it might be worthwhile to pause for a second to reflect on the implications of the discussion so far for the nature and practice of philosophy itself. Philosophy itself is an *a priori* activity, so it follows from the logical positivist account of the *a priori* that the propositions of philosophy are themselves analytic. Philosophy is not a body of factual knowledge, but rather an activity of *analysis*, as the following two quotes from Russell and Carnap make clear:

> "Logical Positivism" is a name for a method, not for a certain kind of result. A philosopher is a logical positivist if he holds that there is no special way of knowing that is peculiar to philosophy, but that questions of fact can only be decided by

> the empirical methods of science, while questions that can be decided without appeal to experience are either mathematical or linguistic.[42]
>
> There is no such thing as a speculative philosophy, a system of sentences with a special subject matter on a par with those of the sciences. To pursue philosophy can only be to clarify the concepts and sentences of science by logical analysis.[43]

In this, Russell and Carnap are echoing some of Wittgenstein's pronouncements on the nature of philosophy in the *Tractatus*:

> 4.112 Philosophy aims at the logical clarification of thoughts. Philosophy is not a body of doctrine but an activity.
>
> A philosophical work consists essentially of elucidations.
>
> Philosophy does not result in "philosophical propositions", but rather in the clarification of propositions.
>
> Without philosophy thoughts are, as it were, cloudy and indistinct: its task is to make them clear and to give them sharp boundaries.

Just as the analytic nature of logic and mathematics is consistent with their utility and interest, so too is that of philosophy itself. Ayer cites Russell's theory of definite descriptions as a good example of the utility of philosophical analysis: the correct analysis of statements involving descriptive phrases ensures that we do not get lost in the undergrowth of Meinong's Jungle (see §2.8 above). There is much that can be said about this extremely influential conception of the nature of philosophy, but we must now move on to assess the plausibility of the logical positivist account of the *a priori* itself.

The main objection to the idea that *a priori* truths are true in virtue of the conventions governing our use of language is due to Quine, in a paper first published in 1936:

> In the adoption of the very conventions whereby logic itself is set up, however, a general difficulty remains to be faced. Each of these conventions is general, announcing the truth of every

> one of an infinity of statements conforming to a certain description; derivation of the truth of any specific statement from the general convention thus requires a logical inference, and this involves us in an infinite regress.[44]

This is somewhat cryptic, but Quine's idea can be unpacked as follows. Take a logical truth such as

(1) If (Jones is English and (If Jones is English then Jones is British)) then Jones is British.

According to logical positivism, this is true in virtue of convention. But which convention? Do we have a separate convention for (1), stipulating that (1) is to be held true come what may? This will obviously not do, since there is an infinite number of logical truths (e.g. we can generate an infinite number of substitution instances of (1), "If (Jacques is French and (if Jacques is French then Jacques is European)) then Jacques is European, and so on, *ad infinitum*), and we can hardly be credited with setting up an infinite number of separate conventions, one for each individual truth of logic. So the convention which governs (1) must be a *general* convention, from which the convention governing (1) is derivable. What might that convention be? Consider

(2) If (P and (If P then Q)) then Q.

Then we might state the convention which covers (1) as follows:

(3) If (1) results from (2) by uniform substitution of P and Q, then that statement is to be held true come what may.

Given this, and

(A) (1) results from (2) by uniform substitution of P and Q

we can now derive the convention specifically governing (1):

(4) (1) is to be held true come what may.

Thus we have generated, on the basis of a general convention, a convention specifically governing (1). But in deriving (4) from this more general convention we have relied on logic itself! In deriving (4) from (3) and (A) we have relied on

(5) If ((1) results from (2) by uniform substitution of P and Q and (If (1) results from (2) by uniform substitution of P and Q then (1) is to be held true come what may)) then (1) is to be held true come what may.

But (5) is of the same logical form as (1)! In order to secure its status as true in virtue of convention we would need

(6) If (5) results from (2) by uniform substitution of P and Q, then (5) is to be held true come what may.

Given

(B) (5) results from (2) by uniform substitution of P and Q

we could then derive

(7) (5) is to be held true come what may.

However, in deriving (7) from (6) and (B) we have relied on logic again! In this derivation we rely upon

(8) If ((5) results from (2) by uniform substitution of P and Q and (If (5) results from (2) by uniform substitution of P and Q then (5) is to be held true come what may)), then (5) is to be held true come what may.

But, again, this is of the same logical form as (1), so in order to secure its status as true in virtue of convention we would need

(9) If (8) results from (2) by uniform substitution of P and Q, then (8) is to be held true come what may.

And now we are off on an infinite regress: deriving the conclusion that (8) is to be held true come may will in its turn presuppose a logical truth, and so on, *ad infinitum*.

The dilemma faced is thus as follows. We either suppose that (5) is true in virtue of convention, in which case our account leads us off on an infinite regress, because we have to presuppose logic itself in order to derive the relevant convention. Or, on the other hand, we suppose that (5) is not true in virtue of convention, in which case we have conceded that the logical positivist account fails with respect to (5) (and by implication, with respect to (1) as well). Either way, the account grinds to a halt. As Quine himself sums matters up, "the difficulty is that if logic is to proceed mediately from the conventions, logic is needed for inferring logic from the conventions".[45]

The logical positivist account of the *a priori* thus looks at the very best an inherently unstable position.[46] The notion of convention has resurfaced in more recent philosophy of language in the work of David Lewis, but we cannot discuss Lewis's account of convention here. In the next section, we move briefly away from Ayer's conception of the distinction between empirically verifiable and analytic statements, and take up another historically influential angle on that distinction, Carnap's famous distinction between "internal" and "external" questions.

3.5 Carnap on internal and external questions

Rudolf Carnap (1891–1970) was one of the most influential members of the Vienna Circle. An appreciation of his work is extremely important, not just for getting to grips with logical positivism, but for understanding many of the subsequent developments in analytic philosophy. In this section, however, we will look only at one strand in Carnap's philosophy, the distinction between what he termed "internal" and "external" questions.

In his "Empiricism, Semantics, and Ontology", Carnap reflects on some of the metaphysical disputes that have traditionally exercised philosophers. For example, there is the dispute between

Nominalists and Platonists concerning the ontological status of abstract entities, such as numbers, sets, and propositions. Traditionally, Platonists about, for example, arithmetic take the truth of statements such as "5 is a prime number" to imply the existence of a range of abstract objects, numbers, not existing in space and time, and completely disconnected from the causal order of concrete entities. Nominalists, thinking Platonism to be ontologically profligate and epistemologically far-fetched, take a different tack: they attempt to analyse statements like "5 is a prime number" in such a way that their truth does not imply the existence of abstract entities. Carnap introduces a distinction which is intended to pull the rug out from under the feet of both sides in this traditional debate. In order to explain this distinction, we need first to explain Carnap's notion of a linguistic framework.

Very roughly, a linguistic framework consists of a vocabulary, together with a set of rules and conventions governing the use of that vocabulary. Consider the case of arithmetic. In arithmetic, a linguistic framework would consist of vocabulary, such as numerals ("3", "4", "5", and so on), variables ("x", "y", "z"), general terms ("is a prime number", "is an integer"), and various conventions (". . . is a number" is true, where any numeral may stand in place of the dots; ". . . is a prime number" is true when and only when the numeral replacing the dots stands for a number which is divisible only by itself and unity; and so on). This is an example of what Carnap calls a "logical framework" (we'll see why in just a moment). But it is also possible to have a "factual framework". For example, we could have a framework consisting of mass terms ("sugar", "salt", "beer"), general terms ("sweet", "spicy", "hoppy"), and linguistic conventions (". . . is soluble" is false if the mass term replacing the dots stands for a substance which does not dissolve in water at normal temperature and pressure).

Using this notion of a linguistic framework, Carnap distinguishes between *internal* and *external* questions:

> It is above all necessary to recognize a fundamental distinction between two kinds of questions concerning the existence or reality of entities. If someone wishes to speak in his language about a new kind of entities, he has to introduce a system

> of new ways of speaking, subject to new rules; we shall call this procedure the construction of a linguistic *framework* for the new entities in question. And now we must distinguish two kinds of questions of existence: first, questions of the existence of certain entities of the new kind *within the framework*; we call them *internal questions*; and second, questions concerning the existence or reality *of the system of entities as a whole*, called *external questions*.[47]

So, for instance, in the arithmetical example, "Is there a prime number between 5 and 9?" is an internal question. Given the linguistic framework consisting of numerals and so on, we can answer it by applying the logical methods of arithmetic. For example, calculation reveals that the only divisors of 7 are itself and 1, so by the convention governing "prime", 7 is a prime number, so that there is a prime number between 5 and 9. We can see now why Carnap called this framework a logical framework: it is because answering the internal questions that are formulable given that framework requires only *a priori* calculation in addition to the application of the relevant conventions. In other words, the framework is a logical framework because the answers to the relevant internal questions are analytic. Matters stand differently in the case of the second linguistic framework we introduced. Consider an internal question such as "Is lead soluble?". In this case, answering the question requires, in addition to an application of the relevant convention, an appeal to empirical evidence and observation. We need to make the relevant observations to check whether lead dissolves in water at normal temperature and pressure. Some of the internal questions that can be answered given this framework therefore depend on *a posteriori* matters, and this explains why Carnap would call this framework a factual one.

Very roughly, then, a linguistic framework is a *logical* framework if all of the internal questions formulable within it admit of analytic answers, but a *factual* framework if at least some of the internal questions formulable within it admit of non-analytic answers.

We can thus see, using the terminology of the previous sections, that the answers to internal questions are literally significant,

since they are either analytic or resolvable by empirical means. What, though, of *external* questions? An example of an external question in the arithmetical case would be "Are there numbers?" or "Do numbers exist?". This is the sort of question which Platonists and Nominalists argue over: the Platonist claims that numbers, construed as abstract objects lying outside of space and time, do actually exist, whereas the Nominalist denies this. Carnap develops an ingenious dilemma which cuts against both sides in the dispute. Both sides in the dispute take themselves to be disagreeing over a substantial matter of fact: the existence of abstract entities. But the question they are debating can be interpreted either as an internal or an external question. If it is an internal question, the debate is not a substantial one, for the existence of numbers follows trivially and analytically from the linguistic framework. "5 is a number" is true by the first convention stated when introducing the framework, and "there exists an x such that: x is a number" immediately follows from this by a step of elementary logic. *This*, clearly, is not what the Platonists and Nominalists have been arguing about:

> If we were to ask them: "Do you mean the question as to whether the framework of numbers, if we were to accept it, would be found to be empty or not?", they would probably reply: "Not at all, we mean a question *prior* to the acceptance of the new framework".[48]

So the question must be an *external* one. And this takes us on to the second horn of Carnap's dilemma. For according to Carnap, external questions are not literally significant. External questions concern decisions we have to make about the adoption of the linguistic frameworks themselves, and these decisions are made purely on the basis of pragmatic considerations regarding the capacity of the framework to help us organise our thinking or predict and explain the course of future experience:

> The acceptance [of a linguistic framework] cannot be judged as being either true or false because it is not an assertion. It can only be judged as being more or less expedient, fruitful, conducive to the aim for which the language is intended . . . The

> acceptance or rejection of abstract linguistic forms, just as the acceptance or rejection of any other linguistic forms in any branch of science, will finally be decided by their efficiency as instruments, the ratio of the results achieved to the amount and complexity of the efforts required.[49]

There may be a substantial question as to which framework best satisfies the relevant pragmatic criteria. But the external questions themselves do not concern any issue of truth or falsity, only decisions based on purely pragmatic considerations of the type mentioned. The traditional metaphysical debate is really only a psuedo-debate, and the questions it raises are pseudo-questions. There are, says Carnap, no substantial, metaphysical questions about matters of ontological commitment: we are only fooled into thinking that there are such questions by confusing external questions for internal questions.

Carnap thus attempts to dissolve traditional metaphysical dispute about the existence of abstract entities, and he suggests that other traditional disputes can be dissolved in the same way (e.g. the debate between realists and subjective idealists about physical objects). The affinities of this position with that of Ayer's are clear. Before returning to Ayer, however, a few remarks about Christopher Hookway's recent discussion of Carnap's distinction are in order.

Hookway, in summing up the epistemological implications of Carnap's position, writes:

> When we attempt to reconstruct any body of knowledge, we can expect to find two things: (1) A framework of analytic principles, whose adoption reflects a practical decision based upon pragmatic considerations. Shared frameworks reflect adoption of conventions to the effect that these principles should be used. (2) A body of internal knowledge, which is justified by reference to the rules and principles that make up that framework.[50]

In summing matters up thus, though, Hookway is in danger of misleading the reader. By framing the contrast in terms of the distinction between analytic principles and internal knowledge, he seems to be suggesting that the distinction between external and internal matters mirrors the distinction between the analytic and

synthetic. But this is not so. As we saw in our exposition of Carnap, *both* of the categories of the analytic and the synthetic fall within the province of internal questions: analytic truths are those which follow from the framework's conventions alone, synthetic truths are those which follow from the framework's conventions in the presence of empirical evidence. *The conventions in the framework themselves are neither true nor false, since their adoption reflects a response to an external question*. That is to say, in referring to the conventions as "analytic principles", Hookway's presentation confuses the conventions which appear in the linguistic framework, the statements of which are neither true nor false, with the analytic truths which are answers to internal questions. The conventions themselves are not analytic truths, because they are not truths at all.[51]

We shall return to Carnap in the discussion of Quine in Chapter 4. Before that, however, we briefly return to Ayer.

3.6 Logical positivism and ethical language

In Chapter 6 of *Language, Truth, and Logic*, Ayer presents the emotivist theory of ethics. What can a logical positivist say about an ethical sentence like "Stealing is wrong"? The criterion of literal significance discussed in §3.2 appears to entail that sentences such as this are not literally significant: they are not analytic, and they do not admit of empirical confirmation or disconfirmation – there are no possible observations that we could make which would verify or disconfirm the sentence. Thus, "Stealing is wrong" is not the sort of thing that can be assessed for truth and falsity: it is not *truth-apt*. So "Stealing is wrong" does not even purport to represent some fact about the world: its semantic function is not to state facts. But what then is its semantic function? According to Ayer's emotivist theory the function of ethical statements is to express feelings or emotions:

> If I say to someone, "You acted wrongly in stealing that money", I am not stating anything more than if I had simply said, "You stole that money". In adding that this action is wrong, *I am not making any further statement about it. I am simply evincing my moral disapproval about it*. It is as if I had said, "You stole

> that money", in a peculiar tone of horror, or written it with the addition of some special exclamation marks. The tone, or the exclamation marks, adds nothing to the literal meaning of the sentence. *It merely serves to show that the expression of it is attended by certain feelings in the speaker*.[52]

It is important to be clear about the status of Ayer's conclusion that moral statements are not literally significant. He sometimes describes this conclusion as the claim that "ethical concepts are mere pseudo-concepts".[53] This might suggest that he was claiming that ethics is *nonsense*. This is further suggested by his remarks on metaphysical statements (such as "We have an enduring self" or "Reality is one substance and not many"): he concludes from the fact that these are neither analytic nor empirically verifiable that they are "nonsensical", and that metaphysics ought to be eliminated (the title of Chapter 1 of *Language, Truth, and Logic* is "The Elimination of Metaphysics"). So why isn't Chapter 6 similarly entitled "The Elimination of Ethics"? Why doesn't Ayer simply conclude from the fact that moral judgements are neither analytic nor empirically verifiable that they are *verbiage*? Ayer realises that this might be taken to be an implication of the emotive theory, and in the Introduction to the second edition he writes:

> In putting forward the principle of verification as a criterion of meaning, I do not overlook the fact that the word "meaning" is commonly used in a variety of senses, and I do not wish to deny that in some of these senses a statement may properly be said to be meaningful even though it is neither analytic nor empirically verifiable.[54]

The idea seems to be this. Although moral judgements are not literally significant, they are not nonsensical, because they possess some other sort of significance: *emotive significance*. Two questions now suggest themselves: (1) by what criterion does Ayer distinguish between putative judgements that possess emotive significance and those which are nonsensical (and which therefore ought to be eliminated)?; (2) can that criterion be stated in such a way that it grants emotive significance to ethical judgements, but refuses it to the putative judgements of the metaphysicians?

We can answer the first question by reflecting further on the generalised version of the verification principle that we mentioned briefly in §3.2. Recall that we could use the verification principle to demarcate not only the class of sentences that were literally significant, but also the classes of sentences which possessed imperatival significance or conative significance. It is not difficult to generalise the verification principle further so that it can be used to distinguish those sentences which possess genuine emotive significance from those which only appear to do so: *a sentence which appears to evince feeling will only be counted as possessing genuine emotive significance if there is some observable behaviour the occurrence of which constitutes the expression of the emotion*.

The problem with this is that when we attend to the sort of feeling which Ayer takes ethical judgements to express, it seems to follow that ethical judgements violate the generalised version of the verification principle. Ayer quite clearly takes the utterance of an ethical sentence to express a special sort of distinctively *ethical* feeling: note the reference to moral disapproval and special exclamation marks in the quote from *Language, Truth, and Logic* (p.107), given above. Also,

> a *special* sort of moral disapproval is the feeling which is being expressed.[55]

> Ethical symbols . . . occur in sentences which simply express *ethical* feeling about a certain type of action or situation.[56]

But is there observable behaviour the occurrence of which would constitute the expression of this special sort of moral or ethical emotion? It is difficult see how Ayer could answer this in the affirmative: we can perhaps imagine patterns of observable behaviour which would express disapproval,[57] but what observable behaviour could possibly manifest the presence of a distinctively *moral* sort of disapproval?[58]

This suggests that the second of our two questions above has to be answered in the negative: ethical sentences then get relegated to the category of nonsense along with the sentences of metaphysics. If this is so, then it is bad news for logical positivism: it shows that *even by its own lights*, logical positivism cannot find room for

the idea that ethical sentences possess *any* sort of significance. It is one thing to suggest that ethical sentences do not have the semantic function which their surface form suggests, another thing altogether to suggest that they have no semantic function whatsoever. Ayer, then, is perhaps closer to having to embrace moral nihilism than he was willing to admit.

3.7 Moderate holism

What can we say about the process of verification of sentences which do possess genuine factual significance, those which are in fact genuine empirical hypotheses? Ayer is a moderate *holist* with respect to such sentences: he thinks that empirical hypotheses are never *individually* subject to the test of experience, but rather face experience *en bloc*. As he puts it:

> When one speaks of hypotheses being verified in experience, it is important to bear in mind that it is never just a single hypothesis which an observation confirms or discredits, but always a system of hypotheses.[59]

Ayer uses the example of an experiment devised to test a particular scientific generalisation. Suppose that the generalisation states that "In conditions C, all A's are B's". Suppose also that we observe, in what appear to be circumstances C, an A which is not B. Then, Ayer argues, we are not compelled to give up on our scientific generalisation: we could also give up the claim that our observation of an A which is not B is veridical; we could give up the claim that the circumstances were actually of type C; and so on. Each of these options will require compensating adjustments in our overall theory, but "so long as we take suitable steps to keep our system of hypotheses free from contradiction, we may adopt any explanation we choose".[60] In practice, we will choose the explanation which best helps us to anticipate the course of our sensations, and though "these considerations have the effect of limiting our freedom in the matter of preserving and rejecting hypotheses", it is still the case that "logically our freedom is unlimited".[61] Ayer sums the point up thus:

> The "facts of experience" can never compel us to abandon a hypothesis. A man can always sustain his convictions in the face of apparently hostile evidence if he is prepared to make the necessary *ad hoc* assumptions. But though any particular instance in which a cherished hypothesis appears to be refuted can always be explained away, there must still remain the possibility that the hypothesis will ultimately be abandoned. Otherwise it is not a genuine hypothesis.[62]

But Ayer's holism is only moderate: it applies only to the subset of literally significant sentences which qualify as such in virtue of possessing factual significance. The other sort of literally significant sentence, the analytic sentences, are exempt from the holistic picture:

> A proposition whose validity we are resolved to maintain in the face of any experience is not a hypothesis at all, but a definition. In other words, it is not a synthetic but an analytic proposition.[63]

In the next chapter we will look at Quine's attempt to generalise Ayer's moderate holism to include even those sentences that have traditionally been regarded as analytic. As we shall see, this thoroughgoing holism has potentially drastic consequences, not only for the logical positivist conception of meaning, but for the notions of meaning and analyticity in general.

Chapter summary

We outlined the main tenets of logical positivism, a philosophical outlook – partly inspired by Wittgenstein's early work – which provides the essential link between Frege and Russell and the philosophy of language in the twentieth and twenty-first centuries. According to logical positivism, a statement is literally significant (capable of being true or false) if and only if it is either (a) an *a priori* tautology or (b) *a posteriori* and in principle capable of empirical verification. Since metaphysics has traditionally aimed to produce statements which are neither empty tautologies nor in principle capable of empirical verification, metaphysics, as

traditionally practised, would be undermined. Moreover, logical positivism puts forward a linguistic theory of necessity: according to this, the *a priori* statements of mathematics, for instance, are necessarily true in virtue of the conventions governing mathematical language. We saw that the attempt to formulate a plausible account of what it is for a statement to be in principle empirically verifiable resulted in failure, and that Quine raised a serious problem for the logical positivist account of the *a priori*. Inter alia, we looked briefly at how logical positivism proposed to deal with ethical judgement and at its commitment to a moderate form of holism about empirical confirmation. We also introduced Carnap's distinction between internal and external questions.

Study questions

1. Give examples of: a statement which is practically verifiable; a statement which is in principle verifiable but not practically verifiable; a statement which is strongly verifiable; a statement which is weakly verifiable but not strongly verifiable.
2. Give a counterexample of your own to the characterisation of factual propositions which Ayer gives in the first edition of *Language, Truth, and Logic*.
3. Let S be (– O1 & O2) v (– N & O3) where O1, O2, and O3 are observation statements that are logically independent of each other and N is some intuitively meaningless statement of your own choosing. (i) Show that S & O1 entails O3. (ii) Explain why this establishes that S is directly verifiable. (iii) Show that S & N entails O2. (iv) Show that O2 does not follow from S on its own. (v) Explain why the above steps establish that N is indirectly verifiable. (vi) Why is this a problem for Ayer?
4. Does Ayer provide any good reasons why we should reject John Stuart Mill's account of mathematical statements?
5. If mathematical statements merely illustrate the way in which we use certain symbols, how could we ever make any interesting discoveries in mathematics, and how could such "discoveries" have any utility?
6. What difficulty is Quine referring to when he writes "the difficulty is that if logic is to proceed mediately from the conventions, logic is needed for inferring logic from the conventions"?

7. Explain Carnap's distinction between logical frameworks and factual frameworks?
8. Do Carnap's linguistic frameworks consist of analytic truths? If not, how much of a problem is this for him?

Further reading

The best place to begin is Ayer's *Language, Truth, and Logic*: this is a concise and brilliantly written exposition of logical positivism. Ayer's editorial introduction to *Logical Positivism* provides a useful summary of the main themes of logical positivism, and a useful short history of the movement. This volume contains many of the classic papers of logical positivism: especially good are Carnap, "The Elimination of Metaphysics through the Logical Analysis of Language" and "The Old and the New Logic"; Schlick, "Positivism and Realism" and "The Turning Point in Philosophy"; Hahn, "Logic, Mathematics, and Knowledge of Nature". For more detailed surveys, see Uebel, "Vienna Circle" and Creath, "Logical Empiricism", both in the *Stanford Encyclopedia of Philosophy*.

Wittgenstein's *Tractatus Logico-Philosophicus* is an obscure and difficult work. For an introduction, see Beaney, "Wittgenstein on Language: From Simples to Samples"; Soames, *Philosophical Analysis in the Twentieth Century*, Chapters 9–11; Kenny, *Wittgenstein*, Chapters 1–6. For more detailed introductions see Pears, *The False Prison*, Volume 1; Schroeder, *Wittgenstein: The Way out of the Fly Bottle*, Chapter 2; White, *Wittgenstein's Tractatus Logico-Philosophicus*; and Morris, *Routledge Philosophy Guidebook to Wittgenstein and the Tractatus*. Recent literature on the *Tractatus* has been dominated by debate about the import of the likes of 6.54 where Wittgenstein refers to philosophical propositions as "nonsensical". See the papers in Crary and Read (eds), *The New Wittgenstein*; and Read and Lavery (eds) *Beyond the Tractatus Wars: The New Wittgenstein Debate*.

For discussion of the emotivist theory of moral judgement, see Miller, *Contemporary Metaethics: An Introduction*, Chapter 3.

A thorough summary of the various objections to the verification principle can be found in §7 of Wright, "Scientific Realism,

Observation, and the Verification Principle", in Wright and McDonald (eds), *Fact, Science, and Morality*. In this paper Wright attempts to formulate a version of the verification principle immune to these objections. Wright's own proposal is discussed in Lewis, "Statements Partly about Observation"; Wright, "The Verification Principle: Another Puncture – Another Patch". A useful full-length study of Ayer's work is Foster, *Ayer*. Quine's main objections to conventionalism can be found in Quine, "Carnap on Logical Truth" and "Truth By Convention". For a more recent discussion of convention, see Lewis, *Convention*. In addition to the collection edited by Wright and McDonald, mentioned above, the following also contain many articles of interest: Hahn (ed.), *The Philosophy of A. J. Ayer*; Philips-Griffiths (ed.), *A. J. Ayer: Memorial Essays*; Gower (ed.) *Logical Positivism in Perspective*. Finally, the two volumes of Ayer's autobiography, *Part of My Life* and *More of My Life*, provide well-written and entertaining background reading, as does Rogers's biography, *A. J. Ayer*.

Notes

1 Wittgenstein, *Tractatus Logico-Philosophicus*, p.3.
2 The *Tractatus* consists of seven main propositions, numbered accordingly. The other propositions consist of comments on these, plus comments on comments, and so on. Thus, 1.1 is a comment on 1, 1.11 is a comment on 1.1, 1.12 is another comment on 1.1, and so on.
3 Matters are actually more complicated here, as Wittgenstein wants to deny that analytic *a priori* propositions, such as the propositions of logic or mathematics, have meaning. See e.g. 6.111 and 6.21. I gloss over this here, and agree with Soames's comment that "For Wittgenstein, tautologies and contradictions don't state anything, or give any information about the world. However, their truth or falsity can be calculated, and understanding them reveals something about our symbol system. Thus, they can be regarded as meaningful in an extended sense" (Soames, *Philosophical Analysis*, pp.234–235). At any rate, I'm here only giving a hint of how Wittgenstein's early work might have influenced the logical positivists, not serious exegesis of that work itself.
4 Ayer, *Logical Positivism*, p.11.

5 Hahn, "Logic, Mathematics", p.159.
6 For ease of exposition, I'll sometimes just speak of statements, rather than sentences expressing thoughts. So, for example, rather than speaking of "sentences which are literally meaningful in virtue of expressing empirically verifiable thoughts", I'll speak of "statements which are literally meaningful in virtue of being empirically verifiable".
7 Ayer, *Language, Truth, and Logic*, p.35. A "pseudo-proposition" is something which appears to express a genuine thought, but in fact expresses no thought.
8 Ibid.
9 Someone might respond that these generalisations of the verification principle to commands, expressions of intention, expressions of desire, and so on are too weak, since they seem to be satisfied by commands and expressions of intention concerning the absolute. For example, what about "Utter a sentence about the absolute!", "I, a metaphysician, intend to write a treatise on the Absolute"? If I say "The absolute is expanding at an ever increasing rate", or write a treatise called *The Nature of the Absolute*, haven't I observably obeyed the command, or implemented the intention, and doesn't this show that these sentences actually satisfy the generalised version of the verification principle? The answer is that it is does not: the logical positivist can respond that you have not succeeded in obeying the command, or implementing the intention. Rather, you have only succeeded in uttering a sentence, or writing a treatise, *containing the word* "Absolute".
10 This generalised version of the verification principle arguably features in many philosophical arguments from the second half of the twentieth century: Wittgensteinian arguments against the possibility of private language, and anti-realist arguments, in the style of Dummett, to mention but two. Perhaps this helps explain Ayer's rather caustic remark in his autobiography that "the verification principle is seldom mentioned and when it is mentioned it is usually scorned; it continues, however, to be put to work. The attitude of many philosophers towards it reminds me of the relation between Pip and Magwitch in Dickens's *Great Expectations*. They have lived on the money, but are ashamed to acknowledge its source" (Ayers, *Part of My Life*, p.156).
11 Ayer, *Language, Truth, and Logic*, pp.38–39.
12 Ibid., p.39.
13 Of course, for this implication, we need the assumption that mere facts about meaning cannot provide an adequate basis for answering the question one way rather than another. This is the analogue

of the assumption, in the immediately preceding case, that the statement in question is not analytically true.

14 Ayer, *Language, Truth, and Logic*, p.13.
15 Ibid.
16 See Church, "Review of *Language, Truth, and Logic*".
17 The material in this section is more demanding than that of the rest of the chapter. Readers completely new to the philosophy of language may wish to skip the section, move directly to §3.4, and return to the present chapter after reading Chapters 4 and 5.
18 Foster, *Ayer*, p.16.
19 Ibid., pp.17, 22.
20 Ibid., p.22.
21 Ibid., p.27.
22 Ibid., pp.29–30.
23 Ibid., pp.28–29.
24 Ibid., p.29.
25 Ibid.
26 For more on the distinction between constitutive and epistemological scepticism, see Chapters 4 and 5.
27 Ayer, *Language, Truth, and Logic*, p.78.
28 Ibid., p.16.
29 Ibid., p.72, emphasis added.
30 Hahn, "Logic, Mathematics", p.148.
31 Ayer, *Language, Truth, and Logic*, p.75.
32 Ibid., pp.75–76.
33 Ibid., p.82.
34 Ibid., p.84.
35 Ibid.
36 Ibid., p.85.
37 Hahn, "Logic, Mathematics", p.152.
38 Ibid., p.153.
39 Ayer, Language, Truth, and Logic, p.86.
40 Hahn, "Logic, Mathematics", p.157.
41 Ayer, *Language, Truth, and Logic*, p.80.
42 Russell, "Logical Positivism", p.367.
43 Carnap, "Old and the New Logic", p.145.
44 Quine, "Truth By Convention", p.351.
45 Ibid., p.352.
46 Note that insofar as Quine's objection is levelled specifically against conventionalism about logic, it may leave scope for a conventionalist account of mathematics. We will have to use logic in deriving individual mathematical truths from the relevant general conventions, but this will not be circular as in the logical case.

47 Carnap, "Empiricism, Semantics, and Ontology", p.242.
48 Ibid., p.245.
49 Ibid., pp.250, 256.
50 Hookway, *Quine*, p.33.
51 Hookway writes as if he is unaware of Carnap's distinction between logical and factual frameworks, and as though once a framework has been chosen, all of the internal questions that can be raised within it are synthetic. Introducing the notion of a linguistic framework, he writes: "Each embodied a system of logical principles, and the terms employed in each were given meaning by analytic principles or 'meaning postulates' *linking claims employing them, directly or indirectly, to observational claims*" (Hookway, *Quine*, p.31, emphasis added). This suggests that he takes all frameworks to be factual: all internal questions are factual, and the analytic truths are themselves part of the framework. Again, this is a misinterpretation of Carnap: according to Carnap, the conventions composing the frameworks are neither analytic nor synthetic, the distinction between analytic and synthetic truths being drawn *within* the framework concerned. Of course, this highlights a difficulty for Carnap: if the framework principles are neither true nor false, how can anything *follow* from them?
52 Ayer, *Language, Truth, and Logic*, p.107, emphasis added.
53 Ibid.
54 Ibid., p.15.
55 Ibid., p.107, emphasis added.
56 Ibid., p.108, emphasis added.
57 Assuming, of course, that Ayer can provide a convincing response on behalf of logical positivism to the general problem of other minds. This is no easy task, obviously, but we can grant Ayer his solution to the other minds problem at this point without jeopardising the objection developed in the text.
58 Could Ayer respond here by claiming that ethical sentences serve to express natural (i.e. not specifically moral) forms of emotion? All Ayer would require in order to save the claim that ethical sentences possess genuine emotive significance would be a solution to the problem of other minds, which he is going to have to solve anyway. Assessing this suggestion would take us too far afield at this point, but it seems unlikely that it would be acceptable to Ayer. In Chapter 6 of *Language, Truth, and Logic*, Ayer is content to use Moore's "open question" argument against theories which attempt to define ethical symbols in naturalistic terms. It is plausible that

the arguments Ayer uses against naturalistic cognitivism would carry over to apply to non-cognitivist theories which attempt to view ethical sentences as expressing only natural (i.e. non-moral) sentiments. For a development of this line of thought, and a more extended exposition of the argument of this section, see Miller, "Emotivism and the Verification Principle".

59 Ayer, *Language, Truth, and Logic*, p.94.
60 Ibid., p.95.
61 Ibid.
62 Ibid.
63 Ibid.

Chapter 4

Scepticism about sense (I)

Quine on analyticity and translation

In the previous chapters we have concentrated on the various ways in which Frege, Russell, and the logical positivists attempted to systematise our intuitive notion of meaning. In this chapter we move on to look at what appears to be a much more negative outlook on meaning: *scepticism*. We will look at three main lines of philosophical scepticism about the notion of meaning, two of which are associated with the influential American philosopher W.V.O. Quine (1908–2000). In §§4.1–4.5 we discuss Quine's attack on the analytic/synthetic distinction; in §§4.6–4.10 we discuss Quine's thesis of the indeterminacy of translation; and in Chapter 5, we discuss Kripke's Wittgenstein's attack on the idea that the expressions of our language possess such a thing as a determinate meaning. These discussions will set the scene for Chapter 6, where we look at a number of attempts to respond to these sceptical challenges.

4.1 Quine's attack on the analytic/synthetic distinction: Introduction

The distinction between analytic and synthetic truth loomed large in Chapter 3. Recall that Ayer defined an analytically true sentence as one which was true purely in virtue of the senses of its constituent expressions, and he attempted to use this definition of analyticity to develop an account of *a priori* truth and the linguistic theory of necessity. Synthetic truths are those whose truth does not consist solely in facts about sense, and on the logical positivist picture these are factually significant, in virtue of being in principle susceptible to empirical verification. Likewise, the distinction between the analytic and the synthetic played a large part in Carnap's distinction between logical frameworks and factual frameworks: a logical framework is a linguistic framework which is such that all of the internal questions formulable within it are analytic; a factual framework is a framework such that at least some of the internal questions formulable within it are synthetic. In this section we introduce Quine's attack on the analytic/synthetic distinction in his 1951 paper "Two Dogmas of Empiricism", arguably the most influential paper in twentieth-century philosophy of language. Quine's main target in this attack was the logical positivist theory of meaning and Carnap's distinction between external and internal questions. But as we will see, Quine's attack extends further than logical positivism, and threatens our intuitive notion of meaning itself.

4.2 The argument of "Two Dogmas" (part I)

Quine's article falls into two main parts. In the first four sections he claims that no philosopher has ever succeeded in giving an acceptable explanation of the notion of analyticity. In the final two sections he explicitly concerns himself with the logical positivist theory of meaning, the verification theory. He argues that to the extent that the logical positivist conception of analyticity can be

made sense of, the concept of analyticity simply fails to have any application: there are no analytic truths. We shall introduce the arguments of Quine's final two sections in §4.4. In this section we outline the shape of Quine's argument in the first four sections of "Two Dogmas" by looking at his rejection of two distinct attempts philosophers have made to explicate the notion of analyticity.

Quine first considers Kant's attempted explication of analyticity in the *Critique of Pure Reason* (1781). According to Kant "Analytic statements are those which attribute to the subject no more than is already conceptually contained in the subject concept".[1] There are two main problems with the Kantian characterisation. First of all, the notion of "conceptual containment" is no more than a *metaphor*, and stands in as much need of explication as the notion of analyticity itself. What exactly does it mean to say that the concept *bachelor* "contains" the concept *unmarried male of marriageable age*? Secondly, the Kantian characterisation limits itself to sentences of the subject-predicate form. The Kantian definition has no application to sentences which are not of this form. For example, consider "Mary drinks with those with whom she herself tipples", or "John eats those he himself devours". Intuitively, these are in some sense true purely in virtue of meaning: we do not need to observe the behaviour of Mary or John to know that they are true.

Kant's attempt to define an analytic truth "as one whose denial results in contradiction" fares no better. Quine writes:

> This definition has small explanatory value; for the notion of self-contradictoriness, in the quite broad sense needed for this definition of analyticity, stands in exactly the same need of clarification as does the notion of analyticity itself. The two notions are the sides of a single dubious coin.[2]

Take a putative analytic truth, such as "If Jones is a bachelor then he is unmarried". Kant's thought is that the analyticity of this statement consists in the fact that its denial results in a self-contradiction. But what is a self-contradiction? One way to characterise self-contradiction would be *syntactically*: a self-contradiction is simply any statement of the syntactic *form* P & –P. Call this the *narrow syntactic notion of self-contradiction*.

This notion may be well-understood, but it is of no help to us in explaining analyticity, since "Jones is a married bachelor" is not of the syntactic form P & –P. So we need a notion of self-contradiction *broader* than the narrow syntactic notion. But, Quine claims, outlining a broad notion of self-contradiction which includes "Jones is a married bachelor" is no easier a task than explaining the notion of analyticity itself.

Both of Kant's attempts at explaining analyticity are thus inadequate. Quine then moves on to consider a more modern account of analyticity, an account due in its essentials to Frege. Very roughly, analytic truths are characterised as those which can be demonstrated or proved using only logical laws and definitions as premises. Logical laws are those statements which are true, and which remain true, under all interpretations of their component parts, excluding the logical vocabulary. The logical vocabulary consists of words like "not, "and", "or", "If . . . then . . .", and so on. "No unmarried men are married" thus counts as a logical law, since it is true, and remains true under all reinterpretations of its non-logical vocabulary (e.g. "No unwashed Englishmen are washed", "No ungodly Scotsmen are godly", and so on). Given this notion of a logical law, we can say that *an analytic truth is any statement which is either (a) a logical law, or (b) derivable from logical laws using only definitions as premises*. Thus "No unmarried men are married" counts as analytic under (a). But what about "No bachelors are married", which is not a logical law (since e.g. "No Welshmen are married" is false)? This counts as analytic under the Fregean characterisation because it is derivable from a logical law using only definitions as premises. The derivation could be set out informally as follows:

(1) No unmarried men are married (Logical law)
(2) Bachelors are unmarried men (Definition of "bachelor")
(3) No bachelors are married (From (1) and (2))

The fact that "No bachelors are married" is derivable in this way shows that it counts as analytic under (b).

Quine spends quite some time arguing that this Fregean characterisation of analyticity is not satisfactory as it stands.

The main problem is that analytic statements of type (b) have to be characterised as those which are derivable from logical truths by means of *correct* definitions. Without this proviso about correctness, Frege's account of analyticity would allow us to prove that, for example, "No philosophers are married" is an analytic truth! Witness

(1) No unmarried men are married (Logical law)
(2) Philosophers are unmarried men (Definition of "philosopher")
(3) No philosophers are married (From (1) and (2))

The problem here is clearly that the definition of "philosopher" is incorrect. But what does it mean to say that a definition is correct or incorrect? Quine's strategy at this point is to argue that answering this question presupposes an answer to our original and as yet unanswered question about analyticity.

One way of answering the question about definition is as follows: a definition is correct if and only if the defined and defining terms are *synonymous*. But now we must give an account of synonymy. What does it mean to say that two expressions are synonymous? Quine considers the suggestion that two expressions "F" and "G" are synonymous whenever they are *intersubstitutable salva veritate*. To say that "F" are "G" are intersubstitutable *salva veritate* is just to say that whenever we have a sentence which contains "F", we can always substitute "G" for "F" without changing the truth-value of the overall sentence (and vice versa, whenever we have a sentence which contains "G", we can always substitute "F" for "G" without changing the truth-value of the overall sentence).

How plausible is this attempt at explicating synonymy? If it works, we can use it to explain the notion of a correct definition, and we can then use that notion to characterise analyticity along Fregean lines. But, according to Quine, this attempt at explicating synonymy is hopeless. Whether the account of synonymy in terms of intersubstitutability is plausible will depend on the richness of the language within which we are working. So suppose, firstly, that our language is an *extensional* language: a language whose sentences are formalisable using only the standard vocabulary of first-order predicate calculus with identity. Then it is easy to see

that intersubsitutability *salva veritate* is not a sufficient condition for synonymy. "Renate" and "Cordate" are co-extensional, and thus intersubstitutable *salva veritate* in such a language, but they are plainly not synonymous, since the first means "Creature with kidneys" while the second means "Creature with a heart".[3]

The problem here is that it is a mere *accident of nature* that "Renate" and "Cordate" are co-extensional and hence intersubstitutable *salva veritate*. In order to rule out this sort of case we need to have the resources within our language to express the thought that two predicates, for example, are more than merely accidentally co-extensional. In other words, we will need to work within a language which includes something like the modal adverb "necessarily", and which is thus richer than standard first-order predicate calculus with identity. This will allow us to say that two expressions are synonymous if and only if they are *necessarily* intersubstitutable *salva veritate*. This would ensure that cases like "Renate" and "Cordate" do not count as synonymous by our account of synonymy: although they are co-extensional, they are not necessarily co-extensional, since we can quite easily imagine possible worlds in which there are creatures with kidneys but no heart, and so on. Thus, "Renate" and "Cordate" are not necessarily intersubstitutable *salva veritate*, and hence do not count as synonymous by our account of synonymy.

But now we face the new problem of making sense of "necessarily". What does it mean to say that "Necessarily P" is true? Quine says that this can mean nothing more than that P is analytic:

> But can we condone a language which contains ["necessarily"]? Does the adverb really make sense? To suppose that it does is to suppose that we have already made satisfactory sense of "analytic". Then what are we so hard at work on right now?[4]

So we are back where we started. In order to explain analyticity along Fregean lines we had to explain the notion of a correct definition, in order to explain the notion of a correct definition we had to explain the notion of synonymy, and in order to explain synonymy we had to explain the meaning of "necessarily", which in turn requires an explanation of analyticity!

The Kantian and Fregean attempts at explaining analyticity thus fail, since they presuppose the notion of analyticity or notions just as unclear as analyticity at some point in their attempted explanations. In the remainder of the first four sections of "Two Dogmas", Quine looks at several other attempts at explaining analyticity, and makes essentially the same objection. He concludes that to the extent that no intelligible characterisation of the distinction has been forthcoming, we have evidence that there is simply no coherent distinction between the analytic and the synthetic.

How plausible is the attack against analyticity which Quine mounts in the first four sections? Before answering this question, we must pause to make three comments about the nature of Quine's attack.

Comment 1. Note that although Quine takes the Kantian and Fregean attempts at explaining analyticity as his stalking horses, his real target is the logical positivist characterisation of analytic truth as "truth purely in virtue of meaning". We can view the Kantian and Fregean explanations as attempts to explain what it is for a statement to be true purely in virtue of meaning. To the extent that they fail, we have not been given an explanation of the logical positivists' characterisation of analyticity.

Comment 2. We can perhaps make an *ad hominem* point against Quine here. Recall Quine's comment that "a logical truth is a statement which is true and which remains true under all reinterpretations of its components other than the logical particles". Now Quine is quite happy to countenance a class of logical truths: his concern is to question whether there is any coherent way in which that notion can be used to characterise a class of analytic truths. But Quine doesn't seem to realise that his arguments against analyticity militate against his own characterisation of logical truth. Consider "$(\forall x)((Fx \ \& \ Gx) \rightarrow Gx)$". Clearly, if anything counts as a logical truth, this does. But "All green banks are banks" is an instance of this, and it is false if the token occurrences of "bank" have different meanings (e.g. if the first token refers to a river bank and the second token refers to a financial institution). Quine thus needs to include in his definition of logical truth some stipulation to the effect that the substituted expressions

are *synonymous*. For example, in the case above, the stipulation requires that the tokens of "bank" that are substituted for the tokens of "F" are synonymous. This shows that if Quine's attack on the notion of synonymy is successful, his own definition of logical truth will be jeopardised.

Comment 3. Quine's attack on the notion of analyticity is, inter alia, an attack on the idea that sentences have such a thing as meaning. This is brought out very clearly in a passage from "In Defense of a Dogma", an early reply to "Two Dogmas" by Paul Grice and Peter Strawson:

> If talk of sentence-synonymy is meaningless, then it seems that talk of sentences having a meaning at all must be meaningless too. For if it made sense to talk of a sentence having a meaning, or meaning something, then presumably it would make sense to ask "What does it mean?". And if it made sense to ask "What does it mean?" of a sentence, then sentence-synonymy could be roughly defined as follows: Two sentences are synonymous if and only if any true answer to the question "What does it mean?" asked of one of them, is a true answer to the same question, asked of the other . . . If we are to give up the notion of sentence-synonymy as senseless, we must give up the notion of sentence-significance (of a sentence having meaning) as senseless too. But then perhaps we might as well give up the notion of sense.[5]

This shows that Quine's attack is far more radical than might initially appear. It is one thing to give up a philosophical distinction, another thing entirely to give up the intuitive notion of sentence-meaning. Quine thus appears to be arguing for a fully fledged version of meaning scepticism: there is no fact of the matter as to what any given sentence or linguistic expression means.

4.3 Criticism of "Two Dogmas" (part I)

The main objection which Quine raises against the Kantian and Fregean attempts to explain analyticity is that they are *circular*: at some point in the explanation they need to presuppose the

notion of analyticity itself, or some other cognate notion – such as correct definition, synonymy, or necessity – which is just as problematic. This objection appears to rest on the assumption that a putative concept is not fully intelligible unless it admits of an explicit, non-circular definition. An explicit definition of a concept P, say, is a statement of a set of conditions necessary and sufficient for the instantiation of P:

> x is P if and only if Fx.

An explicit definition is non-circular when the concept P (or any concept whose definition itself presupposes P) does not appear on the right-hand side, in the specification of F.[6] That this assumption underlies Quine's argument here is clear from the fact that almost all of the attempts at developing, for example, the Fregean definition of analyticity are ruled out because analyticity or one of its cognates appears, or is presupposed, on the right-hand side of the proffered definition. For example, where s is a sentence, the definition:

> s is analytic if and only if s is derivable from logical laws and definitions

is ruled out because the explanation of definition on the right-hand side makes sense only if we eventually presuppose the notion of analyticity itself.

Call this assumption *Quine's Socratic Assumption*.[7] Is it justified? On the face of it, it appears to impose a condition on a concept's being intelligible that is much too strong. Aren't there many concepts which we perfectly well understand but which fail to satisfy this condition? Take colour concepts, for example. We can perfectly well understand sentences containing colour concepts, but no-one

has ever succeeded in giving a non-circular set of necessary and sufficient conditions for, for example, the correct application of "red" or "blue" or "green".

Another consideration which suggests that Quine's Socratic Assumption is unjustified is that in order for a language to contain resources sufficient for an explicit non-circular definition of each of its expressions, it would need to contain infinitely many semantic primitives. Given this, a semantic theory (see §1.7) for the language in question would need to contain infinitely many axioms. But as we shall see when we come to discuss Davidson in Chapter 8, a language is learnable only if a finitely axiomatisable semantic theory for that language is possible. So a language all of whose expressions satisfy Quine's Socratic Assumption would turn out to be in principle unlearnable.

These considerations suggest that Quine's Socratic Assumption can at best constitute only a *local* requirement, a requirement imposed on *some but not all* expressions of the language concerned. Given the local nature of the requirement, then, the onus is on Quine to explain why "analytic" is one of the expressions which have to meet it. What sort of explanation could Quine give here? It would have to rest on some claim to the effect that there were *independent* reasons for thinking "analytic" to be problematic, perhaps that in practice we always find ourselves hesitating over its application. He comes close to offering such a reason in the following passage:

> I do not know whether the statement "Everything is green is extended" is analytic. Now does my indecision over this example really betray an incomplete understanding, an incomplete grasp of the "meanings", of "green" and "extended"? I think not. The trouble is not with "green" or "extended", but with "analytic".[8]

Granted, we do feel some hesitancy here. But in order for Quine to meet our challenge he would have to show not just that we occasionally hesitate over the application of the concepts, but that such hesitancy was *widespread*. Grice and Strawson argue convincingly that this is not the case, since (a) mastery of the concepts *analytic* and *synthetic* can quite easily be imparted to novices, and

(b) those with even a very modest philosophical training tend, non-collusively, to agree over the application of the concepts to new cases (think about your own reaction to the examples involving Mary and John in §4.2).[9] And even in the isolated case in which there does appear to be some hesitancy, the hesitancy is not conclusively traceable to the presence of "analytic":

> If, as Quine says, the trouble is with "analytic", then the trouble should doubtless disappear when "analytic" is removed. So let us remove it, and replace it with a word which Quine himself has contrasted favourably with "analytic" in respect of perspicuity – the word "true". Does the indecision at once disappear? We think not. The indecision over "analytic" (and equally, in this case, the indecision over "true") arises, of course, from a further indecision: viz, that which we should feel when confronted with such questions as "Should we count a point of green light as extended or not?" As is frequent enough in such cases, the hesitation arises from the fact that the boundaries of application of words are not determined by usage in all possible directions. But the example Quine has chosen is particularly unfortunate for his thesis, in that it is only too evident that our hesitations are not here attributable to obscurities in "analytic".[10]

Our conclusion is thus that Quine's Socratic requirement on the legitimacy of concepts appears to be unreasonable and unmotivated. The fact that "analytic" does not meet that requirement does not establish that it is unintelligible, no more than the failure to meet it of "red", "green", and so on, establishes the unintelligibility of our colour concepts. Quine's general line of argument in the first four sections of "Two Dogmas" fails to show that analyticity and its cognate concepts are unintelligible.[11]

4.4 The argument of "Two Dogmas" (part II)

In the final two sections of "Two Dogmas", Quine invites us to consider whether the verificationist theory of meaning propounded by the logical positivists can provide us with a satisfactory elucidation

of analyticity. Recall that the verificationist theory of meaning was the doctrine that the meaning of a statement consists in its method of empirical confirmation or disconfirmation. This seems to furnish us with an appropriate definition of synonymy: two sentences are synonymous if they have the same method of empirical verification. Given this definition of synonymy, we could go on to frame a definition of analyticity: analytic statements are those which are synonymous with logical truths, or, in other words, analytic statements are those that would be confirmed by experience *come what may*.

We can put this thought another way. The truth of a sentence is a function of two things: of its meaning, on the one hand, and of the way the world is, on the other. For example, "grass is green" is true, and this is a function of the fact that it means that grass is green, and that grass is, in fact, green. If the sentence had meant that Newcastle is in Scotland it would have been false. So we can view sentences as having both a factual component and a linguistic component. Analytic sentences can now be characterised as those whose factual components are vacuous or empty.

Quine attacks the story that we have just outlined: he denies that the verificationist theory of meaning can provide us with a plausible characterisation of analyticity. It cannot do this because it rests upon a faulty assumption: the assumption that individual statements, considered in isolation, can admit of confirmation or disconfirmation. This assumption is shown to be faulty by Quine's doctrine of *epistemological holism*, according to which all significant statements face the tribunal of experience not *individually*, but *en masse*: "our statements about the external world face the tribunal of sense experience not individually, but only as a corporate body".[12] When faced with a *recalcitrant* experience – an experience which conflicts with our currently held theory of the world – we have to revise that theory. But in principle any part of the theory can be dropped, subject to this being accommodated by changes in other areas of the theory: which part of the theory should actually be dropped will be determined by purely pragmatic standards, such as the preservation of simplicity and the minimisation of future recalcitrance. Unlike the holism espoused by Ayer (see §3.7), Quine's holism is not modest but *thoroughgoing*:

even the statements that would be held by the positivists to be analytic, such as the statements of logic and mathematics, are, in principle, rationally revisable:

> Total science is like a field of force whose boundary conditions are experience. A conflict with experience at the periphery occasions readjustments in the interior of the field. Truth values have to be redistributed over some of our statements. Reevaluation of some statements entails reevaluation of others, because of their logical interconnections – the logical laws being in turn simply certain further statements of the system, certain further elements of the field. Having reevaluated one statement we must reevaluate some others, which may be statements logically connected with the first or may be the statements of logical connections themselves. But the total field is so underdetermined by its boundary conditions, experience, that there is much latitude of choice as to what statements to reevaluate in the light of any single contrary experience.[13]

The holistic picture implies that there are no statements of theory which are in principle immune to rational revision. It follows from this that the notion of analyticity, as characterised by the verificationist theory of meaning, simply fails to have any application. On the verificationist picture, analytic truths are those statements that are confirmed come what may (that are immune to rational revision in the light of experience) because their factual components are vacuous; synthetic statements are those which hold contingently on experience since their factual components are non-empty. Quine's holism appears to destroy this distinction:

> It is misleading to think of the empirical content of an individual statement – especially if it is a statement at all remote from the experiential periphery of the field. Furthermore it becomes folly to seek a boundary between synthetic statements, which hold contingently on experience, and analytic statements, which hold come what may. Any statement can be held true come what may, if we make drastic enough adjustments elsewhere in the system. Even a statement very close to the

> periphery can be held true in the face of recalcitrant experience by pleading hallucination or by amending certain statements of the kind called logical laws. Conversely, by the same token, no statement is immune to revision. Revision even of the law of excluded middle has been proposed as a means of simplifying quantum mechanics; and what difference is there in principle between such a shift and the shift whereby Kepler superseded Ptolemy, or Einstein Newton, or Darwin Aristotle?[14]

In other words, the distinction between the linguistic and factual components of an individual sentence, upon which the verificationist account of the analytic/synthetic distinction depends, is blurred beyond recognition by holism.[15] A statement is synthetic if it has its own verification conditions and these are not the same as those of a logical truth. A statement is analytic if it has its own verification conditions and these are the same as those of a logical truth. But holism implies that individual statements don't have their own verification conditions (as opposed to whole theories consisting of individual statements). So no statements are either analytic or synthetic.[16] Again, Quine's conclusion is that we have not been given an intelligible distinction between the analytic and the synthetic.

4.5 Criticism of "Two Dogmas" (part II)

Crispin Wright has argued that Quine's holism is implausible, because it is actually *incoherent*: he claims that Quine's holism cannot provide us with a coherent account of scientific methodology. In this section we attempt to reconstruct this objection.

Wright asks us to consider the following example. Suppose we have a theory of the world T together with an underlying logic L. Suppose that, together, they generate the conditional $I \rightarrow P$, "whose antecedent describes certain initial conditions and whose consequent formulates a prediction relative to those conditions".[17] Suppose also that we suffer a recalcitrant experience E, which inclines us to believe, contra T and L, both I and –P. Quine's holism implies that the following are all in principle candidates for revision:

$$\begin{array}{c} I \,\&\, -P \\ T \\ L \\ T \vdash_L I \rightarrow P \end{array}$$

Given that all of these are in principle revisable, which of them ought we to revise in practice? For Quine, this question has to be answered on pragmatic grounds, and on pragmatic grounds alone. So the question we are facing is: how do pragmatic considerations determine whether or not, for example, the rejection of $T \vdash_L I \rightarrow P$ is "a good move"? According to Quine, we answer this by considering whether the rejection of $T \vdash_L I \rightarrow P$ best minimises future recalcitrance (a "good move" is one which best minimises the possibility of future recalcitrance). If it does, we reject it, if it doesn't we keep it, and reject some other part of our overall body of beliefs. However, and this is the crucial point in Wright's objection, *verdicts about recalcitrance are in the same boat as all our other judgements*. That is to say, judgements about recalcitrance are themselves in principle up for revision.

Now suppose that pragmatic considerations do tell us that:

(1) The rejection of $T \vdash_L I \rightarrow P$ best minimises future recalcitrance.

Then holism, together with the fact that judgements about recalcitrance are in the same boat as all the rest, forces us to ask: do pragmatic considerations determine whether the rejection or acceptance of (1) *itself* is a good move? Suppose we get the answer:

(2) The rejection of (1) is a good move (i.e. best minimises the possibility of future recalcitrance).

Once again, holism, together with the fact that judgements about recalcitrance are in the same boat as all the rest, forces us to ask: do pragmatic considerations determine whether the rejection or acceptance of (2) is a good move? Suppose we get the answer:

(3) The rejection of (2) is a good move.

Then holism forces us to ask . . . And so on, *ad infinitum*. Because judgements about recalcitrance are in the same boat as all the rest, we get an *infinite regress* of such questions: and because of the regress, the Quinean methodology can never actually get off the ground. Wright's conclusion is that Quine has failed to provide a coherent account of the methodology we use in revising our theory of the world in the light of the deliverances of experience.[18]

According to Wright, what this shows is that in order to stop the regress, we have to exempt some judgements (e.g. $T \vdash_L I \rightarrow P$) from pragmatic appraisal. These judgements are the analytic judgements. So we are left with a distinction between the analytic and the synthetic after all.

Wright's objection to Quine's holism is subtle and difficult to appraise, and we cannot attempt an evaluation of it here. In the next section we move on to look at the second source of Quine's scepticism about the notion of meaning, his arguments for the indeterminacy of translation.

4.6 Quine on the indeterminacy of translation: Introduction

In the previous sections we looked at Quine's arguments against the analytic/synthetic distinction, and we saw that these arguments were, inter alia, arguments against the very notion of meaning. They were arguments for *meaning scepticism*: there is simply no fact of the matter as to whether a given sentence is analytic or synthetic, no fact of the matter as to what it means; and for any given pair of sentences, there is no fact of the matter as to whether they are synonymous or not. We now move on to look at a different set of arguments for meaning scepticism: Quine's arguments for the indeterminacy of translation.

Suppose that you are a speaker of a language L1, and that you are setting out to translate the sentences of some other language, L2, into your own language. Your aim in this task would be to construct a *translation manual* which correlated each sentence of the language L2 with a sentence of your own language L1. Intuitively,

we would say that there is an objective fact of the matter as to whether the translation manual that you come up with is *correct*. Thus, if L1 is English and L2 is German, a translation manual which correlates "Das ist ein Kaninchen" with "That is a rabbit" is correct, whereas a translation manual which correlates it with "That is an undetached rabbit part" is surely incorrect. Quine claims that this is an illusion: actually, there is no fact of the matter as to which of these two translation manuals is correct. In general, there is no fact of the matter as to whether a given translation manual for a given language is correct. When we are faced with a choice between two different translation manuals, there is no fact of the matter we can "get right" in our choice: some translation manuals may be more useful, more elegant, more natural, more simple, and so on, but these are purely pragmatic considerations.

This, in essence, is the indeterminacy of translation. Quine clearly thinks that the importance of this result lies in the consequences it has for the notion of meaning. The conclusion that it is not an objective factual matter whether a given translation manual for a language is correct or not is supposed to point to the conclusion that there is no objective fact of the matter as to what the sentences of that language mean. This connection between translation and meaning initially seems clear enough: a translation manual has the job of pairing the sentences of one language with the sentences of another, and the manual will be correct if it pairs the sentences of the first language with sentences of the second language which *have the same meaning*. Thus if there is no fact of the matter about the correctness of the translation manual, there will be no fact of the matter about sameness of meaning, and as we saw when discussing analyticity, this entails that there are no facts about meaning at all.

We can now state the doctrine of the indeterminacy of translation in more detail. Given a language L, it is possible, at least in principle, to construct two translation manuals, T1 and T2, both of which fit all of the possible evidence that is relevant to translation, but which are nevertheless incompatible with each other. In fact, Quine would claim that for any language L there will be an infinitely large number of translation manuals which equally fit all of the possible evidence relevant to translation, but which are

pairwise incompatible (that is, if you pick any two of these manuals, they will be incompatible with each other). Quine concludes that there is no fact of the matter as to which manual is correct: if one manual translates "Kaninchen" as "rabbit" and another translates "Kaninchen" as "undetached rabbit part", there may be nothing to choose between them, other than the sorts of purely pragmatic considerations mentioned above. This seems shocking: surely there is a fact of the matter as to which translation manual best captures the nature of the language under consideration? Quine claims not, so we must now examine his arguments for this intuitively shocking claim.

4.7 The argument from below

In fact, there are at least two general lines of argument for the thesis: the "argument from below", which comes from Chapter 2 of Quine's most important philosophical book, *Word and Object*; and the "argument from above", based on general considerations concerning the underdetermination of scientific theory by all possible observational evidence, which comes from Quine's short paper "On the Reasons for the Indeterminacy of Translation". In this and the following section we concentrate on the argument from below; the argument from above is discussed in §4.9.

A crucial figure in Quine's argument is the *radical translator*. This is a translator who is engaged upon the project of translating the language of the natives who belong to some hitherto completely unknown tribe. The only evidence which the radical translator has at his disposal is that which he can obtain from his *observation of the natives' behaviour*: he has no help in his task from dictionaries, or existing partial translation manuals which translate some native sentences into his own language, or from translations of the native language into some other language for which the translator already has a translation manual. The radical translator does not even have access to things like historical studies of the natives' language and culture – he has nothing at all to go on, except what he can observe in the behaviour of the speakers of the native language.

What is the motivation for looking at the process of translation in this very artificial setting? An assumption underlying Quine's entire argument here seems to be this: if there are facts which could determine whether a given translation manual is correct, that is, if there are facts about meaning, then those facts must be determined by facts about the behaviour of the speakers of the language under consideration. What else could constitute a fact about the meaning of an expression? Intuitively, the meaning of any given expression is something settled by *convention*: "grass" only means grass because of the existence of a certain convention among speakers of English. But conventions do not just come magically into being: they are brought into being by the *behaviour* of the various parties to the convention. So, the fact that "grass" means grass, and not snow, must in some sense be determined by our behaviour with the word. *Quine's strategy will be to argue that the facts about behaviour are not sufficient to determine whether a given translation manual is correct: the facts about behaviour will be consistent with the adoption of an altogether different translation manual*. So there is no fact of the matter as to which translation manual is correct.[19]

Since Quine's interest is in the relationship between the behavioural facts and the facts about meaning, using the device of the radical translator allows him to focus on the relationship between the putative facts about meaning and the behavioural facts in their purest form: the radical translator's angle on the facts about behaviour at his disposal is in no danger of being tainted by collateral information, which might already have some implicit semantic import. This is the motivation for looking at translation in this radical context. The task facing the radical translator is thus as follows: *given all of the observable facts about the behaviour of the natives and nothing more, construct a translation manual which yields a translation in your own language of each of the sentences of the natives' language*.

What counts as a fact about the natives' behaviour? What counts as a behavioural fact? Speaking a language is a form of behaviour, so could we count the fact that someone *said that it was raining* as a behavioural fact? Quine would not count this as a behavioural fact, since it already invokes a semantic notion, the notion of a speaker *uttering a sentence with a certain content*.

The whole point of the enterprise is to try to get some handle on what facts about content and meaning could be, so Quine will want to work with a notion of behaviour constrained by the requirement that that behaviour be described in completely non-semantic and non-intentional terms. In fact, Quine works with a very thin notion of the behavioural facts which can legitimately be invoked in the process of radical translation. For Quine, the only facts about native behaviour which can legitimately be counted as evidence for or against the correctness of a given translation manual are facts about what he terms *stimulus meaning*. The stimulus meaning of an expression is defined to be an ordered pair consisting, on the one hand, of those sensory stimulations which prompt assent to the sentence (the *affirmative stimulus meaning*) and, on the other hand, those sensory stimulations which prompt dissent from the sentence (the *negative stimulus meaning*). Thus, another way of describing the task the radical translator faces is: *given all of the facts about the stimulus meanings of the sentences of the native language, construct a translation manual which provides an acceptable translation of those sentences into your own language*. And of course, the indeterminacy thesis is the claim that this task cannot be discharged:

> Manuals for translating one language into another can be set up in divergent ways, all compatible with the totality of speech dispositions, yet incompatible with one another. In countless places they will diverge in giving, as their respective translations of the sentences of the one language, sentences of the other language which stand to each other in no plausible sort of equivalence however loose.[20]

One thing to note at this point is the *restricted* nature of the evidence which Quine allows as relevant to the process of radical translation: all of the evidence the translator is allowed by Quine concerns the dispositions the natives have to assent to (and dissent from) sentences in response to various stimulations of their sensory receptors. No other kind of evidence is allowed.

Note also that the meaning scepticism which is the intended upshot of Quine's argument is *constitutive scepticism* (scepticism

about the *existence of a certain sort of fact*), as opposed to traditional *epistemological scepticism* (which concedes that the sort of fact in question exists, and then questions our right to claim *knowledge* of that sort of fact). It is important to bear this distinction in mind, because it would be easy to make the mistake of seeing Quine as propounding only epistemological scepticism about meaning. Easy, because his argument has a distinctively epistemological flavour: given all of the facts about stimulus meanings, you are still not in a position to *know* which of the competing translation manuals is the correct one. But note that although Quine's argument has this epistemological flavour, it questions whether we could know which manual was correct, even given *idealised epistemological access* to all of the *possible* facts about stimulus meanings and the natives' behavioural dispositions. Even if we imagine our knowledge-acquiring powers with respect to stimulus meanings idealised, we *still* would not be in a position to justify our choice of one of the competing translation manuals over its rivals. It is this idealisation which allows Quine to argue for a constitutive scepticism *via* an epistemological argument.

So, all the evidence the radical translator has to go on in constructing his translation manual are facts about stimulus meanings, although as noted above he has idealised access to all such facts. Let's now think about how he fares in his task of constructing the manual. The translator first of all identifies the native signs for *assent* and *dissent*. He can do this by repeating his own volunteered pronouncements and then taking note of the natives' reaction:

> How is he to recognize native assent and dissent when he sees or hears them? Gestures are not to be taken at face value; the Turks' are nearly the reverse of our own. What he must do is guess from observation and then see how well his guesses work. Thus suppose in asking "Gavagai?" and the like, in the conspicuous presence of rabbits and the like, he has elicited the responses "Evet" and "Yok" often enough to surmise that they may correspond to "Yes" and "No", but has no notion which is which. Then he tries the experiment of echoing the native's own volunteered pronouncements. If thereby he pretty regularly

> elicits "Evet" rather than "Yok" he is encouraged to take "Evet" as "Yes". Also he tries responding with "Evet" and "Yok" to the natives' remarks; the one that is the more serene in its effect is the better candidate for "Yes". However inconclusive these methods, they generate a working hypothesis. If extraordinary difficulties attend all his subsequent steps, the linguist may decide to discard that hypothesis and start again.[21]

Suppose, then, that the translator notices that the natives are generally prepared to assent to the expression "Yo, gavagai" whenever there is a rabbit somewhere close nearby, and that they are generally prepared to dissent from "Yo, gavagai" when there are no rabbits present. In other words, the translator discovers that "Yo, gavagai", and "There is a rabbit", have the same stimulus meaning: the natives assent to and dissent from "Yo, gavagai" in exactly the same circumstances (under the same conditions of sensory stimulation) as we would assent to and dissent from "There is a rabbit". On this basis, the radical translator (let's suppose for the sake of argument that he is English) tentatively proposes to translate "Yo, gavagai" into English as "There is a rabbit".

Now the trouble starts. What if we choose to translate "Yo, gavagai", not as "There is a rabbit", but rather as "There is an undetached rabbit part"? Intuitively, we want to say that this alternative translation is inaccurate: but facts about stimulus meaning – remember, the only facts that we are allowed to appeal to – will be of no help to us in choosing between these two candidate translations. Each of the translations is consistent with all of the relevant evidence about the natives' behavioural dispositions: "There is a rabbit" and "There is an undetached rabbit part" have exactly the same stimulus meanings (since whenever there is a rabbit present there is also an undetached rabbit part present, and vice versa), so that so far as the relevant evidence goes, the translation of "Yo, gavagai" as "There is an undetached rabbit part" is just as good as the translation of "Yo, gavagai" as "There is a rabbit". Since stimulus meanings constitute the only relevant evidence, the only facts relevant to translation, there is no fact of the matter as to which of the translation manuals is the correct one. By the same reasoning, there is no fact of the matter as to

whether "Yo, gavagai" should be translated as "there is a rabbit", or "there is an undetached rabbit part", or "there is a time slice of a four-dimensional rabbit-whole", or "there is an instantiation of rabbit-hood". No doubt, Quine is prepared to admit, in practice we will choose one of these translations over the others, but in so doing we will not thereby be faithful or unfaithful to some *fact* about the meaning of "Yo, gavagai". Our choice will merely be a reflection of pragmatic standards: each of the translations is equally justified by the facts about stimulus meanings, so that as far as the facts of the matter are concerned, all of the translations are equally acceptable.

Quine considers and quickly rejects the following response to the above argument: perhaps it is possible, on the basis of facts about stimulus meaning alone, to tell whether or not "Yo, gavagai" is better translated as "there is a rabbit" or "there is an undetached rabbit part". We can find out which translation is really the correct one by putting certain *questions* to the native. Suppose that we have identified the expression the natives use for *numerical identity* (such as "is the same as"), and also some of the expressions they use as *demonstratives* (such as "that", or "this"). Then surely all we need to do to see which translation is correct is to point to one part of the rabbit (e.g. its tail) and ask, after pointing to its nose, "Si hit gavagai emas sa hat gavagai?"(in other words, "Is this gavagai the same as that gavagai?"). If the native responds with "Evet", which we have already identified as his sign for assent, then this would be genuine behavioural evidence in favour of the translation of "gavagai" as "rabbit" and against the translation of "gavagai" as "undetached rabbit part". And vice versa if he responded with "Yok" to the same query.

Quine points out that this reply assumes that the correct translation of the native expression "emas" is "same rabbit", so that when we put the query above we were asking "Is this rabbit the same rabbit as this rabbit?". The answer "Evet" then spoke in favour of the straightforward "rabbit" translation. But if we take the correct translation of "emas" to be "is an undetached part of the same rabbit as" the answer "Evet" to our query will actually be consistent with the less straightforward "undetached rabbit part" translation. "Si hit emas gavagai sa hat gavagai?" will come out as "Is this undetached rabbit part an undetached part of the same

rabbit as that undetached rabbit part?", and if the natives mean "undetached rabbit part" by "gavagai" the answer "Evet" is precisely what we would expect. So there are ways of rendering even this new source of behavioural evidence compatible with each of the competing translation manuals. Gareth Evans sums up this reply of Quine's very nicely:

> An expression may sensibly be regarded as a predicate only if it interacts with the "apparatus of individuation", and the stimulus conditions that trigger assent to the sentences in which such interaction occurs provide the only empirical evidence that bears upon what extension the expression, as a predicate, should be regarded as possessing. But the identification of the apparatus of individuation in a foreign language is empirically quite underdetermined; the expression that one theory regards as the identity predicate, may, with suitable adjustments be treated by another as an expression for some distinct equivalence relation – both theories assigning to whole sentences a significance completely in accordance with the behavioural data. Consequently, whether an expression is a predicate at all, and if a predicate, what extension it has, are matters underdetermined by all actual and possible observations.[22]

That, then, is the essence of the argument from below. In the next section we examine Evans's own response to this argument.

4.8 Evans and Hookway on the argument from below

The conclusion of the argument from below was as follows. For any language L, there will be indefinitely many translation manuals T1, . . ., Tn, . . . each of which accommodates the facts about the speakers' behavioural dispositions, but which offer what are intuitively incompatible translations of the sentences of L. Since the facts about speakers' behavioural dispositions (the facts about stimulus meaning) are the only facts relevant to translations, it cannot be a factual matter which of the candidate translation manuals is the correct one. And since a translation manual is

supposed to tell us what the speakers mean, it is not a factual matter as to what the speakers mean either.

We shall now look at the response to Quine's argument from below which is developed by Gareth Evans in his paper "Identity and Predication". Evans's response is complex, so we shall simplify it somewhat in order to get across the general line of thought.

Evans's strategy is to deny that it follows from the claim that there is an indeterminacy afflicting the construction of translation manuals that a similar degree of indeterminacy afflicts the notion of meaning. He denies that indeterminacy in the construction of translation manuals for a language implies indeterminacy in the construction of *theories of meaning* (or semantic theories) for that language. Quine clearly thinks that it follows from the indeterminacy of translation that "different theories of meaning could be constructed each of which entails semantical properties for the totality of jungle sentences adequate to the data but which make quite different assignments of semantical properties to those sentences' parts".[23] Evans denies this: according to him the project of constructing a translation manual and the project of constructing a theory of meaning are quite separate enterprises, governed by different sorts of constraints.

But what is a theory of meaning? As explained in §1.7, a semantic theory is a theory relating to a single language which attempts to *state the meaning* of every sentence in that language. Such a theory delivers, for each well-formed declarative sentence of the language, a *theorem* which in some way states the meaning of that sentence. We saw in Chapter 1 that according to Frege, the sense of a sentence is its *truth-conditions*, and we will follow Frege on this point here. Thus, a theorem specifying the meaning of the sentence "Snow is white" would be a theorem which stated the truth-conditions of that sentence, namely:

"Snow is white" is true if and only if snow is white

Thus, we can say that a correct semantic theory for a language should issue a theorem which states the truth-conditions of each well-formed declarative sentence in the language. So what is the

difference between a translation manual and a semantic theory? Evans writes:

> A translation is one thing, a theory of meaning another. A manual for translation aims to provide, for each sentence of the language under study, a way of arriving at a quoted sentence of another language which has the same meaning. A theory of meaning, on the other hand, entails, for each sentence of the language under study, a statement of what it means. A translator states no semantical truths at all, nor has he any need of the concepts of truth, denotation, and satisfaction. Semantical truths relate expressions to the world, and can be stated only by using, not mentioning, expressions of some language or other.[24]

More importantly, Evans claims that there are constraints on the construction of theories of meaning which rule out the indeterminacy which Quine claims to have found in translation manuals. The constraint which rules out the indeterminacy in theories of meaning concerns *structure*. A theory of meaning has to be structured in a way in which a translation manual does not, and it is the presence of this structure in the theory of meaning which rules out indeterminacy. The translator can divide a sentence or phrase into parts in any way he finds convenient for his practical task of setting up a mapping between expressions. Quine realises that finding structure in sentences (constructing *analytical hypotheses*, as he calls them) is in fact how the practical business of constructing a translation manual will have to proceed:

> [The translator] segments heard utterances into conveniently short recurrent parts, and thus compiles a list of native "words". Various of these he hypothetically equates to English words and phrases, in such a way [as to respect the facts about stimulus meanings]. Such are his analytical hypotheses.[25]

Evans's point is that the only constraint the translator has to respect on the choice of analytic hypotheses is that they generate the right answers about stimulus meanings of the whole sentences in which the relevant expressions occur. The theorist of meaning,

however, is not so unconstrained in projecting structure into his theory: the semantic theory must be so structured that (a) the truth-conditions of the sentences of the language are shown to be dependent upon the semantic properties of the *parts* of those sentences, and (b) the structure of the theory *mirrors* the structure of the linguistic *abilities* of the actual speakers of the language.

Let's try to get clearer on this. The theorist of meaning aims to give a systematic account of the truth-conditions of the sentences of the language under study: he must show how those truth-conditions depend functionally on the semantic properties of those sentences' parts. Thus, for example, he will want to construct a theory of meaning which shows how the truth-condition of, for example, "Lightning is a white cat" is a function of the semantic properties of "Lightning", "white", "cat", and so on. But the semantic theorist is not free to assign semantic properties to those parts in any way that he chooses. He must assign semantic properties to those parts in such a way that the following additional constraint is respected:

The mirror constraint: A theorem specifying the truth-condition of a sentence S should be derivable from a specification of the semantic properties of S's subsentential expressions s1, . . ., sn if and only if actual speakers of the language are able to move to an understanding of S on the basis of training with and exposure to s1, . . ., sn as they appear in sentences other than S.

Suppose that speakers of English who understand "Lightning is a white dog" and "Archie is a black cat" are able, without further training and exposure, to understand "Lightning is a white cat": given this, the mirror constraint says that the theory of meaning must show how the truth-condition of "Lightning is a white cat" is dependent upon the semantic properties of "white", "cat", etc. In other words, the theory of meaning must have axioms for "white", "cat", etc., from which the theorem spelling out the truth-condition for "Lightning is a white cat" is derivable. The structure

thus projected into the theory of meaning mirrors the structure of the speakers' linguistic abilities: just as, on the basis of their prior understanding of the subsentential expressions "Lightning", "white", and "cat", and so on, they are able to understand the sentence "Lightning is a white cat", the theory of meaning shows us how the truth-condition of "Lightning is a white cat" is derivable from the semantic axioms for "Lightning", "white", "cat", and so on.

What is the relevance of this to the indeterminacy of translation? In other words, how does the imposition of the mirror constraint on theories of meaning ensure that they are not prey to indeterminacy? Suppose that we set out to construct a theory of meaning (as opposed to a translation manual) for the native language, and decide, on the basis of facts about stimulus meaning, to interpret "blap" and "gavagai" as meaning *white* and *rabbit* respectively. That is, we decide to include the following as axioms in our theory of meaning:

(1) (∀x)(x satisfies "gavagai" iff x is a rabbit)
(2) (∀x)(x satisfies "blap" iff x is white).[26]

Now Quine, wishing to claim indeterminacy in the theory of meaning, will claim that there will be an alternative theory of meaning for the native language, adequate to all of the behavioural data, incorporating not (1), but rather an axiom along the following lines:

(3) (∀x)(x satisfies "gavagai" iff x is an undetached part of a rabbit).

How can Evans rule out this alternative theory of meaning? Evans claims that, because of the imposition of the mirror constraint, the alternative theory of meaning incorporating (3) will give us an incorrect account of the truth-conditions of some complex sentences in which "gavagai" features. Suppose the theorist of meaning discovers that after exposure to or training with sentences containing "gavagai" and sentences containing "blap", speakers of the native language are able to use and understand "blap gavagai". Then the mirror constraint requires that the theorist of meaning give an account of the truth-conditions of "**a** is a blap gavagai" which

shows how this condition is systematically dependent upon the satisfaction conditions of "blap" and "gavagai". That is, we must be able to generate a truth-condition for "**a** is a blap gavagai" on the basis of the axioms which deal with "blap" and "gavagai". Now, on the basis of our axioms (1) and (2), we will get the following truth-condition for "**a** is a blap gavagai":[27]

(4) "**a** is a blap gavagai" is true iff **a** is a white rabbit.

But what kind of truth-condition for "**a** is a blap gavagai" will a theory of meaning containing axiom (3) generate? If we want to hold on to axiom (2) it will have to be something like:

(5) "**a** is a blap gavagai" is true iff **a** is a white undetached rabbit part.

But this won't do! It gets the native stimulus meaning for "**a** is a blap gavagai" entirely wrong. We notice that the natives are prepared to assent to "**a** is a blap gavagai" only when **a** is a wholly white rabbit; whereas if (5) accurately captured its truth-condition, we would expect to find the natives assenting to "**a** is a blap gavagai" in the presence of an otherwise brown rabbit with a white foot. But they don't: in fact they dissent from "**a** is a blap gavagai" in the presence of such a rabbit.

Is there a way out of this for Quine? Perhaps the trouble arose only because we also attempted to hold on to axiom (2) for "blap", treating "blap" as meaning white. Maybe Quine can hold on to (3), and yet respect the facts about stimulus meaning, by dropping (2) and switching instead to

(6) $(\forall x)$(x satisfies "blap" iff x is a part of a white animal).

This, together with (3), would give us the following truth-condition for "**a** is a blap gavagai":

(7) "**a** is a blap gavagai" is true iff **a** is an undetached rabbit part which is a part of a white animal.

Given these adjustments we can account for the conditions under which "**a** is a blap gavagai" is assented to: the native assents to this when a wholly white rabbit is around because there is an undetached rabbit part around which is a part of a white animal; but he dissents from it when there is a brown rabbit with a white foot around because although we have an undetached rabbit part around it is not a part of a white animal.

Is this enough to give Quine his desired alternative but incompatible translation scheme? As Evans points out, it is not. The reading of "blap" encapsulated in (6) now precludes any account of the functioning of "blap" as it appears conjoined with terms other than "gavagai": it cannot account, for example, for the natives' applications of "blap" to white huts or white pieces of paper.

This, then, is the crux of Evans's argument against Quine. The theorist of meaning has to assign semantic properties to the parts of sentences in such a way that the observed assent conditions of those sentences can be accommodated, and in such a way that the assignment is consistent with the occurrence of those parts in all of the contexts in which they appear. But given this, we can see that the interpretation of "gavagai" as "undetached rabbit part" faces the following dilemma. *If we leave the satisfaction conditions of the expressions with which "gavagai" couples unchanged then we get an account of the truth-conditions of the complex sentences which contain "gavagai" which is out of step with their observed assent conditions; but if we change the satisfaction conditions of some of the other expressions in such a way as to accommodate these assent conditions, then we get an account of the truth-conditions of other sentences containing "blap" which is out of step with their assent conditions.*

Evans thus sees the extra constraints on theories of meaning concerning the systematic relation of sentences' truth-conditions to the semantic properties of their parts as ensuring that a theory of meaning cannot include an axiom to the effect that "gavagai" means *undetached rabbit part*. If we construe "gavagai" as ranging over undetached rabbit parts we do not get a workable account of how the truth-conditions of complex sentences containing "gavagai" depend systematically on the semantic properties of their parts.

So the indeterminacy which Quine claimed to find in translation does not vitiate theories of meaning.

Christopher Hookway has suggested that there may be a response available to Quine which would allow him to say that there is still an indeterminacy even in the case of theories of meaning. Contrary to what Evans claims, there is a way of amending the axiom for "blap" which (a) allows us to interpret "gavagai" as meaning *undetached rabbit part*, (b) is faithful to the stimulus meaning of the sentence "**a** is a blap gavagai", and (c) does not preclude the possibility of "blap" meaningfully occurring with terms other than "gavagai". Hookway's suggested Quinean axiom for "blap" is:

(8) (∀x)(x satisfies "blap" if and only if either (a) "blap" occurs together with "gavagai" and x is part of a white animal, or (b) "blap" occurs in some other context and x is white).

Having a "disjunctive" axiom like (8) allows us to keep axiom (3) for "gavagai", while remaining faithful to the stimulus meaning of "**a** is a blap gavagai", and yet, through clause (b), allows "blap" to occur with terms other than "gavagai". Quine will of course admit that we will prefer a simpler axiom for "blap", such as (2), but again he will say that this preference merely reflects our pragmatic inclinations, rather than our faithfulness to any fact of the matter.

How might Evans respond to this? He might respond that there is a way of distinguishing, on empirical grounds, between a theory of meaning containing axiom (2) and a theory of meaning containing axiom (8) (this response is suggested by Hookway himself). The mirror constraint is designed to ensure that a theory of meaning for a language mirrors the structure of the linguistic abilities of speakers of that language. At the level of the semantic axioms, what this means is that each axiom will attribute to a speaker a disposition corresponding to the relevant primitive expression.[28] Thus, a theory containing axiom (2) will attribute a *single* dispositional state underlying the speaker's uses of "blap". And a theory containing axiom (8) will attribute two dispositional states underlying the speaker's uses of "blap": a state underlying the use of "blap" as it occurs together with "gavagai", and another state

underlying the use of "blap" as it appears together with expressions other than "gavagai". So which theory should we accept? It looks as if it is going to be the theory that contains axiom (2) since we find, by empirical observation, that the initial training the natives get with "blap" is sufficient to equip them to use it in all contexts, and that there is no need for separate training with it as it appears conjoined with "gavagai". This suggests that there is only a single dispositional state underlying the speaker's uses of "blap", so that there are good empirical reasons to opt for the theory of meaning containing (2) rather than that containing (8).[29]

4.9 The argument from above

In this section we move on to look at Quine's other main line of argument for the indeterminacy of translation, the "argument from above". This is presented in Quine's short paper "On the Reasons for the Indeterminacy of Translation", which appeared ten years after the publication of *Word and Object*. Quine writes:

> My gavagai example has figured too centrally in discussions of the indeterminacy of translation. Readers see the example as the ground of the doctrine, and hope by resolving the example to cast doubt on the doctrine. The real ground of the doctrine is very different, broader and deeper.[30]

The argument from above is thus conceived by Quine to be independent of and more fundamental than the argument from below. It depends on the idea that physical theory is underdetermined by all possible observational evidence. As Quine puts it:

> Theory can still vary though all possible observations be fixed. Physical theories can be at odds with each other and yet compatible with all possible data even in the broadest sense. In a word they can be logically incompatible and empirically equivalent. This is a point on which I expect wide agreement, if only because the observational criteria of theoretical terms are commonly so flexible and so fragmentary.[31]

This follows from Quine's *epistemological holism*, discussed above in §4.4. As we saw, the epistemological holist asserts that significant statements face the tribunal of experience not *individually*, but *en masse*: "our statements about the external world face the tribunal of sense experience not individually, but only as a corporate body". So suppose we are faced with a *recalcitrant* experience – an experience which conflicts with our currently held physical theory in conjunction with a set of hypotheses ("auxiliary hypotheses") describing experimental set-up, laboratory conditions, state of the experimenter, etc. According to epistemological holism we have the choice – at least in principle – of giving up some part of our physical theory, or revising our auxiliary hypotheses, or maybe even giving up the claim that a recalcitrant experience has occurred. Likewise, if we decide to give up some part of our physical theory, *which* particular part we give up will be underdetermined by all actual and possible observations: we can hold on to any part of our physical theory, provided we are willing to make the requisite adjustments elsewhere, among the auxiliary hypotheses or wherever. It follows that given any set of actual or possible observations, we will have – at least in principle – a choice between a range of competing theories, all of which can be chosen consistently with the observational data subject to appropriate revisions elsewhere in the set of our empirical (or even logical or mathematical) beliefs.

Quine notes that although most people will agree that physical theory, at some level or other, will be underdetermined by all possible observational evidence, there will be disagreement about the precise level at which the underdetermination kicks in:

> Some will acknowledge such slack only in the highest and most speculative reaches of physical theory, while others see it as extending even to common-sense traits of macroscopic bodies.[32]

Quine does not have to enter into this debate at this point, since his aim in the argument from above is to convince us that the indeterminacy of translation extends as far as the underdetermination of physical theory by possible observations, however far the latter underdetermination is thought to extend:

> What degree of indeterminacy of translation you must then recognize . . . will depend on the amount of empirical slack that you are willing to acknowledge in physics. If you were one of those who saw physics as underdetermined only in its highest theoretical reaches, then . . . I can claim your concurrence in the indeterminacy of translation only of highly theoretical physics. For my own part, I think the empirical slack in physics extends to ordinary traits of ordinary bodies and hence that the indeterminacy of translation likewise affects that level of discourse.[33]

How, then, does Quine intend to get from the premise that physical theory is underdetermined by all possible observations, to the conclusion that translation is indeterminate? The whole of the argument is contained in one paragraph of "On the Reasons".[34] This time we are attempting radical translation of some foreign physicist's theory. Following Robert Kirk's excellent presentation in *Translation Determined*, the argument is best broken down as follows:

[1] The starting point of the process of radical translation is the equating of observation sentences of our language with observation sentences of the foreigner's language, via an inductive equating of stimulus meanings.
[2] In order afterward to construe the foreigner's theoretical sentences we have to project analytical hypotheses.
[3] The ultimate justification for the analytical hypotheses is just that the implied observation sentences match up.
[4] Insofar as the truth of a physical theory is underdetermined by observables, the translation of the foreigner's physical theory is underdetermined by translation of his observation sentences.

Therefore:

[T] Translation of physical theories is indeterminate at least to the extent that physical theories are underdetermined by all possible observations.

Some comments are in order here. First of all, Quine is adamant that the indeterminacy of translation which the argument purports to establish is not just another example of the underdetermination of theory by observational evidence, as applied to the special case of translation manuals:

> The indeterminacy of translation is not just an instance of the empirically underdetermined character of physics. The point is not just that linguistics, being a part of behavioral science and hence ultimately of physics, shares the empirically underdetermined character of physics. On the contrary, the indeterminacy of translation is additional.[35]

Quine obviously has to say this, since the fact that translation is indeterminate is supposed to establish *that there are no facts of the matter* about the correctness or otherwise of translation manuals. But he certainly does not want to conclude, from the fact that physics is observationally underdetermined, that there are no facts of the matter about physics. As Hookway points out, Quine endorses J. J. C. Smart's claim that "the physicist's language gives us a *truer* picture of the world than the picture of common sense".[36] Now physics can only give us a truer picture of the world if it is possible to have a *true* picture of the world, and to say that it can give us a true picture of the world is to say that there are *facts* about physics. So Quine must say that the indeterminacy of translation is "additional". But how can he say this? Chomsky has expressed well the thought that the "indeterminacy" Quine claims to find in translation is just *another* instance of the underdetermination of theory by observational data:

> It is quite certain that Quine has failed to show that serious hypotheses concerning a native speaker's knowledge of English . . . will "go beyond the evidence". Since they go beyond mere summary of data, it will be the case that there are competing assumptions consistent with the data. But why should all of this occasion any surprise or concern?[37]

Chomsky thus attempts to argue that no invidious distinction can be drawn between linguistics (or translation) and physics.

However, Kirk spells out well why the indeterminacy of translation is indeed additional, and thus how Quine gets the better of Chomsky here. The point to appreciate is that Quine's claim is that:

> Theories of translation are not only underdetermined *as* physics is underdetermined, but underdetermined even *by* the totality of truths expressible in terms of physics.[38]

Likewise, Hookway says that:

> Quine has several times stressed that, while theory choice in physics is not determined by all possible evidence, his point now is that, once we have made a choice of physical theory, our choice of translation manual is still open.[39]

Given Quine's physicalism, the claim that incompatible translation manuals are consistent with the totality of *physical facts* does entail that there is no fact of the matter as to which translation manual is correct. This contrasts with, for example, chemical theory: although this too is underdetermined by all possible observational evidence, once we have made a choice of physical theory, our choice of chemical theory would *not* remain open.

Now for some comments on individual premises. [1] seems plausible enough. I begin the process of radical translation by trying to match my observation sentences[40] with the observation sentences of the foreigner on the basis of facts about sameness of stimulus meaning. And as Kirk puts it:

> In general the sentences of physical theory, regardless of the extent to which they are supposed to be underdetermined by observables, are not observation sentences. So there will not be much – if any – equating of theoretical sentences in the early stages.[41]

Likewise, [2] seems plausible enough. As Kirk says "There seems no way to construct a useful translation manual at all without devising analytical hypotheses".[42] Kirk notes that on one way of reading [3], it is simply a *restatement* of the *argument from below*. All it says is that the choice of analytical hypotheses is

underdetermined by all of the possible facts about stimulus meanings: in other words, there will be incompatible analytical hypotheses which are each compatible with all of the facts about stimulus meanings. But, as Kirk stresses, it is very important to note that [3] cannot be allowed to have this significance in the argument from above: if we read [3] in this way, the indeterminacy of translation is *already* derived at [3] and the rest of the argument becomes redundant. But the argument from above is explicitly constructed by Quine with the aim of convincing those who find the argument from below *unconvincing*. And the argument from above is supposed to be the real ground of the indeterminacy doctrine, not the argument from below. So we must not read [3] as a restatement of the argument from below.

How plausible is the argument from above? It is difficult to assess, since it is unclear precisely how Quine sees the move from the premises to the conclusion being effected, and he gives us little in the way of elucidation. In any event, we shall now see that Kirk has an apparently powerful objection against the argument from above. *The objection is that given that [3] is not to be read as a restatement of the argument from below, there is a non-sequitur in the move from [4] to [T].*

Kirk invites us to consider Fred the Physicist. Fred is trying to translate Chinese physics. He believes (a) that physics is underdetermined by observational evidence only at the level of quark theory and above, and (b) that physics is not underdetermined by observational evidence below this (e.g. at the level of protons, electrons, and positrons etc.). Quine's argument from above has the aim of convincing him that his translation of the Chinese physics *at the level of quarks and above* is indeterminate. Also, Quine must grant Fred, for the sake of argument, that his translation of Chinese physics below the level of quark theory is not indeterminate. Kirk claims that Fred can now maintain that he can determinately translate Chinese quark theory on the basis of his determinate translation of Chinese physics at all of the levels lower than quark theory:

> Fred maintains that there is a large class C of theoretical sentences of Chinese physics [those dealing with levels lower

> than quark theory] which (a) are not observation sentences, yet (b) are determinately translatable; and that his supposedly determinate translations of all members of C supply him with a sufficiently solid basis to ensure that his translations of Chinese high-level particle physics [quark theory and beyond] are determinate too.[43]

Kirk's claim is that unless Quine takes [3] to already establish the indeterminacy of translation, he has nothing to say in response to this suggestion about how Fred can determinately translate Chinese quark theory. Kirk notes that the conclusion [T] can actually be spelled out as:

> [T*] To the extent that the truth of a physical theory is underdetermined by observables, the translation of the foreigner's physical theory is underdetermined by the totality of facts.

Kirk argues that [T*] would follow from [4] only given the assumptions that the only facts relevant to translation are facts about the translations of observation statements and that those facts do not serve to determine a uniquely correct translation. But those assumptions amount to the argument from below. If Quine does not make those assumptions, as he must not if the argument from above is to not to assume the argument from below as one of its premises, he appears to have nothing to say in response to the question: "Why should the determinacy of translation of higher-level theoretical sentences not be guaranteed by behavioural dispositions relating to lower-level theoretical sentences?".[44]

Thus, Kirk argues, the argument from above is either invalid (if [3] is not read as a statement of the argument from below); or valid only at the cost of presupposing the argument from below (if [3] is read as a statement of the argument from below). Kirk concludes that Quine has no compelling argument for the indeterminacy doctrine which is "broader and deeper" than the argument from below.[45]

4.10 Conclusion

We have now examined both the argument from below and the argument from above for the indeterminacy thesis, and we have found that there are serious objections that either argument must face. Deciding whether these objections are ultimately compelling is outwith the scope of the present work. Instead, we should end by noting an inherent weakness in Quine's arguments. The only facts that are allowed as possible candidates for constituting the correctness of a translation manual are observable facts about the behavioural dispositions of the speakers, where the notion of "behaviour" involved is a very austere one (see §4.7). As Kripke has put it:

> Quine bases his argument from the outset on behavioristic premises . . . Since Quine sees the philosophy of language within a hypothetical framework of behavioristic psychology, he thinks of problems about meaning as problems of disposition to behavior.[46]

So even if Quine's arguments can get over the hurdles we have considered, they will even then fail to convince those who fail to find this restriction to thin behavioural facts, and the associated behaviouristic psychology, compelling.[47] It would be much more interesting if we could find arguments which have the same conclusions as Quine's – that there are no facts which render ascriptions of meaning true or false – which do not depend on such behaviouristic assumptions.[48] We consider one such argument – that developed by Kripke's Wittgenstein – in the next two chapters.[49]

Chapter summary

We considered four arguments about meaning developed by Quine. First, the argument in the first four sections of "Two Dogmas of Empiricism", to the effect that no-one has ever given a non-circular explanation of the notion of analyticity. Second, the argument in the last two sections of "Two Dogmas" to the effect that since individual statements do not possess their own verification conditions,

the logical positivist theory of meaning implies that no statements are either analytic or synthetic. Third, the "argument from below", developed in Chapter 2 of Quine's *Word and Object*, to the effect that there are no facts of the matter about the correct translation of the sentences of a language. Fourth, the argument in Quine's paper "On the Reasons for the Indeterminacy of Translation", to the effect that translation is indeterminate to the extent that physical theory is underdetermined by all possible observations. We looked at a range of challenges to these arguments from Gareth Evans, Crispin Wright, and Robert Kirk.

Study questions

1. If the notion of synonymy makes no sense, can we hold on to the idea that a sentence can have a determinate meaning?
2. How plausible is Quine's Socratic Assumption?
3. Does the fact that "any statement can be held true come what may" imply that all statements are synthetic?
4. What is the difference between a translation manual and a theory of meaning for a language? Is the difference significant for the cogency of Quine's "argument from below"?
5. Outline how Quine's idea that physical theory is underdetermined by all possible observations is a consequence of the epistemological holism of the last two sections of "Two Dogmas".
6. How convincing is Kirk's case that Quine has no compelling argument for the indeterminacy doctrine which is "broader and deeper" than the argument from below?

Further reading

On analyticity, the reader should certainly start with a careful reading of Quine's "Two Dogmas of Empiricism", since we have only been able to give a flavour of his arguments from this classic paper. After that, the thing to read is Grice and Strawson's "In Defense of a Dogma", which contains much interesting and useful argument in addition to that mentioned in the text. Grice and Strawson are criticised by Harman in his "Quine on Meaning

and Existence (I)". Another useful paper is Putnam's "The Analytic and the Synthetic". Chapters 2 and 3 of Hookway's *Quine* contain much useful background material on Quine's relationship to the logical positivist tradition. Wright's "Inventing Logical Necessity" contains some excellent exposition and commentary on "Two Dogmas".

An important development in the recent literature is Boghossian's distinction between metaphysical and epistemological notions of analyticity. A sentence S is metaphysically analytic if it is true (false) simply in virtue of its meaning, whereas a sentence S is epistemically analytic if it is possible to have a justification for holding S simply in virtue of grasping S's meaning. Boghossian agrees with Quine that the metaphysical notion is hopeless, but he thinks that the epistemological notion might be viable and may even help to explain *a priori* justification. See Boghossian, "Analyticity" (and his new "Postscript: Further Thoughts about Analyticity 20 Years On").

For a recent defence of the notion of analyticity, see G. Russell, *Truth in Virtue of Meaning*. See also Fodor and LePore, "Analyticity Again".

It is worth noting here that the arguments of the final two sections of "Two Dogmas" have received more acceptance than the arguments of the first four sections, which are now generally acknowledged to be unsuccessful. So suppose, for the sake of argument, that Wright's criticisms of holism are unsuccessful. Is there any way a defender of Quine can hive off the arguments in the final two sections from the arguments of the first four sections – in other words, can Quine concede that the arguments in the first four sections are unconvincing without thereby giving the game away to Carnap and the logical positivists? See Hookway, *Quine*, pp.37–47, and Wright, *Truth and Objectivity*, pp.154–55, for arguments to the effect that Quine may indeed be able to do this.

On indeterminacy of translation, the classic texts are Chapter 2 of Quine's *Word and Object*, and his paper "On the Reasons for the Indeterminacy of Translation". The most important papers in the secondary literature are Evans's "Identity and Predication" and Chomsky's "Quine's Empirical Assumptions". Hookway's *Quin*e, Chapters 8–10, contains much superlatively clear exposition and valuable discussion of the main issues (the influence of Hookway's presentation on this chapter should be clear enough). Kirk's *Translation Determined* is a very detailed discussion of the

issues; Chapter 6 is extremely good on the "argument from above". For a debate about the role of simplicity considerations in cutting down indeterminacy, see Wright, "Indeterminacy of Translation" and Miller ,"Meaning Scepticism", §2.4.

See also Orenstein, *W. V. Quine*, Chapter 6; Hylton, *Quine*, Chapters 3 and 6; and the papers by Creath and Kirk in Gibson (ed.), *The Cambridge Companion to Quine*. There are also many relevant papers in Hahn and Schilpp (eds), *The Philosophy of W. V. Quine*. This volume also contains a short intellectual autobiography by Quine. A very useful overview of Quine's philosophy can be found in Kemp, *Quine: A Guide for the Perplexed*.

Notes

1 Quine, "Two Dogmas of Empiricism", pp.20–21.
2 Ibid., p.20.
3 See Chapter 1, where we note that in Frege's logical language, the semantic value of a predicate is a function (understood in extensional terms), and expressions with the same semantic value can be intersubstituted *salva veritate*.
4 Quine, "Two Dogmas", p.30.
5 Grice and Strawson, "In Defense of a Dogma", p.146.
6 We're using "concept" in an informal – as opposed to Fregean – sense here.
7 So called because it underlies the strategy famously adopted by Socrates in Plato's dialogues. For example, in *The Republic* Socrates asks for explicit non-circular definitions of "good" and "justice".
8 Quine, "Two Dogmas", p.32.
9 Grice and Strawson, "In Defense of a Dogma", pp.149–153.
10 Ibid., p.153.
11 The argument of this section derives from Wright's excellent paper "Inventing Logical Necessity".
12 Quine, "Two Dogmas", p.41.
13 Ibid., p.42.
14 Ibid., p.43.
15 Note that insofar as the distinction between linguistic and factual components of a sentence corresponds to Frege's distinction between sense and semantic value, Frege's distinction is also endangered by Quine's argument.

16 So it would a mistake to take sections 5 and 6 of "Two Dogmas" to be arguing that all statements are synthetic. (Note, though, that there are some reasons for thinking that Quine doesn't distinguish sufficiently clearly between non-factualism about "analytic" and an error theory about "analytic", see Boghossian, "Analyticity", pp.586–588).
17 Wright, "Inventing Logical Necessity", p.192.
18 Clearly, nothing turns on the deliverance by the method that e.g. the rejection of (1) is a good move. The argument would go through as before if we got the answer that the rejection of (1) was a bad move.
19 I'm not suggesting that Quine himself is motivated by this idea about convention, just that it is one way you could motivate the sort of behaviouristic restriction he works with.
20 Quine, *Word and Object*, p.27.
21 Ibid., pp.29–30.
22 Evans, "Identity and Predication", pp.27–28.
23 Ibid., p.27.
24 Ibid., p.25.
25 Quine, *Word and Object*, p.68.
26 Note that to say that x satisfies "gavagai" is just to say that x is in the extension of "gavagai" (or that x is paired with the truth-value T in the extension of "gavagai").
27 I'm simplifying here somewhat. Technically, we ought to have an axiom spelling out the reference of the name "**a**", and a compositional axiom governing the coupling of names with predicates. But it should be obvious how the semantic theory described in the text can be extended to include these. I leave this as an exercise for the reader.
28 For a development of this idea see Evans, "Semantic Theory and Tacit Knowledge". Note that when we speak of an object possessing a *disposition*, we could mean either (a) a *behavioural regularity* which that object exhibits, or (b) an *underlying state* of the object which *causally explains* the behavioural regularity in question. When we say that e.g. sugar is disposed to dissolve in water we could be referring either to the behavioural *regularity* (dissolving in water in standard conditions of temperature and pressure) which sugar displays; or we could be referring to some *underlying state* which sugar possesses and which explains *why* sugar exhibits the behavioural regularity in question. Evans makes it clear that he is working with the second, "full-blooded" notion of a disposition. The notion of a disposition is a very important and controversial one in many areas of contemporary

philosophy, including philosophy of mind, philosophy of science, and metaphysics. For a comprehensive treatment, see Mumford, *Dispositions*.

29 For further discussion, see Hookway, *Quine*, esp. pp.155–162. There may be a case for thinking that Hookway has not described Evans's response as well as he might have. As noted in the text, Hookway says that according to Evans, acceptance of a theory of meaning containing (8) will involve the attribution to speakers of two dispositional states underlying their uses of "blap". We can then check empirically to see whether they actually have two such dispositions, and see thereby whether the acceptance of the relevant theory of meaning is empirically justified. But maybe this is not the best way to frame Evans's response. In "Semantic Theory and Tacit Knowledge" (see the previous note) Evans's requirement is that *there be a disposition corresponding to each expression for which the theory provides a distinct axiom*. Now in the theory of meaning containing axiom (8) there is only one disposition corresponding to "blap" attributed to speakers: the disposition to assent to "Yo, blap gavagai" in the presence of undetached rabbit parts which are parts of white animals, and "Yo, blap x" for "x" other than "gavagai" when x is white. So on Evans's account, the acceptance of the theory containing (8) requires the attribution of only one disposition, the disposition described above, and not two dispositions as claimed by Hookway. But now the real problem with the acceptance of the theory containing (8) becomes apparent. As mentioned in the previous note, Evans is explicity working with a full-blooded notion of a disposition, as an underlying state which causally explains the obtaining of a behavioural regularity. But what sort of unitary underlying state of a subject could underlie a "bifurcated" regularity like the regularity which corresponds to (8)? This is a difficult question to answer, but the onus is surely on Evans's opponent to spell out what sort of state could underlie a disposition like that which corresponds to (8).

30 Quine, "On the Reasons", p.178.

31 Ibid., p.179. Quine also writes on p.182 that "The gavagai example was at best an example only of the inscrutability of terms, not of the indeterminacy of translation of sentences". According to the inscrutability of reference thesis, we can have translation manuals that respect all of the facts about stimulus meaning but differ over the assignments of referents to the subsentential expressions in the language. These translation manuals needn't diverge with respect to truth-value: in Quine's "gavagai" example, "There is a rabbit" and "There is an undetached rabbit part", though

inequivalent, are both true. According to the indeterminacy of translation thesis proper, we can have translation manuals that respect all of the facts about stimulus meaning but which are inequivalent in the sense of *differing in truth-value*. To keep things simple I've ignored the distinction between inscrutability and indeterminacy proper in the text.

32 Ibid.
33 Ibid., p.181.
34 Ibid., pp.179–180.
35 Ibid., p.180.
36 See Hookway, *Quine*, p.203, and passim for much useful discussion of Quine's realism about physics.
37 Chomsky, "Quine's Empirical Assumptions", p.67.
38 Kirk, *Translation Determined*, p.136.
39 Hookway, *Quine*, p.137.
40 Quine, *Word and Object*, §10.
41 Kirk, *Translation Determined*, p.143.
42 Ibid.
43 Ibid., p.145.
44 Ibid., p.146. Of course, Quine cannot appeal at this point to the claim that translation of the lower-level theoretical sentences is indeterminate: Fred does not accept that the lower-level sentences are underdetermined by possible observations, so the argument from above has to grant him the idea that they have determinate translations.
45 For an attempt to undermine Kirk's objection to the argument from above, see Miller "Meaning Scepticism".
46 Kripke, *Wittgenstein on Rules*, p.56.
47 Note that some philosophers have claimed that Quine is not committed to a behaviouristic psychology. See e.g. Hylton, *Quine*.
48 This is especially important, given the manifest inadequacies of behaviourist philosophies of mind and psychology. For an account of the standard objections to behaviourism, see Carruthers, *Introducing Persons*.
49 Note that I have concentrated here on Quine's views, as presented in only three works: "Two Dogmas", *Word and Object*, and "On the Reasons". Readers should note, though, that this is really only a snapshot of three very important time slices of Quine's work. For Quine's latest views, see his *The Pursuit of Truth*, esp. Chapter 3.

Chapter 5

Scepticism about sense (II)

Kripke's Wittgenstein and the sceptical paradox

In this chapter we move on to look at another form of scepticism about sense, that developed by Kripke's Wittgenstein in Kripke's *Wittgenstein on Rules and Private Language*.[1] Let Jones be a representative speaker of English, and consider sentences such as "Jones means *addition* by '+'", "Jones understands the '+' sign to mean *addition*", and "The sense that Jones associates with the '+' sign is such that it stands for the *addition* function". KW argues for a form of constitutive scepticism about such claims: there is no fact of the matter which constitutes Jones's meaning one thing rather than another by the "+" sign, no fact of the matter which constitutes his attaching one sense rather than another to the "+" sign. Another way of putting the conclusion would be: ascriptions of meaning or sense do not themselves possess sense, they do not have truth-conditions, and are neither true nor false. Ascriptions of meaning thus appear to be meaningless. This is KW's sceptic's "sceptical paradox". KW goes on to try to neutralise the impact of the sceptic's conclusions: even though there are no facts in virtue

of which ascriptions of sense and meaning are true or false, we can still find a place for them by viewing them as possessing *some non-fact-stating* role.[2] This is KW's "sceptical solution". In this chapter we'll outline KW's sceptic's arguments for the sceptical paradox, and KW's sceptical solution to that paradox. We'll see that the sceptical solution is problematic, so that some more direct response to the sceptical paradox is called for. In the next chapter, we'll look at a number of such responses that have been developed by contemporary philosophers of language.

5.1 The sceptical paradox

The conclusion of KW's sceptic's argument is that there is no fact of the matter in virtue of which sentences such as "Jones means *addition* by '+'" or "Smith means *green* by 'green'" are either true or false. The general strategy which the sceptic adopts in arguing for this conclusion is as follows. First, it is argued that if the species of fact in question is to be found, then it must be found within some particular areas. Once this has been done, our knowledge-acquiring powers are imagined to be *idealised* with respect to those areas: we are given *unlimited epistemic access* to the areas in question. Given this, and the ensuing conclusion that even under *these* conditions the sought-after facts still elude our grasp, that any particular claim about the character of these facts still cannot be justified by us, it follows that there simply were *no* such facts there in the first place. For if there were any such facts, given unlimited epistemic access we would surely have found them, and would surely have been able to justify at least some claims concerning their character. But we cannot do this, so, the argument goes, there cannot be any such facts.[3]

KW's sceptic outlines his argument with an example from simple arithmetic, and asks "In virtue of what fact did I mean, in the past, the addition function by my use of the '+' sign?".[4] In order to make the question vivid, he imagines the following example. Suppose that "68 + 57" is a computation that I have never performed before. Since I've performed at most only finitely many computations in the past, we can be sure that such an example

exists (even if you have performed this computation before, just suppose, for the sake of argument, that you haven't: the argument would work just as well for any other computation which you haven't actually performed). Also, the finitude of my previous computations ensures that there is an example where both of the arguments (in this case 68, 57) are larger than any other numbers I've previously dealt with (again, even if this is not the case in the present example, we can easily enough imagine one for which it is the case, and nothing turns on this).

Now suppose that I perform the computation and obtain "125" as my answer. After checking my working out, I can be confident that "125" is the correct answer. It is the correct answer in two senses: first, it is correct in the arithmetical sense, since 125 is indeed, as a matter of arithmetical fact, the sum of 68 and 57; and it is correct in the metalinguistic sense, since the "+" sign really does mean the addition function. (You can imagine how these two senses of correctness might come apart: if the "+" sign really stood for the subtraction function, 125 would still be the sum of 68 and 57, but the correct answer to the question "68 + 57 = ?" would now be "11".) But is my confidence that I have given the correct answer justified? KW's "bizarre sceptic" argues that it is not:

> This sceptic questions my certainty about my answer, in what I just called the "metalinguistic" sense. Perhaps, he suggests, as I used the term "plus" in the past, the answer I intended for "68 + 57" should have been "5"! Of course the sceptic's suggestion is obviously insane. My initial response to such a suggestion might be that the challenger should go back to school and learn to add. Let the challenger, however, continue. After all, he says, if I am now so confident that, as I used the symbol "+", my intention was that "68 + 57" should turn out to denote 125, this cannot be because I explicitly gave myself instructions that 125 is the result of performing the addition in this particular instance. By hypothesis, I did no such thing. But of course the idea is that, in this new instance, I should apply the very same function or rule that I applied so many times in the past. But who is to say what function this was? In the past I gave myself only a finite number of examples instantiating this function.

> All, we have supposed, involved numbers smaller than 57. So perhaps in the past I used "plus" and "+" to denote a function which I will call "quus" and symbolize by "⊕". It is defined by
>
> $$x \oplus y = x + y, \text{ if } x, y < 57$$
> $$= 5 \qquad \text{otherwise.}$$
>
> Who is to say that this is not the function I previously meant by "+"?[5]

KW's sceptic's challenge is thus: cite some fact about yourself which constitutes your meaning *addition* rather than *quaddition* by the "+" sign. Any response to this challenge has to satisfy two conditions. First, it has to provide us with an account of the type of fact which is constitutive of the meaning of "+". Second, it has to be possible to *read off* from this fact what constitutes *correct* and *incorrect* use of the "+" sign. It must show why the answer "125" to the problem "68 + 57 =?" is *justified* or is the answer I *ought* to give.

The sceptic is thus challenging us to provide an acceptable answer to the question: In virtue of what fact are you now justified in answering "125" to the query "What is 68 + 57?"? In accordance with the general strategy outlined above, the sceptic begins his argument by allowing us unlimited epistemic access to two areas, and inviting us to find a suitable meaning-constituting fact from within either of those two areas. The areas in question are *(a) our previous behaviour, linguistic and non-linguistic; and (b) the entire contents of our previous mental histories*.[6]

Nothing from the finite pool of my previous behaviour will do, since ex hypothesi I have never dealt with numbers larger than 57, and "+" and "⊕" have the same extensions for numbers smaller than 57. Anything which is a "correct" answer to "x + y = ?" will also be a "correct" answer to "x ⊕ y =?" so long as x, y < 57. Enlarging the pool of previous behaviour won't make a difference, since no matter how it is enlarged, a "deviant" interpretation of "+", such as that which takes it as standing for the quaddition function, will always be possible: even if we enlarge the pool of previous behaviour so that we have encountered numbers larger than 57, there will always be some number which is larger than those we have

previously encountered, and the sceptic can use this to construct an analogue of the quaddition interpretation.

At this point, KW's sceptic imagines the following protest:

> [O]ur problem arises only because of a ridiculous model of the instruction I gave myself regarding "addition". Surely I did not merely give myself some finite number of examples from which I am supposed to extrapolate the whole [function] . . . Rather I learned – and internalized instructions for – a *rule* which determines how addition is to be continued. What was the rule? Well, say, to take it in its most primitive form: suppose we wish to add x and y. First count out x marbles in one heap. Then count out y marbles in another. Put the two heaps together and count out the number of marbles in the union thus formed.[7]

Isn't it this fact about what I had previously learnt and internalised that constitutes the fact that I meant *addition* and not *quaddition*? Alas, the sceptic has a response to this suggestion:

> True, if "count", as I used the word in the past, referred to the act of counting (and my other past words are correctly interpreted in the standard way), then "plus" must have stood for *addition*. But I applied "count", like "plus", to only finitely many past cases. Thus the sceptic can question my present interpretation of my past usage of "count" as he did with "plus". In particular, he can claim that by "count" I formerly meant quount, where to "quount" a heap is to count it in the ordinary sense, unless the heap was formed as the union of two heaps, one of which has 57 or more items, in which case one must automatically give the answer "5".[8]

Thus, if I follow the "counting" rule properly, I really ought to answer "5" when asked "68 + 57 =?". We are back to where we started: citing something like a *general thought or instruction* in response to the sceptical challenge won't work, because the sceptic can always respond by giving a deviant interpretation of the symbols of the general thought or instruction itself. And the point can be generalised: any set of instructions that come before the mind require interpretation as much as the linguistic expression whose

understanding they are supposed to facilitate, and are thus as susceptible to deviant interpretation as that original expression. Clearly, invoking instructions for interpreting the instructions will send us off on a fruitless infinite regress. Kripke takes himself here to be expounding Wittgenstein's remarks in *Philosophical Investigations* on "a rule for interpreting a rule":

> [A]ny interpretation still hangs in the air along with what it interprets, and cannot give it any support. Interpretations by themselves do not determine meaning.[9]

We'll return to this in due course.

It thus looks as if no fact about my previous behaviour will do the trick. Does the search within our *mental histories* fare any better? Does the possession of some *mental image*, or some other specific mental item possessed of its own distinctive qualitative character provide what we are after? KW's sceptic argues that it does not.

Firstly, it is not a *necessary* condition for understanding that some particular item come before one's mind when one hears or uses a given expression. As a matter of empirical fact, it seems to be the case that no one "mental entity" comes before one's mind when one correctly understands a linguistic expression. And even in cases where there does seem to be an empirical regularity between a particular expression and a particular such item we can still perfectly well *conceive* of someone understanding the expression in the absence of that item (indeed, we can conceive of someone understanding the expression even when *no* such item is present at all). KW's sceptic's conclusion is that it is no more necessary for understanding an expression that I have an inner mental picture before my mind than it is that I have a concrete physical picture ready to hand – just as I can perfectly well understand "cube" without having a drawing of a cube in front of me on the table, so I can understand it without having to call up a mental image.

Secondly, neither is it a *sufficient* condition for meaning a sign in a particular way that some item, be it a picture or otherwise, come before one's mind. *The essential point is that the picture does not by itself determine the correct use of the associated word,*

because the picture thus associated is really just another sign whose meaning also requires to be fixed. There is no logical route from the properties of an image or picture to the meaning of an associated word, because of the possibility of deviant applications of the word consistent with some interpretation of the picture; the relation of any picture or image to the associated word can be construed in such a way that *any* future pattern of use of the expression can count as correct. Whatever comes before the mind can be made to accord with a deviant application of the expression. Thus, mental images set no standards for the correct use of an expression: one cannot "read off" from a mental image what counts as the correct use of an associated expression. In order to drive this point home, we can imagine someone at the hands of an omnipotent but curious experimenter, conditioned by the experimenter to have the image of a cube every time he hears the word "cube"; but again, this will not be sufficient to determine what he understands by the word in question. We might of course admit that images *naturally suggest* certain applications, but the important point is rather that they do not logically determine them. It is possible that two people might understand a word differently – they might go on to use it differently – even though the same images occurred to them on hearing it.[10]

Perhaps the fact that the subject means addition by "+" consists in the fact that when he uses or hears the expression he has an irreducible experience, with its own distinctive quale, known by introspection? On this sort of view "the experience of *meaning addition* has its own irreducible quality, as does that of feeling a headache. The fact that I mean addition by 'plus' is to be identified with my possession of an experience of this quality".[11] Arguably, this special qualitative experience is a philosophers' fiction: think of your uses of "+" in the past and ask yourself seriously whether they were accompanied by a characteristic sort of feeling (can you describe it?). KW argues that even if there were such an experience it would fail to satisfy the second of the two conditions which candidate answers to the sceptic have to meet:

> [S]uppose I do in fact feel a headache with a very special quality whenever I think of the "+" sign. How on earth would this

> headache help me figure out whether I ought to answer "125" or "5" when asked about "68 + 57"? If I think the headache indicates that I ought to say "125", would there be anything about it to refute a sceptic's contention that, on the contrary, it indicates that I should say "5"?[12]

KW's sceptic considers a further reply, which, drawing on the subject's past mental life, might be adduced in response to the sceptical argument. This is that in the paragraphs above we have concentrated too exclusively on what might be called "quotidian" states of the mind – states such as mental images, sensations, headaches, and other introspectible mental states with their own distinctive phenomenological character. Perhaps meaning addition is not a state of these kinds, but rather an irreducible state, "a primitive state . . . a state of a unique kind of its own".[13] This response, however, is castigated by the sceptic on the grounds that it is "desperate", and leaves "completely mysterious" the character of the primitive state claimed to constitute understanding. For one thing, it is not an introspectible state, and yet its possessor is thought to be aware of it with a high degree of certainty whenever he has it. How could this be possible? It is also completely mysterious, moreover, how "a finite object contained in our finite minds" could be such as to reach out to a potentially infinite number of future uses of an expression and determine whether or not they are correct in the light of that expression's meaning. It is a mystery what the relationship is between these future uses and the putative primitive state; the response as it stands gives no clue as to what the nature of this relationship is, or how it is forged.

KW's sceptic now considers a suggestion which at first seems to issue in a plausible refutation of his sceptical argument. The suggestion is that in limiting us to facts concerning our *past* actual behaviour and our previous *occurrent* mental states, KW has already guaranteed himself success: rather, we should consider *dispositional* facts about language users. These will enable us to distinguish between the hypothesis that I meant *addition* in the past, and the hypothesis that I meant *quaddition* in the past. The claim that I meant *addition* would be true if I was disposed in the past, when asked to compute "x + y", to produce the *sum* of

the two numbers. Similarly, the claim that I meant *quaddition* would be true if I was disposed, when faced with the same query, to respond with the result of *quadding* the two numbers. Thus, were my knowledge-acquiring powers concerning my dispositional properties sufficiently idealised, I would no doubt see that I was disposed to give the answer "125" to the query, even though in fact I never did because I was never asked to. KW's sceptic's claim that I meant some function other than addition would thus be refuted.

KW's sceptic, however, has two powerful objections to the dispositionalist account. The first objection is that the dispositionalist account completely fails to take account of the *normativity* of meaning – in effect, it fails to satisfy the second of the two conditions laid down earlier in this section on any candidate fact for meaning addition. We say that a competent language speaker *ought*, given his previous meanings and his intentions to remain faithful to them, to respond to, for example, arithmetical questions in certain determinate ways, and we believe that the response he ought to give is logically independent of the response that he *did* give, or *would have given*, had he actually been faced with the query. But the dispositionalist account leaves no room for such a distinction between the answer he ought to have given and the answer he would have given, for according to the dispositional account the answer he ought to have given simply collapses into the answer that he would have given. In a nutshell, the dispositionalist response appears to involve the intuitively unacceptable equation of *competence* and *performance*. Kripke sums the problem up as follows:

> Suppose I do mean addition by "+". What is the relation of this supposition to the question how I will respond to the problem "68 + 57"? The dispositionalist gives a *descriptive* account of this relation: if "+" meant addition, then I will answer "125". But this is not the proper account of the relation, which is *normative*, not descriptive. The point is *not* that, if I meant addition by "+", I *will* answer "125", but rather that, if I intend to accord with my past meaning of "+", I *should* answer "125". Computational error, finiteness of my capacity, and other disturbing factors may lead me not to be *disposed* to respond as I *should*,

> but if so, I have not acted in accordance with my intentions. The relation of meaning and intention to future action is *normative*, not *descriptive*.[14]

This problem shows up further in the fact that intuitively we want to be able to leave room for the possibility that someone is *systematically mistaken*, in the sense that he is disposed to make mistakes. For example, someone might be disposed to systematically miscarry when carrying out addition problems, and we want to leave room for the possibility that such a person means addition but is giving answers out of line with those that he ought to give: we do not want to be committed to the conclusion that he in fact means some different arithmetical function and is after all counting correctly.[15]

The second problem for the dispositional suggestion is that dispositions are, like the totality of our previous linguistic behaviour, finite. For example, it simply is not true that I am disposed in such a way that I will always give the sum of two numbers when faced with "What is x + y?": some numerals will simply be too long for me to handle and some will be so long that I will die before having an opportunity to respond to the query. Given this, it is easy for KW's sceptic to construct unwelcome interpretations that are nevertheless compatible with all of the dispositional facts concerning me – for example, perhaps I meant *skaddition*, where this is defined as:

$$x * y = x + y, \text{ if } x, y \text{ are small enough for me to handle,}$$
$$= 5, \text{ otherwise.}$$

I would then be unable to cite any fact about my dispositions that could constitute my meaning addition and which would be incompatible with the hypothesis that I meant skaddition. Dispositions are thus unable to fix the meaning of the "+" sign.

So none of the facts we have considered – facts about our previous behaviour, facts about our behavioural dispositions, facts

about general thoughts or instructions, facts about "quotidian" mental states such as mental images, facts about *sui generis* and irreducible states of meaning and understanding – appear to be plausible candidates for constituting the fact that we mean *addition* rather than *quaddition* by "+". We appear to be facing the conclusion that there is no fact of the matter as to what we meant in the past by "+", and since our present understanding of "+" will be up for retrospective viewing in the future, it follows that there is no fact of the matter as to what we mean at present by "+" either. And, since the sceptical argument could, without any loss, be rerun for any other speaker and any other linguistic expression, it follows that there are no facts of the matter as to what anyone means by any expression.[16] The notion of meaning apparently has, as Kripke puts it, "vanished into thin air".[17]

Before continuing, it is perhaps worthwhile to pause to note how KW's sceptic's argument relates to some of the Fregean theses about sense that we outlined in Chapter 2. Recall that according to thesis 8, the sense of an expression determines its semantic value. In the case of an incomplete sign such as ". . .+. . .", this means that its sense has to determine which function it stands for (or, alternatively, its extension). KW's sceptic's argument in effect proceeds as follows: no fact could constitute Jones's attaching a particular sense to "+" because all of the possible facts that we could cite as constituting Jones's grasping a sense are themselves compatible with "+" standing for different functions (or having incompatible extensions). For example, facts about Jones's dispositions are compatible with "+", as he understands it, standing for either the addition function or the quaddition function. So, no facts about Jones's dispositions – nor any of the other sorts of fact considered – determine the extension of "+" as he understands it. So since the sense of "+" determines its extension, none of these facts can constitute Jones's grasping its sense. So, given the assumption about idealised epistemic access, there is no fact of the matter as to the sense of "+". Note also that some of the other Fregean theses play a role in the argument: for example, thesis 20 is what lies behind the normativity objection to dispositionalism. KW's sceptic can perhaps be viewed as arguing that it is impossible to find any fact which could satisfy all of the main Fregean theses concerning sense: even if

dispositional facts could determine the extension of ". . .+. . .", they would still fall down on the normativity requirement.[18]

5.2 The sceptical solution and the argument against solitary language

Kripke distinguishes between two ways in which we might respond to KW's sceptic's argument. One way he describes as a "straight solution". This would consist of a demonstration of the thesis that the sceptic called into question: the production of a suitable meaning-constituting fact of the sort the sceptic questions. On the other hand there is what Kripke describes as a "sceptical solution". Such a solution would consist of two parts. First, an admission that the sort of fact questioned by the sceptic is in fact non-existent. Secondly, in a more positive vein, an argument to the effect that the area of discourse does not have to be viewed as fact-stating in order for it to be legitimate. In other words, in the special case of, for example, meaning, it would be argued that our practice involving sentences which ascribe meaning and understanding does not require for its tenability the sort of justification which the sceptic demands, namely, an account of the facts which would render them true or false. We can justify our practice of ascribing meaning and understanding in other terms, without viewing ascriptions of meaning as having truth-conditions. It is this latter sort of solution, a sceptical solution, which Kripke sees Wittgenstein himself advocating in *Philosophical Investigations*.[19]

KW thus admits, at the outset of the sceptical solution, that sentences ascribing meaning do not have truth-conditions, that there are no facts or states of affairs in virtue of which ascriptions of meaning are true or false. As textual support for his claim that Wittgenstein rejected the idea that sentences had to be viewed as possessing truth-conditions in order for our practices with them to be deemed legitimate, Kripke cites Michael Dummett's claim that the emphasis in Wittgenstein's early work, *Tractatus Logico-Philosophicus*, upon facts and truth-conditions was replaced in the later *Philosophical Investigations* by a completely different emphasis.[20] In the later work, the emphasis is on describing the conditions

under which sentences are deemed to be *justified* or *assertable*, and on the *role or utility* that so deeming them has within our lives. As far as meaning goes, then, the change in emphasis which Kripke sees Wittgenstein as recommending is this: we are enjoined not to look for "entities" or "facts" corresponding to sentences ascribing meaning, but rather to look at the circumstances under which such ascriptions are actually made, and the utility that resides in making them under these conditions. If these conditions and the corresponding role and utility can be specified adequately, then, KW suggests, we will have provided a sceptical solution to the sceptical paradox raised by his sceptic.

What are the relevant conditions, utility, and role? Consider the sentence "Jones means *addition* by '+'". If Jones is considered in isolation from any linguistic community – as a speaker of a solitary language – then the conditions under which this will be asserted will correspond to those under which Jones himself would assert "I mean addition by '+'". There will thus be no distinction between the assertion conditions of "Jones believes that he means *addition* by '+'" and "Jones means *addition* by '+'". Thus, whatever seems right to Jones will be right, "and that only means that here we can't talk about 'right'".[21] Thus, in the case of an individual considered in isolation, the sceptical solution will be powerless to salvage any point to our practices with the notion of meaning, since any such reconstruction must respect the fact that the meaning an individual speaker attaches to an expression is normative with respect to his inclinations to apply it in certain ways. Kripke takes this to be the upshot of Wittgenstein's famous claim that "to think one is obeying a rule is not to obey a rule. Hence, it is not possible to obey a rule 'privately': otherwise thinking one was obeying a rule would be the same thing as obeying it".[22]

Matters stand differently when Jones is considered, not as a solitary individual, but as a member of a linguistic community. The utterance of "Jones means addition by '+'" is then considered to be justified when Jones has performed satisfactorily often enough with "+", where this is taken to involve nothing more than that Jones has satisfactorily often enough come up with the answer that most of the rest of his fellow speakers in the community are disposed to give. The utterance of the sentence thus marks the

community's acceptance of Jones into its midst, and it marks also the community's conviction that Jones can generally be trusted to act as they do in transactions which involve the use of the "+" sign. The utility of uttering ascriptions of meaning under such conditions is clear. They allow us to discriminate between those people we can trust in our transactions involving "+", and those we cannot. More generally, they allow us to discriminate between people who are members of our general linguistic community and those who are not. A grocer who "means addition by '+'" is one who can be trusted to treat me as I expect when I go in to his shop to buy five apples. And note, crucially, that in the communal setting there is a distinction between the conditions under which it is assertable that Jones believes that he means addition, and the conditions under which it is assertable that he actually does mean this: it might indeed be the case that Jones believes that he means addition, but at the same time the rest of the community finds that his use of the "+" sign is out of step with theirs, so that the assertion conditions for "Jones means addition by '+'" are not in fact met.

To sum up then, the sceptical solution admits that discourse involving meaning is not fact-stating, but attempts to legitimise it by finding it an appropriate non-fact-stating role to play. This involves spelling out the assertability conditions of ascriptions of meaning, and showing that asserting them under these conditions plays a useful role in our lives. It emerges that these assertion conditions involve an essential reference to a linguistic community, since the community underwrites the "seems right/is right" distinction essential to any conception of meaning. The preservation of a legitimate role for ascriptions of meaning is thus claimed by KW to deliver a demonstration of the impossibility of a solitary language.[23]

5.3 Boghossian's argument against the sceptical solution

Two sorts of general criticism can be levelled against KW's "sceptical solution". First, that the sceptical solution does not succeed in establishing the impossibility of solitary language. Secondly, that such a sceptical solution about meaning is *internally incoherent*.

We shall concern ourselves here only with criticisms of the second, more general, type.[24] We'll first look at an argument developed by Paul Boghossian, which purports to establish that the sceptical solution is incoherent. We'll argue that Boghossian's argument is unsuccessful. In §5.4, we'll look at an objection to the sceptical solution that has been raised by Crispin Wright, and suggest that it is at best inconclusive. In §5.5, we'll outline Jose Zalabardo's (apparently successful) argument against the sceptical solution.

Boghossian's objection forms part of a general line of attack against what he terms *irrealist* conceptions of content:[25] against both *error-conceptions* of content – which claim that ascriptions of meaning possess genuine truth-conditions but are nevertheless *systematically false* – and against *non-factualist* conceptions of content, according to which ascriptions of meaning do not even possess genuine truth-conditions.[26] KW's sceptical solution falls into the latter, non-factualist, category of irrealist conception, so we'll concentrate here on Boghossian's argument against non-factualist accounts of meaning.

Boghossian notes that it is constitutive of non-factualism about a given discourse that it denies that significant declarative sentences from that discourse have truth-conditions. He also notes that on a *deflationary conception of truth*, for a sentence to possess truth-conditions it is sufficient that it be disciplined by norms of correct usage and that it possess the syntax distinctive of declarative sentences. It would seem to follow, therefore, that anyone wishing to advance a non-factualist thesis about a given region of discourse would have to operate with a conception of truth richer than the deflationary: he would have to operate with what Boghossian terms a *robust conception of truth*, a conception "committed to holding that the predicate 'true' stands for some sort of language-independent property, eligibility for which will not be certified purely by the fact that a sentence is significant and declarative".[27] Sentences such as "Jones means addition by '+'" are subject to norms of discipline: ordinarily, for example, the fact that Jones has generally replied with the sum would be taken to render an utterance of it to be correct, while this would not be the case if he gave apparently random answers to queries involving "+". They also have the syntax distinctive of declaratives in that their negations are syntactically

well-formed ("It is not the case that Jones means addition by '+'") and they can figure in the antecedents of conditionals ("If Jones means addition by '+' then everything will be OK"). So according to the deflationary conception they must have truth-conditions. But non-factualism about meaning claims that they do not have truth-conditions. So non-factualism about meaning presupposes that the deflationary conception of truth is false: non-factualism presupposes, in the sense above, a robust conception of truth.

Boghossian now attempts to show that non-factualism about meaning also entails that truth is *not* robust. It would then follow that non-factualism about meaning presupposes both that truth is robust and also that truth is not robust, and this would demonstrate the incoherence of the non-factualist position. Boghossian notes that non-factualism consists in the following pair of claims. For any S, p:

(1) There is no property corresponding to the predicate "has truth-condition p",

and

(2) "S has truth-condition p" is not truth-conditional.[28]

Boghossian now argues that it follows from (1) that

(3) There is no property corresponding to the predicate "is true".

Why does this follow? Boghossian argues as follows. First of all, whether or not a sentence is true is a function of two things: the truth-condition it possesses, and the way the world is. For example, the fact that the sentence "Grass is green" is true is a function of the fact that it has the truth-condition

(4) "Grass is green" is true iff grass is green

and of the fact that grass is, as a matter of fact, green. Secondly, if one of the things which determines whether a is F is non-factual, then it will follow that there is no fact of the matter as to

whether a is F either. This can be illustrated by using an example of Crispin Wright's:

> If among the determinants of whether it is worth while going to see a certain exhibition is how well presented the leading exhibits are, then, if questions of good presentation are not considered to be entirely factual, neither is the matter of whether it is worth while going to see the exhibition.[29]

If whether an exhibition (a) is worth while going to see (F) is determined in part by whether the leading exhibits are well presented, and if there is no fact of the matter as to whether the leading exhibits are well presented, then there will be no fact of the matter as to whether the exhibition is worth while going to see (i.e. no fact of the matter as to whether a is F). Applying this line of argument to the case of truth and truth-conditions: since the truth-value of a sentence is a function of its truth-condition, if there is no fact of the matter as to which truth-condition it possesses, it will follow that there is no fact of the matter as to its truth-value either. In other words, it follows from non-factualism about meaning that, for any S,

(5) "S is true" is not truth-conditional.

But this is just to say that there is no property, truth, in virtue of the instantiation (or non-instantiation) of which sentences are true (or false). In other words, it follows that

(6) There is no property corresponding to the predicate "is true".

But if there is no property corresponding to the predicate "is true", then *a fortiori* there is no language-independent property corresponding to the predicate "is true". In other words, truth cannot be robust.

So the difficulty for non-factualism about meaning is clear: it requires, via (2), the presupposition that truth is robust. But at the same time it entails, in the manner outlined above, that truth cannot be robust. Boghossian suggests that this constitutes a *reductio ad absurdum* of non-factualism about meaning.

We'll now attempt to develop a response to Boghossian's argument. This response attempts to show that the non-factualist about meaning can formulate his thesis in such a way that it does not entail the contradiction which Boghossian focuses on.

Jane Heal, in her book *Fact and Meaning: Quine and Wittgenstein on the Philosophy of Language*, has suggested two conditions which a region of discourse has to satisfy in order for it to have a factual subject matter. First, the sentences of that region of discourse must satisfy the *law of non-contradiction*: it must not be possible for two incompatible sentences within that region to be equally fully acceptable. Second, the relevant sentences must satisfy a condition of *minimal doxastic independence*: the fact that a given sentence is true must not follow simply from the fact that some speaker believes it to be true. According to Heal, a demonstration that a region of discourse failed to satisfy either of these conditions would be sufficient to establish non-factualism about the sentences of that discourse.

Interestingly, KW's sceptic's argument does appear to take the form of a challenge to the claim that the law of non-contradiction is satisfied by sentences purporting to ascribe meaning. KW's sceptic's claim is, after all, that there is no way of showing that the hypothesis that you mean *addition* is more acceptable than the incompatible hypothesis that you mean *quaddition* (or vice versa). Given Heal's necessary conditions on factuality and KW's sceptic's arguments, it follows straightforwardly that sentences ascribing meaning are non-factual.

Is this way of viewing the route to non-factualism about meaning susceptible to Boghossian's argument? It seems not: it now appears that a non-factualist about meaning can state his thesis without presupposing any commitment to a robust notion of truth. All the non-factualist has to argue is that the significant, declarative sentences which constitute putative ascriptions of meaning do not satisfy the law of non-contradiction in Heal's sense. They are therefore non-factual. Now where in this is the commitment to the view that truth is robust? *In arguing thus the non-factualist does not seem to have committed himself to the view that truth is some language-independent property*. Or at least, it requires some more substantial argument to show that he is so committed: at a minimum, Boghossian requires a substantial argument to show that a commitment to robust truth – in the form of a commitment

to a view of truth as a language-independent property – emerges in any account of the rationale for taking non-satisfaction of the law of non-contradiction to entail non-factuality.

To put the point slightly differently, we can now represent KW's sceptic as arguing in the following way:

(7) A discourse must satisfy Heal's law of non-contradiction in order to be factual

(8) Discourse about meaning does not satisfy Heal's law of non-contradiction (the claim that you mean *addition* by "+" can no more be justified than the claim that you mean *quaddition* by "+")

So

(9) Discourse about meaning is non-factual.

Where does the commitment to a view of "true" as standing for a language-independent property surface in this formulation of KW's sceptic's argument? (8) is just a statement of the conclusion of the sceptical argument, which we are supposed to be granting in order to evaluate his sceptical solution, so the only route open to Boghossian would seem to be to argue that the commitment to robust truth is contained implicitly in (7). And there is our unanswered question: why should laying down satisfaction of the law of non-contradiction as a necessary condition on factuality harbour a commitment to a view of truth as a language-independent property? The onus is on Boghossian to answer this question.

We can thus tentatively suggest that KW does have a way of stating his non-factualist thesis about meaning in a manner which avoids the outright contradiction described by Boghossian. The internal incoherence of the sceptical solution has thus not yet been established.

5.4 Wright's objections to the sceptical solution

Perhaps the argument advanced by Boghossian, although not ultimately successful as it stands, does contain, in one of its steps, the resources for another argument which does undermine the sceptical solution. This is the argument, originally given by Wright, to

the effect that if meaning is a non-factual matter, it follows that the truth of sentences is likewise a non-factual matter: in other words, it follows that no sentences whatsoever have a factual subject matter. Now this is quite striking: non-factualism about meaning entails global non-factualism, non-factualism about all regions of discourse. Doesn't this show that there must be something wrong with non-factualism about meaning?

There are (at least) two considerations which seem to suggest an affirmative answer to this question. Firstly, if all sentences are non-factual, this includes the conclusion of the sceptical argument itself together with the sentences which comprise KW's sceptical solution. Then the worry, as expressed by Wright, is that "a statement of the sceptical argument, for instance, is not itself to be projective".[30] Does this pose a genuine problem for KW? It is difficult to avoid agreeing with Boghossian's point that it is at least not transparent that there is a genuine difficulty here:

> A global [non-factualism] would have to admit that it is no more than assertible that no sentence possess a truth condition. But what is wrong with that? If there is an instability here, it is not a transparent one.[31]

Why cannot KW get by with the claim that, although the conclusion of the sceptical argument is itself lacking in truth-conditions, it is nevertheless assertable under certain conditions, and that its assertion has some point or role to play in our lives? At the very least, more substantial argument is required to show that this sort of story could not be sufficient for KW's purposes.

There is, however, perhaps a different line of argument lurking in Wright's objection. Surely the claim that sentences ascribing meaning are non-factual is intended, by KW, to show that such sentences *compare unfavourably* with some other body of sentences, those that are genuinely factual and possess genuine truth-conditions. In other words, KW's argument depends on the existence of a *distinction* between factual and non-factual sentences, and it is then supposed to emerge as a substantial conclusion that ascriptions of meaning fall on the non-factual side of this divide. The suggestion would then be that if, as suggested by Wright's argument above, it follows that *no* sentences are factual, it is

difficult to see what distinction there could be between factual and non-factual sentences: KW, it would be suggested, owes us an account of what this distinction consists in, and the objection is that given the consequences of his sceptical conclusion, he will be unable to do so. The sceptical conclusion and sceptical solution depend on a distinction which they themselves seem to obliterate.

Is this line of argument conclusive? This is not clear. For one thing, someone claiming that there are no square circles can hardly be criticised if he is unable to give a substantial account of what it would be for a circle to be square. Why should it be any different for a sceptic who claims that there are no factual discourses? To be sure, in the case of square circles we can say that something would be a square circle if it satisfied a certain set of mathematical equations, which turn out to be *a priori* unsatisfiable. But likewise, in the case of the sceptic, we can say what it is for the sentences of a discourse to have truth-conditions – for example, satisfying Heal's two conditions (§5.3) on factuality – even though these turn out as an *a priori* matter never to be jointly satisfiable. Secondly, and perhaps more importantly, it seems in any case that the onus is not on KW to give substance to the factual/non-factual distinction. All he has to do is take *our* favoured conception of factuality, and show that on that conception, ascriptions of meaning fail to be factual: if it follows from this that the notion of factuality in play has no instances, KW can simply reply – can't he? – that this is so much the worse for us, and that his argument not only shows that there are no facts about meaning, it goes further and shows that our notion of factual discourse has no instances. So it is by no means clear that the consequence of the sceptical solution Wright focuses on constitutes a difficulty for KW, rather than just an additional difficulty for someone concerned to avoid the conclusion of the sceptical argument. At best, Wright's objection to the sceptical solution appears to be inconclusive.

5.5 Zalabardo's objection to the sceptical solution

The final objection against the sceptical solution that we shall consider is perhaps more damaging. Curiously enough, it is suggested by an objection which KW's sceptic himself raises in the course of

giving his argument for the sceptical paradox. At one stage KW's sceptic considers a reply to the sceptical argument which has it that a choice between the incompatible, but apparently equally acceptable, hypotheses that it was *addition* that was meant and that it was *quaddition* that was meant, could be made by some kind of appeal to the *simplicity* of the respective hypotheses: perhaps it is the simplest function that should be deemed to have been meant. KW's sceptic, however, is scornful of this suggestion, but not merely because "simplicity is relative, or that it is hard to define, or that a Martian might find the quus function simpler than the plus function".[32] These are indeed problems, but KW's sceptic suggests that the difficulty with the suggestion is in fact more basic, and that it relies on a misunderstanding of the nature of the conclusion of the sceptical argument. The conclusion proper of the sceptical argument was not that there are two competing genuine hypotheses about what was meant: that conclusion was only supposed to be a "dramatic device", designed to facilitate the statement of the sceptical argument. The real conclusion of the sceptical argument was that there is no fact of the matter as to whether a determinate meaning attached to a given expression. Keeping the real nature of the sceptical conclusion in mind then helps us to see that appeal to simplicity considerations won't give us a reply to the sceptic. As Kripke puts it:

> [S]implicity considerations can help us to decide between competing hypotheses, but they can never tell us what the competing hypotheses are. If we do not understand what two hypotheses *state*, what does it mean to say that one is "more probable" because it is "simpler"? If two competing hypothesis are not genuine hypotheses, not assertions of genuine matters of fact, no "simplicity" considerations will make them so.[33]

In other words, if two ascriptions of meaning do not have truth-conditions, what does it mean to say that one of them is more probably true because it is simpler?

What we shall now suggest is that the sceptical solution offered by KW itself falls prey to a similar misunderstanding to that which apparently vitiates the appeal to simplicity considerations.[34]

The objection has been made forcibly by Jose Zalabardo (though he fails to notice any similarity between his own objection and the one which KW's sceptic makes against the simplicity suggestion). Zalabardo too points out that "the result does not concern an indeterminacy as to which rule is being followed, but the notion of rule itself".[35] If the problem were merely one of choosing between alternatives, then facts about agreement of communal responses, in the manner described by KW, might indeed constitute a solution to that problem. But given that this is not the real sceptical problem, the crucial question is whether "Kripke's description of the assertibility conditions of content could provide the meaning of the ascriptions".[36] And it would appear that the answer to this question has to be negative, for three reasons. Firstly, if ascriptions of meaning amount to no more than claims about agreement of responses within the linguistic community, it would seem to follow that "all judgements would collapse into judgements about the ascribers (or the whole community's) inclinations to consider certain situations as in some respect similar to each other".[37] In other words, and absurdly, that all our judgements would be about our fellow speakers' inclinations and the agreement or disagreement among them. Secondly, this would really only be another attempt at a straight solution, an attempt to spell out the truth-conditions of ascriptions of meaning in terms of communal responses, and the distinctive nature of the sceptical solution would be lost. Thirdly, this suggestion would itself be susceptible to KW's sceptical argument. This point has been well made by Crispin Wright:

> Could it yesterday have been true of a single individual that he associated with the sentence "Jones means addition by '+'" the sort of assertion conditions Kripke sketches? Well, if so, that truth did not consist in any aspect of his finite use of that sentence or its constituents; and, just as before, it would seem that his previous thoughts about that sentence and its use will suffice to constrain within uniqueness the proper interpretation of the assertion conditions he associated with it only if he is granted correct recall of the content of those thoughts – exactly what the sceptical argument does not grant. But would not any truths concerning assertion conditions previously associated by

> somebody with a particular sentence have to be constituted by aspects of his erstwhile behaviour and mental life? So the case appears no weaker than in the sceptical argument proper for the conclusion that there *are* no such truths; whence, following the same routine, it speedily follows that there are no truths about the assertion conditions that any of us presently associates with a particular sentence, nor, *a fortiori*, any truths about a communal association.[38]

It seems that the assertability conditions used in the sceptical solution must not be viewed as constituting the truth-conditions of ascriptions of meaning. But if KW makes no attempt to provide an account of the truth-conditions of ascriptions of meaning, then it is utterly unclear what role is being performed by the facts about communal inclinations: we cannot view them as selecting from the candidate ascriptions of meaning on the basis of their agreement or disagreement with the truth-conditions of those ascriptions, so on what basis do the facts about communal inclinations select favoured options from among the possible ascriptions of meaning? In other words, if two ascriptions of meaning don't have truth-conditions, what does it mean to say that one of them agrees more closely than the other with the linguistic inclinations of the community?

This, then, would appear to be the most powerful objection so far to KW's sceptical solution. The sceptical solution too is mistaken in taking the epistemic sceptical problem – the problem of knowing which ascription of meaning correctly describes a given speaker – to be more than a "merely dramatic device", used only *en route* to a more damaging and more pervasive constitutive sceptical conclusion: that ascriptions of meaning themselves possess no truth-conditions.

In conclusion, we can say that although the objections discussed in §5.3 and §5.4 do not conclusively establish the implausibility of KW's sceptical solution, the objection discussed in the present section does suggest that the sceptical solution is a non-starter. This means that we will have to think seriously about finding a *straight* solution to KW's sceptical paradox: an account of the fact which constitutes a speaker's meaning one thing rather than another by an expression. We look at a number of attempts at such solutions in the next chapter.[39]

5.6 The normativity of meaning?

In §5.1 we saw that KW's sceptic attempts to use the "normativity of meaning" to block a dispositionalist response to his sceptical paradox. In this section, we'll probe a little deeper into whether meaning is normative in any sense that might justify irrealism about meaning and content. We can do this by considering very briefly a well-known "normativity" argument in metaethics that is directed against moral realism, and then asking whether it can be transposed plausibly to the case of meaning.

In order to do this, we can define *semantic realism* as the view that ascriptions of meaning to linguistic expressions (e.g. "+" means addition), or linguistic understanding to speakers (Jones understands "+" to mean addition), are apt to be assessed in terms of truth and falsity, and are, in many instances, true. So defined, semantic realism is a form of *cognitivism* about meaning, holding that judgements ascribing linguistic meaning or understanding express *beliefs*, states apt for assessment in terms of truth and falsity.

One argument used in metaethics against ethical cognitivism, the view that moral judgements express beliefs, is *the argument from motivational internalism*. One way to set this up is to start with the idea that the following three propositions generate a contradiction:

(1) The judgement that doing X in C is morally right expresses a belief (cognitivism).
(2) It is a conceptual truth that: if an agent S judges that doing X in C is morally right, then so long as S is practically rational, S will be motivated to X in C (motivational internalism).[40]
(3) Motivation is always a matter of having both beliefs and desires, where beliefs and desires are "distinct existences" in the sense that there is no conceptual or necessary connection between any particular belief and any particular desire (the Humean theory of motivation).[41]

and then to jettison (1), ethical cognitivism, on the basis of (2) and (3).[42]

Why think that (1), (2), and (3) generate a contradiction? Take a particular moral agent Smith, and suppose that he is "practically rational", in other words not weak-willed, depressed, or prone to any other kind of malady known to sap one's motivation to act in accord with one's moral judgements. Suppose also that Smith judges that vegetarianism is morally right. By (1), this expresses the belief that vegetarianism is morally right. By (2), since Smith is ex hypothesi practically rational, it follows as a matter of conceptual necessity that Jones is motivated not to eat meat. By (3), this means that it follows as a matter of conceptual necessity that Smith has a desire not to eat meat. So, there is a conceptual and necessary connection between Smith's belief that vegetarianism is morally right and his desire not to eat meat. But this is precisely the sort of connection between a particular belief and a particular desire outlawed by the assertion in (3) that beliefs and desires are distinct existences.

Thus, defenders of the Humean theory of motivation (3) and motivational internalism about ethical judgement (2) can reject ethical cognitivism (1). Is there any scope for a parallel argument against cognitivism about meaning? In giving an exposition of the claim that meaning is normative, Boghossian appears to suggest that there is at least prima facie scope for such an argument. Boghossian writes:

> To be told that "horse" means *horse* implies that a speaker ought to be motivated to apply the expression only to horses[43]

and also

> [A dispositional view of meaning] must show how a [meaning-constituting] disposition could amount to something called a *correctness* condition, *something we would be inherently motivated to satisfy*.[44]

Boghossian's talk of "inherent motivation" and so on suggests that he sees KW's claim that meaning is normative as involving

the following kind of motivational internalism about semantic judgement: it is a conceptual truth that if an agent S judges that "magpie" means *magpie*, then so long as S is practically rational, S will be motivated to apply "magpie" to an object only if that object is a magpie. This would generate the following argument against cognitivism about meaning:

(4) The judgement that "magpie" means *magpie* expresses a belief.
(5) It is a conceptual truth that: if an agent S judges that "magpie" means *magpie*, then so long as S is practically rational, S will be motivated to apply "magpie" to an object only if that object is a magpie.
(6) Motivation is always a matter of having both beliefs and desires, where beliefs and desires are "distinct existences".

However, even granting the plausibility of the argument from motivational internalism against ethical cognitivism in the moral case, the argument from (5) and (6) to the rejection of (4) in the case of meaning is surely highly implausible. The problem is that (5) has no plausibility whatsoever. It is easy to describe a situation in which a practically rational agent judges that "magpie" means magpie but has no motivation whatsoever to apply the expression "magpie" to magpies only: imagine a case where you judge that you mean magpie by "magpie" but are confronted by a thug who threatens to kill you unless you apply "magpie" to things that aren't in fact magpies. Surely in this scenario, your lack of motivation to apply "magpie" to the magpie in front of you doesn't entail that you are prone to practical irrationality? While it may be debatable whether someone who sincerely makes a moral judgement but fails to be motivated to act in accord with it must as a matter of conceptual necessity be practically irrational, the analogous claim in the case of meaning, at least when construed as (5), is simply a non-starter.[45]

In order to get a plausible form of motivational internalism along the lines suggested by Boghossian, we would need something along the lines of

(7) It is a conceptual truth that: if an agent S judges that "magpie" means *magpie*, then so long as S is practically rational *and has a desire not to convey false information*, S will be motivated to apply "magpie" to an object only if that object is a magpie.

This is not obviously implausible in the manner of (5), but of course the reference to a desire not to convey false information now means that there is no hope of combining it with (6) to yield a problem for (1): there will be no necessary connection implicated between a mere belief and a particular desire, but rather between a belief-desire pair and a further desire, which is perfectly consistent with (6) and the Humean theory of motivation. So the normativity argument suggested by Boghossian's remarks on motivation faces the following dilemma. If motivational internalism is construed along the lines of (5), then in tandem with (6) it would generate a prima facie argument against (4). However, as argued above, (5) is obviously implausible. On the other hand, if motivational internalism is construed along the lines of (7), it may well be plausible, but now – because of the reference to a background desire – can no longer combine with (6) to pose even a prima facie problem for (4).[46]

So, there appears to be no straightforward argument from motivational internalism against cognitivism about meaning, at least in the form suggested by Boghossian.[47]

5.7 "Factualist" interpretations of Kripke's Wittgenstein

In this chapter we have followed the "standard" interpretation of KW as proposing, in the sceptical solution, a *non-factualist* account of ascriptions of linguistic meaning and understanding. In recent years, the standard interpretation has been challenged by a number of philosophers, including George Wilson, Alex Byrne, David Davies, Jussi Haukioja, and Martin Kusch.[48]

The alternative – "revisionary" – interpretation of KW sees him instead as advocating a particular form of *factualism* about ascriptions of meaning: they do have truth-conditions, it is just that they don't have the kind of truth-conditions that the sceptical argument attacks. In other words, there are facts about meaning, but not the kind of facts undermined by the argument of KW's sceptic. The revisionary interpretation is rich and extremely interesting, and the various authors mentioned above do not share a single, unified interpretation of KW: although the revisionary interpretation merits an extended discussion and evaluation, we cannot attempt this here, and will instead offer some reasons for doubting the revisionary interpretation as expounded by George Wilson. We will offer a modified version of the standard interpretation designed to accommodate some of Wilson's more plausible contentions.

According to Wilson, KW's main target is what he calls the Classical Realist account of what the truth-conditions of ascriptions of meaning consist in:

(CR) If a speaker X means something by a term T, then there is a property *P* that governs the correct application of T for X.[49]

So, for example, I mean something by the term "green", so there is a property, *being green* (or *greenness*), that governs the correct application of "green" for me, in the sense that, as I understand it, "green" correctly applies to an object if and only if that object instantiates the property *greenness*.

Now, if this is indeed the case with me, there must be some fact about me that constitutes the fact that it is greenness that provides the standard of correctness of "green" for me, and not, say, *grueness* (where an object is grue if and only if it is either green before time t or blue thereafter). As Wilson puts it "The existence of conditions of applicability for a term must be intelligibly *grounded* in facts about the speaker's psychological and/or social history".[50] In general:

(G) If there is a property *P* that governs the correct application of T for speaker X, then there are facts about X that constitute *P* as the condition that governs speaker X's use of T.

Wilson calls this the "Grounding Constraint", and sees KW's sceptic as attempting to argue that the grounding constraint cannot be satisfied: there are no facts about me that constitute the property of *greenness* (as opposed to something like *grueness*) as governing the correct application of "green". Since this holds in general, KW's sceptic can draw what Wilson calls the "Basic Sceptical Conclusion":

(BSC) There are no facts about a speaker X that constitute any set of properties as conditions that govern X's use of T.

According to Wilson, in addition to BSC, KW's sceptic goes on to infer the "Radical Sceptical Conclusion":

(RSC) No-one ever means anything by a term.

Now, does KW himself (as opposed to KW's *sceptic*) accept RSC in addition to BSC? As we saw above, Kripke distinguishes between "straight" and "sceptical" solutions to the sceptical problem developed by KW's sceptic. Kripke writes "[A] proposed solution to a sceptical philosophical problem [is] a *straight* solution if it shows that on closer examination the scepticism proves to be unwarranted".[51] In other words, a straight solution is one which finds some fault in the sceptic's reasoning or denies one of the premises in the sceptic's argument. A "sceptical solution", in contrast "begins on the contrary by conceding that the sceptic's negative assertions are unanswerable".[52] In other words, a sceptical solution begins by accepting the negative *conclusion* of the sceptical argument. Given that KW is himself a proponent of a sceptical solution, and given that one of the conclusions drawn by his sceptic is RSC, it would be easy to move from these facts to the conclusion that KW himself also accepts RSC. And given this – that no-one *really* ever means anything by a term – one might see KW as pressured in the direction of non-factualism in the same way that the claim that nothing is really right or wrong might pressure one in the direction of an emotivist brand of non-factualism about ethics. Wilson, though, suggests that we should resist the temptation to see KW as himself conceding RSC as part of his sceptical solution. After all, Kripke describes

the claim that all language as meaningless as "incredible and self-defeating" and "insane and intolerable".[53] Rather, suggests Wilson, we should see KW not as conceding RSC but rather as *rejecting* RSC and then using this as a basis for the rejection of Classical Realism. So, according to Wilson, we should distinguish between the overall argument propounded by KW's sceptic and that propounded by KW himself:

KW's sceptic: CR, G, BSC; therefore, RSC.

KW: G, BSC, not-RSC; therefore, not-CR.

In saying that in the sceptical solution KW is accepting the sceptic's negative conclusion, Wilson suggests, Kripke means that KW is conceding the Basic Sceptical Conclusion BSC, not that KW is conceding the Radical Sceptical Conclusion RSC. The real target of KW's overall argument is thus not factualism about ascriptions of meaning, but rather the Classical Realist position embodied in CR. Factualism about meaning is not undermined, because although KW asserts both:

(S) No sentence has classical realist truth-conditions

and

(S*) Ascriptions of meaning do not have classical realist truth-conditions

he is not committed to rejecting either:

(S**) There are no facts about X that assertions of the form "X means ∅ by T", even when they are correct by ordinary criteria, describe truly

or

(S***) There are no facts that constitute X's meaning something by a term T.

It is not clear whether Wilson thinks that we are entitled to say that there are facts constitutive of meaning, or that some ascriptions of meaning can be viewed as true descriptions of genuine matters of fact, because of some minimalist or deflationary account of facts and truth-aptitude: some of his remarks suggest not, while others suggest that he does in general advocate a deflationary conception of truth-aptitude.[54] We will now see that, either way, there is no factualist account of ascriptions of meaning that is adequate as an expression of KW's views.

Suppose, then, that Wilson intends to defend the factualist interpretation of KW, without recourse to a deflationary account of truth-aptitude, in the following way. We can legitimately claim that there are facts constitutive of X's meaning something by a term T, or that assertions of the form "X means Ø by T" sometimes truly describe matters of fact, by utilising the positive account of the normativity of meaning that KW develops in the sceptical solution as an alternative to the allegedly bogus account provided by Classical Realism. The crucial difference between the Classical Realist account of normativity and that proffered by KW in the sceptical solution turns on the different approaches they adopt in accounting for "normative conditionals" concerning meaning, such as:

(NC) If Jones means green by "green", his answer to "Is this emerald, at time t+, green" should be "Yes".

In the general case, the Classical Realist would explain this kind of normative conditional as follows:

> Since meaning something by T is thought to involve there being properties that govern X's application of the term, X must or should ascribe T to any item that has just those properties. That is, X must ascribe T to an object from this class if, respecting the standards or correctness she herself has set, her ascription is to be, in fact, correct.[55]

According to the Classical Realist, if Jones *now* means green by "green", then this entails that there is something true of him *now* that determines, modulo facts about the colours of the various

candidate items, whether "green" applies correctly or not: namely, whatever fact about him that determines that it is greenness (and not, say, grueness) that provides the standard of correctness for his use of "green". And, of course, we are conceding here that this conception of normativity has been undermined by the negative argument of KW's sceptic.

The alternative account of normativity developed in the sceptical solution, according to Wilson, turns on replacing the Classical Realist's emphasis on facts and truth-conditions with an account framed in terms of *assertability* conditions: "What the normative conditional requires of [Jones] constitutes an *assertability condition* for judgements about whether her use of the word means ['green']".[56] In Kripke's words:

> By such a conditional we do not mean, on the Wittgensteinian view, that any state of Jones guarantees her correct behaviour. Rather, by asserting such a conditional we commit ourselves, if in the future Jones behaves bizarrely enough (and on enough occasions), no longer to persist in our assertion that she [means green by "green"].[57]

So, for example, if after time t+ Jones begins to apply "green" to, for example, Glasgow Rangers football shirts, we will withdraw our assertion that he means green by "green" (perhaps he means grue by "green", in which case "green" as he means it is correctly applicable to Glasgow Rangers shirts at t+). Wilson contends that there is nothing in the switch from the Classical Realist account of normativity and the normative conditionals to KW's account in terms of assertability conditions that forces us to embrace either (S**) or (S***): in rejecting the idea that ascriptions of meaning purport to describe facts as conceived of by the Classical Realist, we are not committed to denying that they purport to describe facts *tout court*, since there is no reason to deny that, as a matter of fact, we do employ normative conditionals in the manner outlined in Wilson's version of the sceptical solution:

> The new Sceptical Solution tries to explain the content of meaning ascriptions in terms of their role and utility in the relevant

> language games, and the normative conditionals about meaning, which the Classical Realist misconstrues as describing a super-rigid semantical determination, are explained in terms of the requirements that our use of standard criteria for meaning ascriptions engender and enforce.[58]

We will now see that this interpretation of the line adopted by KW in the sceptical solution is inadequate, since the (non-Classical Realist) notion of meaning-fact which it delivers itself falls prey to the argument of KW's sceptic, so that on this interpretation KW, like the Classical Realist, turns out to be committed to the "insane and intolerable" view encapsulated in the Radical Sceptical Conclusion RSC.

We can see this by running a simple analogue of the argument that in Wilson's view KW's sceptic deploys against Classical Realism.

> (WSS) If X means something by a term T, then there are requirements enforced and engendered by our standard criteria for ascriptions of meaning, that govern the correct application of T for X.

(WSS) is a simple consequence – perhaps even a restatement of – the account of normativity that Wilson discerns in the sceptical solution. The crucial point for our current concern is that, just as it was reasonable for KW's sceptic to impose a grounding constraint G on the Classical Realist account of normativity, an analogous constraint can likewise be imposed on Wilson's version of the sceptical solution:

> (GWSS) If there are requirements, enforced and engendered by our standard criteria for ascriptions of meaning, that govern the correct application of T for speaker X, then there are facts about X that constitute those requirements as the requirements that govern speaker X's use of T.

And, just as KW's sceptic asked of the Classical Realist, we can ask: what fact about Jones (or his speech community) constitutes the requirement (i) *"green" correctly applies to an object if and only if that object is green*, and not the requirement (ii) *"green" correctly applies to an object if and only if that object is either green before time t or blue thereafter*, as the requirement enforced and engendered by our standard criteria for ascriptions of meaning that governs Jones's use of "green"? As before, no fact about Jones or his speech community will be found capable of playing this role. So, it follows that:

> (BSC*) There are no facts about a speaker X that constitute any requirements as requirements, enforced and engendered by our standard criteria for ascriptions of meaning, that govern X's use of T.

So it follows, as before, that:

> (RSC) No-one ever means anything by a term.

So, the "facts" about meaning allegedly delivered by the sceptical solution turn out to be just as evanescent as those sought by the Classical Realist.

It should be clear enough what has gone wrong here. Wilson has failed to notice that any requirements allegedly "enforced and engendered by our standard criteria for ascriptions of meaning" will, if introduced as part of a factualist account of meaning ascriptions, fall prey to the argument that KW's sceptic deployed against the Classical Realist. This has happened because Wilson has neglected the stricture – emphasised in §5.5 above – that any assertability conditions introduced into the sceptical solution must not be viewed as providing truth-conditions for ascriptions of meaning, or as leaving conceptual space for facts about meaning, on pain of attempting a straight solution to KW's sceptic's negative conclusion.[59]

The argument just developed suggests that Wilson's interpretation seriously underestimates the power of the negative argument developed by KW's sceptic. The position attributed to Wilson above did not see his account of the sceptical solution as utilising a minimalist or deflationary account of facts or truth-conditions in defence of factualism about ascriptions of meaning. Could such minimalism provide the basis for an alternative route to a factualist interpretation of the sceptical solution that avoids the charge that it, too, would lead to the "insane and intolerable" Radical Sceptical Conclusion?[60]

According to a minimalist conception of truth-aptitude a discourse consists of sentences that are truth-apt – and hence potentially describe matters of fact – so long as it satisfies the following two conditions:

Discipline: The discourse is disciplined, in the sense that there must be standards operative with respect to which uses of the sentences are judged to be appropriate or inappropriate. There must be acknowledged standards for the proper and improper use of sentences of the discourse.

Syntax: The sentences of the discourse possess the right sort of syntactic features. In particular, they must be capable of conditionalisation, negation, embedding in propositional attitudes, and so on.[61]

One could perhaps take KW's suggestion that Classical Realist truth-conditional accounts of meaning have to be superseded by a theory of meaning that focuses on assertability conditions and their role and utility as pointing towards a minimalist conception of truth-aptitude. Indeed, both Wilson, and more unequivocally David Davies, are aware of the possible connection between their factualist account of KW and the minimalist conception.[62] The thought would be, presumably, that according to the sceptical solution ascriptions of meaning satisfy the two conditions relating to discipline and syntax, and hence, by the minimalist account,

are truth-apt, and when true, descriptive of matters of fact; but that they are not true in any sense that goes beyond the minimal, deflationary notion, so that because they are *merely* minimally truth-apt and, in some cases, *merely* minimally true, they still compare differentially with statements from more "robust" areas of discourse.[63]

It is arguable, though, that the view KW adopts in his sceptical solution cannot be so formulated. This is because the assertability conditions – standards of appropriateness – that provide for merely minimal truth-aptitude are normative in a sense directly put under pressure by the conclusion of KW's sceptic's negative argument.[64] Since they are standards that sort X's uses of T into the categories *correct* and *incorrect*, in adopting the minimalist conception of truth-aptitude we are in effect taking for granted the notion of correctness questioned by the sceptic's negative argument. This means that in this context, adopting the minimalist conception of truth-aptitude in defence of a factualist account of ascriptions of meaning is effectively indistinguishable from the non-reductionist account of meaning as primitive and *sui generis* that KW's sceptic rejects as "mysterious and desperate" (see §5.1). Given that KW is said to agree with his sceptic's objection to this kind of non-reductionist view, the conclusion, once more, is that we do not have here a sound basis for a defence of a factualist account of ascriptions of meaning.[65]

There is, then, no plausible way in which KW's sceptical solution can be viewed as underwriting a factualist account of ascriptions of meaning. A non-factualist account of ascriptions of meaning appears to be the only option for a proponent of KW's sceptical solution. But can we dissociate the non-factualist account from Wilson's suggestion that it illegitimately and implausibly runs together the Basic Sceptical Conclusion BSC and the "insane and intolerable" Radical Sceptical Solution RSC?

In fact, it is relatively straightforward to do so. We need to begin by characterising factualism:

(F) Ascriptions of meaning are factual.

We then yoke the notion of a fact about meaning to Classical Realism:

Ascriptions of meaning are factual if and only if (If a speaker X means something by a term T, then there is a property *P* that governs the correct application of T for X).

When KW's sceptic's argument is taken together with

(G) If there is a property *P* that governs the correct application of T for speaker X, then there are facts about X that constitute *P* as the condition that governs speaker X's use of T

we reach

(BSC) There are no facts about a speaker X that constitute any set of properties as conditions that govern X's use of T.

And then, as above

(RSC) No one ever means anything by a term.

KW's sceptic can thus be viewed as arguing

KW's sceptic: F, F iff CR, G, BSC; therefore, RSC

Given that the notion of a fact about meaning has been yoked to Classical Realism, we can view KW as arguing instead:

KW: G, BSC, not-RSC; therefore, not-CR; F iff CR; therefore, not-F

There is thus an interpretation of KW's sceptical conclusion that views it as rejecting *both* Classical Realism and factualism about ascriptions of meaning, without spiralling into the insane and

intolerable scepticism of RSC. Given the problems with the factualist characterisations of KW's sceptical solution, then, it seems that the most plausible interpretation of KW's sceptical solution views it as rejecting, inter alia, both Classical Realism and factualism about meaning. Given the problems canvassed earlier in this chapter for non-factualism about meaning, it remains imperative to seek a straight solution to KW's sceptic's negative arguments. It is to such solutions that we turn in Chapter 6.[66]

Chapter summary

In this chapter we outlined Kripke's Wittgenstein's sceptical argument for the conclusion that there are no facts in virtue of which ascriptions of meaning, such as "Jones means addition by '+'", are true or false. We saw how Kripke's Wittgenstein considers and rejects a number of candidate meaning-constituting facts ("straight solutions"), including facts about mental images and qualitative mental states, facts about non-qualitative mental states, and facts about how speakers are disposed to use linguistic expressions. We saw that in these arguments KW relies on the idea that meaning is normative, that in some sense the meaning of an expression implies facts about how it *ought* to be used. We saw that KW himself argues that we can give an account of our practice of ascribing meanings which is not undermined by the admission that there are no meaning-constituting facts (a "sceptical solution"), and takes this to imply the impossibility of a language spoken by a solitary speaker. We outlined arguments against this kind of non-factualist view of meaning: Boghossian's argument that it presupposes and denies a robust conception of truth, Wright's argument that it implies an unstable form of global non-factualism, and Zalabardo's idea that it treats the sceptical paradox as an essentially epistemological rather than constitutive form of scepticism. We then investigate one dimension along which meaning might be normative – to do with its capacity to provide intrinsic motivation for linguistic use – and argued that it appears not to be normative in this sense. Finally, we argued that the prospects for Wilson's "factualist" construal of KW do not look good.

Study questions

1. Why couldn't the fact that you mean cube by "cube" consist in an association between the world and a mental image of a cube? Give at least two reasons for your answer.
2. Could the fact that you mean addition by "+" be constituted by facts about how you are disposed to use it? Give at least three reasons for your answer.
3. Explain the difference between a "straight solution" and a "sceptical solution" to a sceptical paradox like that developed by KW. Is KW's "sceptical solution" a convincing reply to the sceptical paradox about meaning?
4. Is there any philosophically interesting sense in which meaning is normative?
5. Is the reply to Wright's objection to the sceptical solution adumbrated in §5.4 guilty of viewing the sceptical solution as embracing RSC in addition to BSC? If so, is there another reply available to the defender of the sceptical solution?

Further reading

The key text here is Kripke, *Wittgenstein on Rules and Private Language*. Further reading on responses to the sceptical paradox will be provided at the end of Chapter 6. Boghossian's objections to the sceptical solution can be found in "The Rule-Following Considerations" (§§III–IV) and "The Status of Content". For further discussion, see Devitt, "Transcendentalism about Content"; Boghossian, "The Status of Content Revisited"; Blackburn, "Wittgenstein's Irrealism"; and Wright, *Truth and Objectivity*, Chapter 6. Wright's own objections to the sceptical solution can be found in his "Kripke's Account of the Argument against Private Language". Zalabardo's objection can be found in his "Rules, Communities, Judgements". Other works of interest include Blackburn, "The Individual Strikes Back"; Goldfarb, "Kripke on Wittgenstein on Rules"; and McGinn, *Wittgenstein on Meaning*, Chapter 4. It should be pointed out that it is almost universally accepted that whatever the philosophical merits of KW's

arguments, they do not accurately reflect the views held by the actual, historical Wittgenstein. See McGinn, *Wittgenstein on Meaning*, Chapters 1 and 2 for a good explanation of why not. For "factualist" interpretations of KW, see Wilson, "Kripke on Wittgenstein on Normativity", "Semantic Realism and Kripke's Wittgenstein", "Rule-Following, Meaning and Normativity", and "On the Skepticism about Rule-Following in Kripke's Version of Wittgenstein"; Byrne, "On Misinterpreting Kripke's Wittgenstein"; Davies, "How 'Sceptical' is Kripke's 'Sceptical Solution'"; Haukioja, Rule-Following, Response-Dependence and Realism; and Kusch, *A Sceptical Guide to Meaning and Rules*, We'll return to the question of Wittgenstein's own views on meaning in the next chapter. For Boghossian's latest view on the normativity issue see his "The Normativity of Meaning" and "Is Meaning Normative?". For an excellent and comprehensive discussion of the issue see Hattiangadi, "Is Meaning Normative?" and *Oughts and Thoughts: Scepticism and the Normativity of Content*. For useful defence of the idea that meaning is normative see Whiting, "The Normativity of Meaning Defended", "Is Meaning Fraught With Ought?", and "What is the Normativity of Meaning?". For a contrasting approach see Hattiangadi, "The Normativity of Meaning".

Notes

1 Kripke is careful not to present this as straightforward exegesis of Wittgenstein. He writes "the present [work] should be thought of as expounding neither 'Wittgenstein's' argument nor 'Kripke's': rather Wittgenstein's argument as it struck Kripke, as it presented a problem for him" (Kripke, *Wittgenstein on Rules and Private Language*, p.5). In what follows I'll just speak of "Kripke's Wittgenstein", and abbreviate this by "KW".

2 Note that the interpretation of KW as propounding a form of non-factualism about ascriptions of meaning has been challenged by a group of philosophers offering instead a "revisionary" interpretation of KW that does not involve viewing him as a non-factualist. For discussion of this alternative interpretation of KW, see §5.7 below.

3 Note that KW's sceptic, like Quine, is arguing for a constitutive scepticism via an epistemological route. See §4.7.

4 Kripke says that the problem is set up in this way, with respect to past meanings, in order to enable the sceptic to intelligibly formulate his argument at all – to allow himself to present his sceptical argument to us, the sceptic at this point does not call our present meanings into question: "Before we pull the rug out from under our own feet, we begin by speaking as if the notion that at present we mean a certain function . . . is unquestioned and unquestionable. Only past usages are to be questioned. Otherwise, we will be unable to formulate our problem"; Kripke, *Wittgenstein on Rules and Private Language*, pp.13–14). Of course, once the sceptical conclusion has been established with respect to past meanings, it can be generalised to encompass present meanings too, for we can always imagine ourselves running the sceptical argument tomorrow about what we presently mean by the "+" sign.
5 Kripke, *Wittgenstein on Rules and Private Language*, pp.8–9.
6 Note that although KW's sceptic's argument has obvious affinities with Quine's argument from below for the indeterminacy of translation, it is much stronger than Quine's argument, insofar as it allows us access to a wider range of facts in our search for the facts which constitute meaning. Quine rules out an appeal to facts of the sort mentioned in (b) from the start, whereas KW's sceptic's argument is that even if we suppose ourselves to have idealised epistemic access to these sorts of fact, we will still be unable to find a fact which can constitute our meaning one thing rather than another.
7 Ibid. p.15.
8 Ibid. p.16.
9 Wittgenstein, *Philosophical Investigations*, §198.
10 I am here merely summarising the rich battery of arguments which Colin McGinn, drawing extensively on Wittgenstein's texts, develops against the suggestion that understanding might be constituted by the occurrence of mental images, etc. See McGinn, *Wittgenstein on Meaning*, pp.1–7 for some excellent exposition. See also the Fregean argument that is developed against Locke's imagist theory of meaning in §2.3.
11 Kripke, *Wittgenstein on Rules and Private Language*, p.41.
12 Ibid., pp.41–42.
13 Ibid., p.51.
14 Ibid., p.37.
15 The issue about the alleged normativity of meaning is actually considerably more complex than indicated here. See §5.7 and the relevant papers mentioned under Further reading.

16 The passages from Wittgenstein which inspired KW's sceptic's conclusion to this effect are known as "the rule-following considerations". The notion of intending to follow a rule in a certain way is analogous to meaning something by a linguistic expression. Suppose I intend to follow the rule "add 2" when writing out the following arithmetical series: 2, 4, 6, 8, 10, ... Intuitively, later on in the series, certain continuations (e.g. 24, 26, 28) are determined to be correct by the rule which I am following, and certain continuations are determined to be incorrect by that rule (e.g. 34, 35, 37). This is the analogue of the applications of a predicate being determined as correct or incorrect by the sense of the predicate. Just as KW's sceptic will claim, of a predicate, that there are no facts of the matter as to which applications are correct and which are incorrect, he will also claim that there are no facts of the matter as to which continuations of the arithmetical series are correct or incorrect. Recall that in §2.4 we said that, generally, the sense of an expression could be thought of as a rule for determining its semantic value; this shows how KW's sceptic's conclusion applies to Frege's notion of sense.

17 Ibid., p.22.

18 Clearly, KW's arguments would still be effective against views which dispense with the notion of sense and equate the meaning of an expression with its referent or semantic value: if there were no fact of the matter as to which function "+" refers to, on this type of view there is straightforwardly no fact of the matter as to what "+" means. For some useful discussion, see Boghossian, "Is (Determinate) Meaning a Naturalistic Phenomenon?".

19 Talk of a "sceptical solution" is a direct reference to the Scottish eighteenth-century philosopher David Hume. Kripke takes Hume to be arguing that e.g. there are no facts in virtue of which statements about causal relations are true or false, but as attempting to neutralise this conclusion by arguing that such statements can be viewed as having some legitimate non-fact-stating role. See e.g. Hume, *An Enquiry Concerning Human Understanding*. It is an open question whether Kripke accurately represents Hume's views on causation. For a comprehensive discussion of these views, see Beebee, *Hume on Causation*; and for some commentary Miller, "Hume".

20 The reference is to Dummett's paper, "Wittgenstein's Philosophy of Mathematics".

21 Wittgenstein, *Philosophical Investigations*, §258.

22 Ibid., §202.

23 Kripke attempts to use these remarks in an exegesis of Wittgenstein's famous "private language argument". But the

issues about the relationship between arguments against solitary language (a language which is in fact spoken only by a solitary individual) and arguments against private language (a language which is necessarily unintelligible to everyone except its speaker) are complicated. See Wright, "Does PI §258–60 Suggest a Cogent Argument" for some useful discussion.

24 For a useful brief survey of criticisms of the first type, see Boghossian, "The Rule Following Considerations", §3. For an extended discussion, see Wee, Wittgenstein, Objectivity and Rule-Following.

25 Boghossian's arguments can be found in "Rule-Following Considerations", §4, and "The Status of Content".

26 The view that sentences of a certain discourse do not possess truth-conditions is called non-factualism with respect to that discourse, since it amounts to the claim that there are no possible facts in virtue of which the sentences of that discourse might be true or false. Given the assumption that the sense of a sentence is its truth-condition, non-factualism about meaning thus amounts to the claim that sentences ascribing meanings (truth-conditions) to other sentences do not themselves possess truth-conditions.

27 Boghossian, "Rule-Following Considerations", p.162.

28 Note that (1) is really just a reformulation of (2): to say that there is no property corresponding to the predicate "has truth-condition p" is just to say that there is no property in virtue of the instantiation (or non-instantiation) of which the sentence "S has truth-condition p" is true (or false). Which is just to say that "S has truth-condition p" – an ascription of meaning to a sentence – does not itself possess truth-conditions.

29 Wright, "Kripke's Account", p.104.

30 Ibid., p.105. To say that a statement is projective is just to say that it doesn't aspire to state a fact.

31 Boghossian, "Rule-Following Considerations", p.161.

32 Kripke, *Wittgenstein on Rules and Private Language*, p.38.

33 Ibid.

34 In fact, Wright, "Kripke's Account", p.108, footnote 6, argues that this rebuttal of the simplicity suggestion begs the question because it *assumes* non-factualism about ascriptions of meaning, a thesis which, at the relevant stage in the dialectic, is still up for proof. Even if correct this point does not damage the application of the objection to the sceptical solution itself, for there we are ex hypothesi assuming the conclusion of the sceptical argument.

35 Zalabardo, "Rules, Communities, Judgements", p.36.
36 Ibid., p.38.
37 Ibid., p.39.
38 Wright, "Kripke's Account", p.770.
39 It seems to me that the only way KW can avoid the argument of the present section is to give an account of the meaning of ascriptions of meaning along *non-cognitivist* lines (see §9.2). This would interpret judgements about meaning as expressing *attitudes* (perhaps confidence that other speakers will give the same answers as oneself), in much the same way that ethical non-cognitivists view moral judgements as expressing non-cognitive attitudes rather than beliefs. The assertability conditions would then not be viewed as constituting the truth-conditions of ascriptions of meaning: the semantic function of such ascriptions would be viewed as that of expressing attitudes, rather than stating facts, so that the first horn of Zalabardo's dilemma would be avoided. And the second horn would be avoided because although we would not be attempting to capture the sense of ascriptions of meaning in terms of truth-conditions, we would be attempting to capture their sense by viewing them as having a different semantic function, that of expressing attitudes. Talk of an ascription of meaning agreeing more closely than another with communal linguistic inclinations would then be cashed out in terms of clashes of non-cognitive sentiment, as in the account of moral disagreement favoured by ethical non-cognitivism. There seems to be no reason why viewing KW as a non-cognitivist in this sense would affect the argument against the possibility of solitary language: the claim against solitary language would be recast as the claim that in such a language there would be no distinction between the conditions under which Jones expresses confidence that he means addition by "+" and the conditions under which such confidence would be rightfully expressed. Having said that, however, viewing KW as such a non-cognitivist does little to make the sceptical solution more palatable, given the long list of worries and problems faced by non-cognitive theories in other areas. For a resume of such problems, see Hale, "Can There Be a Logic of Attitudes?" and Miller, *Contemporary Metaethics*, Chapter 3. For an attempt to respond to Kripke along broadly "non-cognitivist" lines, see Gibbard, *Meaning and Normativity*. For an argument that non-cognitivist/non-factualist theories of meaning can't constitute a plausible reply to the sceptic, see Miller, "Rule-Following Skepticism".

40 Define an "amoralist" as someone who is practically rational, makes a moral judgement, but is completely unmotivated to act in accord with that judgement. Then, motivational internalism is effectively the view that amoralists are conceptually impossible. Motivational externalists on the other hand hold that amoralists are at least conceptually possible. See Smith, *The Moral Problem*, Chapter 3.

41 Suppose Jones is motivated to open the fridge door; one way of explaining this would be to attribute to Jones a belief that there is beer in the fridge together with a desire for some beer. In effect, the Humean theory of motivation holds that all explanations of motivation involve the ascription of desires and means–end beliefs, as in this case, and also that beliefs and desires are "distinct existences". For a defence of the Humean theory of motivation, see ibid., Chapter 4.

42 This way of setting up the argument derives from Smith's presentation in ibid. Note that Smith himself does not endorse the argument, but rather attempts to defuse it by arguing that initial appearances notwithstanding there is actually no contradiction generated by (1), (2), and (3).

43 Boghossian, "Rule-Following Considerations", p.170.

44 Ibid., second emphasis added.

45 It would be implausible to suggest that in this case the practically rational agent is motivated to act as he does because, although he is in fact motivated to apply "magpie" only to magpies, this motivation is overpowered by a stronger motivation to stay alive. It seems to me far-fetched that such an agent would feel any motivational conflict in the situation described. Compare this with the case of a conflict between the dictates of morality and the dictates of prudence, at least as conceived by a particular agent: here it does seem plausible that there would at least be the sort of motivational conflict absent in the semantic case.

46 One way to attempt to patch up Boghossian's version of the argument would be to distinguish between *true* use and *correct* use in the following sense (see Millar, "The Normativity of Meaning"). Suppose that a speaker means magpie by "magpie", is confronted by a pigeon o, and desires to express the belief that o is a magpie. Then there is a sense in which the speaker's applying "magpie" to o is correct, even though false. One might then formulate a modified version of motivational internalism as the view that it is a conceptual truth that if an agent S judges that "magpie" means *magpie*,

then so long as S is practically rational, S will be motivated to use "magpie" in patterns that are linguistically correct. I consider this suggestion in Miller, "Semantic Realism and Motivational Internalism".

47 We've looked at only one metaethical "normativity" argument in this section. Other metaethical arguments that would have to be considered in an exhaustive discussion of the normativity of meaning include the argument from the instrumental theory of rationality and Moore's "open question argument". See Miller, "Meaning Scepticism", §4.2. See also Miller, "Argument From Queerness".

48 See the Further reading section in this chapter.

49 This is a simplified form of Wilson's characterisation of Classical Realism: where we talk of a single property, Wilson talks of a set of properties.

50 Wilson, "Kripke on Wittgenstein on Normativity", p.239.

51 Kripke, *Wittgenstein on Rules and Private Language*, p.66.

52 Ibid.

53 Ibid., p.71 and p.60 respectively.

54 For Wilson's denial that he is operating against the background of minimalism or deflationism, see Wilson, "Semantic Realism and Kripke's Wittgenstein", pp.121–22. He appears to retract this in Wilson, "Rule-Following, Meaning and Normativity", pp.162–163. Davies is quite explicit about the commitment to minimalism in his account of the sceptical solution; Davies, "How 'Sceptical' is Kripke's 'Sceptical Solution'?", p.131.

55 Wilson, "Kripke on Wittgenstein on Normativity", p.253.

56 Ibid., p.254.

57 Kripke, *Wittgenstein on Rules and Private Language*, p.254.

58 Wilson, "Kripke on Wittgenstein on Normativity", p.95.

59 In effect, then, Wilson's version of the sceptical solution falls prey to Wright's argument about assertability conditions, outlined in §5.5 above.

60 It is worth noting that although Wright thinks that KW's intended position is non-factualist, he argues that KW ought to have formulated his position as the view that ascriptions of meaning are merely minimally truth-apt and merely minimally true; Wright, *Truth and Objectivity*, p.212.

61 See ibid. for the minimalist account of truth-aptitude mentioned here.

62 See n.54 above for the relevant references.

63 For discussion of what this additional "robustness" might consist in, see Wright, *Truth and Objectivity* and Chapter 9 in this volume.
64 As noted by Edwards, "Debates about Realism", p.64.
65 The claim here is not that the primitivist non-reductionist position attacked by KW is identical to the view that ascriptions of meaning are merely minimally truth-apt and true, rather that in this particular context (i.e. replying to KW's sceptic) simply helping oneself to the notion of semantic correctness in the characterisation of a minimal fact is open to the same objection that KW directs against the primitivist. Note that Davies, "How 'Sceptical' is Kripke's 'Sceptical Solution'?", p.137, acknowledges the non-reductionism implicit in his interpretation but he doesn't consider whether it falls prey to the difficulties raised by KW's sceptic.
66 A suggestion worth pursuing that to my knowledge has not been canvassed in the literature would be that KW's real target is the grounding constraint G, but I cannot consider this suggestion here. Although I have argued against the factualist interpretation of KW, I do think that Wilson and the other factualists have done a great service to defenders of the non-factualist line in forcing them to get clearer on their interpretation. Certainly, the version of the present chapter in the first edition of this book suffered from unclarity concerning the distinction between BSC and RSC. I pursue the factualist interpretation further in Miller, "Kripke's Wittgenstein, Factualism and Meaning" (I've drawn on parts of this in this section).

Chapter 6

Saving sense

Responses to the sceptical paradox

In the previous chapter, we outlined Kripke's Wittgenstein's sceptical conclusion: there are no facts in virtue of which ascriptions of meaning, such as "Jones means addition by '+'", are either true or false. We saw that this conclusion threatened us with a sceptical paradox to the effect that no-one ever means anything by any linguistic expression, and also that KW's own attempt at rehabilitating meaning in the face of this conclusion – the sceptical solution – faces severe difficulties. We now look at a number of attempted "straight" solutions to the sceptical argument, solutions which try to meet the sceptic head on by giving an account of the facts which constitute meaning. In §6.2 and §6.3 we look at attempts to defend the sort of dispositionalist theories of meaning which we saw KW attacking in §5.1. In §6.4 we look at Jerry Fodor's "asymmetric dependency" account of meaning. In §6.5 we look at Colin McGinn's attempt to construe meaning and understanding in terms of the possession of *abilities or capacities*. In §6.6 we outline Crispin Wright's attempt to respond to the sceptical argument by claiming that facts about meaning are "judgement-dependent",

and in §6.7 we'll look at Philip Pettit's "ethocentric" view. In §6.8, we'll look at (what is arguably) Wittgenstein's own response to the sceptical paradox, before finishing in §6.9 with a discussion of Hannah Ginsborg's recent invocation of primitive normative attitudes as a way of replying to KW's sceptic. Before looking at these responses to KW's sceptical argument, we pause to make some brief comments on the relationship between scepticism about the notion of meaning as applied to expressions of public language, and scepticism about the content of states of mind.

6.1 Linguistic meaning and mental content

So far in this book we have been concerning ourselves with questions about the meaning of expressions in a language. But it is not only linguistic expressions that are capable of possessing meaning. Some mental states – beliefs, desires, intentions, and wishes, etc. – are also normally thought to possess meaning or content. For example, the content of my belief that Edinburgh is in Scotland is *that Edinburgh is in Scotland*. Just as the sentence "Edinburgh is less impressive than Glasgow" is *about* Edinburgh, and says that it is less impressive than Glasgow, my belief is *about* Edinburgh, and is to the effect that it is in Scotland. This is also true of the other types of mental states mentioned. For example, I can have an intention with the content *that I will finish this book on time*; I can have a wish with the content *that the price of beer will be halved in the next budget*; and I can have a desire with the content *that I win next week's National Lottery*. Philosophers call mental states such as beliefs, desires, wishes, intentions, hopes, and so on, *propositional attitudes*.[1] We can say that just as sentences have linguistic meaning, propositional attitudes have *mental content*.

Having made this observation, the following question now arises: what is the relationship between linguistic meaning and mental content? Do we have to explain the notion of linguistic meaning in terms of the notion of mental content? (i.e. Do we have to explain what it is for a linguistic expression to have meaning in terms of the propositional attitudes of speakers of the language?), or can we explain the notion of mental content in terms of the

notion of linguistic meaning? (i.e. can we explain what it is for a speaker to have a propositional attitude with a certain content in terms of the linguistic meaning of expressions of a language?). Philosophers of language have divided over the correct answers to these questions: Paul Grice has led philosophers who attempt to explain linguistic meaning in terms of the content of propositional attitudes, while Michael Dummett has been the main proponent of the view that the content of propositional attitudes must be explained in terms of the notion of linguistic meaning.[2] In addition, philosophers such as Donald Davidson have argued that there is no explanatory priority either way: linguistic meaning and mental content must be explained *together*, or not at all.

Deciding which of these three types of view on the relationship between language and thought is correct is an extremely difficult task, and in this book we'll only be able to give a brief outline of some of the arguments for each type of position (for the Gricean View, see Chapter 7; Davidson is discussed in Chapter 8; and Dummett is discussed in Chapter 9). In this section, our concern is with a more limited question: what is the relationship between scepticism about linguistic meaning and scepticism about mental content? For example, if KW's sceptical argument is cogent, does it carry over to the case of mental content: would it follow that there are no facts of the matter in virtue of which we have propositional attitudes with one content rather than another? Alternatively, is there space for a sceptical argument about the contents of thoughts, which leaves the meanings of linguistic expressions alone?

The answer to the last of these questions is "yes", and the answer to the second is "no" regardless of what type of view on the explanatory priority issue is adopted. Suppose one adopted the Gricean account, on which linguistic meaning is to be explained in terms of the propositional attitudes possessed by speakers. Since linguistic meaning is explained *in terms of* mental content, KW's sceptic's argument, which is targeted on linguistic meaning, will carry over to threaten the notion of mental content: if A is explained in terms of B, and the notion of A proves to be incoherent, then we cannot avoid the conclusion that B is incoherent also. For example, if the notion of responsibility is explained in terms of the notion of free will, then given the conclusion that the notion of responsibility is

incoherent, we will be forced to the conclusion that the notion of free will is incoherent also.[3] Likewise for the Dummettian view, on which the content possessed by propositional attitudes is to be explained in terms of the use of language. If A is explained in terms of B, then any argument that B is incoherent will thereby threaten the coherence of A: if Divine Grace is explained in terms of God, and the latter notion proves to be incoherent, then there will be nothing left of the former. So given the Dummettian view, KW's sceptical argument, in threatening the notion of linguistic meaning, will also threaten the notion of mental content. These points are well summarised by Paul Boghossian:

> There would appear to be no plausible way to promote a *language-specific* meaning scepticism. On the Gricean picture, one cannot threaten linguistic meaning without threatening thought content, since it is from thought that linguistic meaning is held to derive; and on the [Dummettian] picture, one cannot threaten linguistic meaning without *thereby* threatening thought content, since it is from linguistic meaning that thought content is held to derive. Either way, [mental] content and [linguistic] meaning must stand or fall together.[4]

The point is even clearer in the case of the Davidsonian view that there is no explanatory priority either way: if linguistic meaning and mental content must be explained together, or not at all, then any argument against the coherence of one notion is straightforwardly an argument against the coherence of the other.

In addition to the points above, there is another reason why it is difficult to restrict the conclusion of the sceptical argument to one or other of linguistic meaning and mental content. This concerns the nature of the sceptical argument itself: the argument turns, not on the fact that sentences belong to language, but just on the assumption that they possess meaning. All of the arguments that KW's sceptic ran in §5.1 could be rerun with mental items substituted for linguistic expressions, at no loss to the plausibility of the arguments. As Boghossian puts it:

> The real difficulty with the suggestion that one may sustain differential attitudes toward mental and linguistic content stems from the fact that the *best* arguments for the claim that nothing

> mental possesses content would count as *equally* good arguments for the claim that nothing linguistic does [and vice versa]. For these arguments have nothing much to do with the items being *mental* [or *linguistic*] and everything to do with their being contentful: they are considerations of a wholly general character, against the existence of items individuated by content.[5]

Getting out of the predicament that we were left in at the end of Chapter 5 is thus all the more pressing. If KW's sceptic's argument is successful, it follows that there is no fact of the matter as to whether we have propositional attitudes with specific contents. And this is a difficult conclusion to live with. Given that the content of propositional attitudes plays an essential role in the explanation of human *action* – for example, Jones reached for the fridge because he had a belief that there was beer in the fridge and a desire for a beer – the whole picture of human beings as *agents* comes under threat. So are there any plausible straight solutions to the sceptical paradox?[6]

6.2 Sophisticated dispositionalism

Recall that in §5.1 KW's sceptic argued against the idea that Jones's meaning addition by "+" could consist in his being disposed to respond to the relevant arithmetical queries with the sum of the numbers involved. Firstly, our actual dispositions are finite, and so unable to determine a unique function as the semantic value of "+". Secondly, dispositional facts are facts about what we *will* do, not about what we *ought* to do, and as such cannot capture the normativity of meaning. In this section we'll consider whether a dispositionalist theory of meaning can offer any plausible responses to these objections.

The dispositional theory considered in §5.1 was extremely crude. We arrive at a more serious dispositional theory if we follow Boghossian in the reflection that:

> All dispositional properties are such that their exercise – the holding of the relevant counterfactual truth – is contingent on the absence of interfering conditions, or equivalently, on the presence of the ideal conditions.[7]

For example, common salt possesses a disposition to dissolve when placed in water, but it will not dissolve in a sample of water which is already saturated; so to say that salt is water soluble is really to say that it is disposed to dissolve in water *under certain conditions*, one consequence of whose obtaining is that the water in question is not already saturated. The reflection suggests a sophisticated dispositionalism, the basic idea of which is that we can avoid the problems posed for the dispositional theory considered in §5.1, if we concentrate on what we would be disposed to do *under some suitably specified set of ideal, or at least ceteris paribus, conditions*. The problem which the finiteness of our actual dispositions posed might then be avoided if we could include some specification in these conditions to the effect that the speaker lives long enough to hear out the relevant arithmetical enquiry: it seems plausible to say that if I were to live long enough, even for numerals "n" and "m" which are in fact too large for me to take in in my normal life span, I would, in response to an appropriate query, utter a numeral denoting the sum of the two numbers rather than "5". This would rule out the interpretation of "+" as standing for the skaddition function (see §5.1). Moreover, it might also be plausible that such a sophisticated dispositionalism can provide us with something that captures the normativity of meaning: if the conditions specified are genuinely *ideal*, or at least conditions of *proper* functioning, then doesn't it follow from the fact that I would respond in a certain way under conditions of that type that I have a *reason* for responding in that fashion? For example, if someone in ideal conditions for appraising arithmetical claims were to say that a certain answer to an arithmetical problem was appropriate, and moreover, I know that the conditions are in fact ideal, doesn't it follow that I *ought* to accept the verdict given on the arithmetical problem? We also get the necessary distinction between competence and performance. Competence would be a matter of acting as one would act under ideal conditions; this could come apart from actual performance, in the cases where the ideal conditions fail to obtain. One could thus be systematically disposed to make a mistake, because the ideal conditions could systematically fail to obtain.

What is required, then, is a specification of the ideal conditions: a set of conditions such that *under those conditions*, Jones is

disposed to respond to the query "x + y = ?" with z, if and only if the numeral z stands for the *sum* of x and y. Since the dispositionalist is attempting to reduce facts about meaning to facts about dispositions, it is crucial that in spelling out the ideal conditions, no appeal is made to facts about meaning: this would put the cart before the horse, since the story about how speakers are disposed to behave under those conditions is supposed to tell us what constitutes the facts about meaning. So, no use of semantic or intentional materials is to be made in the specification of the ideal conditions.

In order to make the discussion simpler, we'll move away from the arithmetical example, and consider a simple observational predicate, "magpie". Intuitively, "magpie" stands for *magpies*: bullfinches, sparrows, people, and tables do not fall within its extension. Suppose that Jones means magpies by the predicate "magpie". The dispositionalist will say that this fact about Jones is constituted by the fact that he is disposed, under ideal conditions, to apply the predicate "magpie" to all and only those objects which are magpies. In other words, Jones's meaning *magpie* by "magpie" is constituted by the fact that he is disposed, under ideal conditions, to token the belief that x is a magpie, if and only if x is in fact a magpie. Can sophisticated dispositionalism spell out a set of ideal conditions such that this is true, in a way which avoids the use of prior semantic and intentional materials?

Boghossian has argued that it cannot, so that even sophisticated dispositionalism is ultimately bound to fail. Boghossian's objection takes off from the observation that the fixation of a speaker's beliefs is a *holistic* affair. He writes:

> Belief fixation is typically mediated by background theory – what contents a thinker is prepared to judge will depend upon what other contents he is prepared to judge . . . just about any stimulus can cause just about any belief, given a suitably mediating set of background assumptions.[8]

Boghossian notes that a speaker may token the belief *there is a magpie* when there is some other kind of bird present, due to the presence of a belief to the effect that there are no birds apart from magpies in the relevant neighbourhood, and so on. But

> A dispositional theorist has to specify, without use of semantic or intentional materials, a situation in which a speaker is disposed to think *lo, a magpie* only in respect of magpies.[9]

The sophisticated dispositionalist wishes to identify someone's meaning such and such with facts of the form: *S is disposed to token the belief B under conditions C*, where the C are to be specified non-semantically and non-intentionally. Boghossian's argument focuses on the fact that these conditions will have to include some proviso to the effect that certain other clusters of background beliefs B1, . . ., Bn, . . . are *absent*. Thus, in specifying the optimal conditions for the meaning-constituting disposition concerning "magpie", we will require non-intentional and non-semantic optimal conditions for the range of background beliefs which figure in the clusters B1, . . ., Bn, . . .: otherwise we will not be able to stipulate, in non-intentional terms, the conditions under which each member of that range of beliefs is absent. Since "there looks to be a potential infinity of such mediating background clusters of belief", Boghossian claims that "what is needed is precisely what a dispositional theory was supposed to provide: namely, a set of naturalistic necessary and sufficient conditions for being a belief with a certain content", and concludes on this basis that "if there is to be any reductive story about meaning at all, it cannot take the form of a dispositional theory".[10]

Let's try to get clearer on what the problem here is supposed to consist in. Suppose we are trying to give an account of the conditions under which my forming the belief *lo, a magpie* covaries invariably with (and only with) the presence of magpies. In order to do so, we have to specify conditions under which each cluster in the potentially infinite set of clusters of beliefs B1, . . ., Bn, . . . is absent. Then we require "a non-semantically, non-intentionally specified situation in which it is guaranteed that none of this potential infinity of background clusters of belief is present". And this requires just what we were supposed to be providing, namely, "a set of naturalistic necessary and sufficient conditions for being a belief with a certain content".[11]

Is the fact that the set of clusters of belief which have to be ruled out is *infinitely* large playing any significant role in Boghossian's

argument? It looks as if the argument would go through equally well even if the set was finite. In order to give an account of the conditions under which my forming the belief *lo, a magpie* covaries invariably with (and only with) the presence of magpies we have to specify conditions under which each cluster in the set of clusters of background beliefs B1, . . ., Bn, . . . is absent. Then we would require a non-semantically and non-intentionally specified situation in which it is guaranteed that no member of this set of background clusters of beliefs is present. And this requires just what we were supposed to be providing, namely a set of naturalistic necessary and sufficient conditions *for being a belief with a certain content*. (This would seem to go through even if there was only *one* such cluster which had to be ruled out. In order to rule it out we would have to give naturalistic necessary and sufficient conditions for the various beliefs which appear in the cluster – in other words, naturalistic necessary and sufficient conditions for being a belief with a certain content. And this is just what we are in the process of attempting to provide.)

So the argument doesn't seem to hinge on the *infinite* nature of the range of clusters of background beliefs, but rather on the fact that it is clusters of *beliefs* which have to be ruled out, so that the attempted reduction of facts about meaning to dispositional facts turns out to be viciously circular. If the belief *lo, a magpie* were itself a member of one of the sets of background beliefs whose absence was relevant for a member of one of the B1, . . ., Bn, . . ., say B, then it would follow on the one hand that (a) we require non-semantic and non-intentional optimal conditions for B in order to specify the non-semantic and non-intentional optimal conditions for *lo, a magpie*, while on the other that (b) we require non-semantic and non-intentional optimal conditions for *lo, a magpie* in order to specify the optimal conditions for B. The sophisticated dispositionalist account would thus be viciously circular.

If Boghossian's real objection to sophisticated dispositionalism is that it harbours a vicious circularity, then there is a move at the dispositionalist's disposal which will neutralise this worry about circularity. In the next section we describe this move, which consists in wedding a sophisticated dispositionalist account of meaning to a reductive analysis of the sort developed by David Lewis.

The result is a form of what might be called *ultra-sophisticated dispositionalism*, a version of the dispositionalist view which is immune to the worry about vicious circularity we have just adumbrated. We'll then argue that although Boghossian does not consider the possibility of such a position in his original argument against sophisticated dispositionalism, that argument nevertheless possesses the resources for a rebuttal of even this ultra-sophisticated dispositionalism.

6.3 Lewis-style reductionism and ultra-sophisticated dispositionalism

There are reasons for thinking that the worry about vicious circularity *must* be surmountable. Isn't it an accepted fact that a similar problem arises when we consider the attribution of beliefs and desires to an agent? Simplifying somewhat, it is widely held that an intentional action in general requires both beliefs and desires to be present, so that when we attempt to give, for example, a constitutive account of the belief which partially rationalises a certain action, we stand in need of a similar account of the appropriate desire; but when we attempt to give a constitutive account of the desire, we find that we stand in need of a similar account of the relevant belief. But we do not think that this signals the impossibility of providing a constitutive account either of belief or desire. We would instead perhaps attempt to give an account of what beliefs and desires are which would break up into the following two components. (A) an account of how the various propositional attitude states relate to each other, consisting of an account of the platitudinous relations which constrain the relations which obtain between them. (B) an account of the interpretative principles which link the holistic system of propositional attitudes described in (A) to the non-intentional and non-semantic facts which ground their ascription.[12]

The general method here is an example of Lewis's style of reductive analysis.[13] Let's look at how Lewis's method applies to the case of colour terms. Just as the dispositionalist about meaning is attempting to give a reductive account of meaning, Lewis's method attempts to give us a reductive account of colour. Lewis's story takes off from the fact that there are a large range of

platitudes about the colours, and that it is by virtue of coming to treat them as platitudes that we are enabled to master colour vocabulary. Examples of such platitudes would be "redness causes us, under certain circumstances, to have experiences of redness", "red is more similar to orange than to blue". Lewis has tried to show how we can obtain an analysis of the concept of being red from this range of platitudes. We first of all run through the various platitudes and rewrite them so that the references to colours appear in property-name style. So the two platitudes mentioned above would be rewritten as "the property of being red causes us, under certain circumstances, to have experiences of the property of redness", "the property of being red is more similar to the property of being orange than it is to the property of being blue". Having done that for all of the colours, we go on to represent the result of conjoining the totality of platitudes as a relational predicate "T" true of all the various colour properties. That is, the conjunction will be represented by T[r g b . . .], where "r", "g", and so on stand for the properties of being red, green, and so on. Having done this we remove the property-names of the various colours and replace them with free variables so that we get T[x y z . . .].[14] Then, if there actually are colours, there is a unique set of properties related to the world and to each other in exactly the way that the conjunction of platitudes we just formed says there are. In other words, if there actually are colours then it is true that

$$\exists x\, \exists y\, \exists z \ldots \{T[x\ y\ z \ldots] \ \&\ ((\forall x^*)\ (\forall y^*)\ (\forall z^*) \ldots T[x^*\ y^*\ z^* \ldots] \text{ iff } (x = x^*,\ y = y^*,\ z = z^* \ldots))\}.$$

Then we could go on to say that the property of being red can be defined in the following manner: the property of being red is the x such that

$$\exists y\, \exists z \ldots \{T[x\ y\ z \ldots] \ \&\ ((\forall x^*)\ (\forall y^*)\ (\forall z^*) \ldots T[x^*\ y^*\ z^* \ldots] \text{ iff } (x = x^*,\ y = y^*,\ z = z^* \ldots))\}.$$ [15]

This is a reductive analysis of red: it defines it in purely non-colour vocabulary, since no colour vocabulary appears on the right-hand side of the definition.

The important point for our purposes here is the ease with which Lewis's style of analysis achieves this, despite the fact that the platitudes which it uses as its data in the case of, for example, redness, refer to *other* colours such as blue and orange, in whose governing platitudes in turn reference is going to be made to redness. For example, the set of platitudes for orange is going to contain "the property of being orange is more similar to the property of being red than it is to the property of being blue". *But it is just this sort of circularity which Boghossian sees as fatal to the prospects of a reductive account of belief*: the objection was after all that the optimality conditions for one belief were going to refer to (the absence of) other beliefs, in the specifications of whose optimality conditions reference was going to have to be made to the (absence of) the original belief. *This shows that the sort of circularity which Boghossian points to needn't worry the sophisticated dispositionalist. It just shows that he will need to become an ultra-sophisticated dispositionalist, and adopt the Lewis-style analysis we've just sketched*: the worry about circularity can be avoided by treating the clauses that rule out certain background clusters of belief as akin to the various platitudes about colour that form the raw material for the Lewis-style reductive analysis of colour.

Is ultra-sophisticated dispositionalism plausible? Does it give us a plausible way of reducing facts about meaning to facts about dispositions? The answer to these questions is "no". We'll now see that even though ultra-sophisticated dispositionalism can avoid the worry about circularity by adopting a Lewis-style method of reductive analysis, it is nevertheless open to a serious objection.

Recall that in our presentation of Boghossian's objection in §6.2, the play with the fact that the range of defeating background beliefs was potentially infinite appeared to be idle. The real objection, it was suggested, concerns *vicious circularity*. The emphasis on circularity is predominant in the presentation of Boghossian's argument in "The Rule-Following Considerations". But in Boghossian's later paper "Naturalizing Content", the emphasis switches from circularity to the fact that the range of

combinations of background beliefs that have to be ruled out is potentially infinite. The worry now centres on our ability to *specify* the appropriate range of optimality conditions. Specifying the optimality conditions would involve giving an account of the various clusters of background beliefs which would have to be absent in order to ensure that the disposition to form the belief *lo, a magpie* really did covary with (and only with) the presence of magpies. This account would be of the form of

(*) – Bel 1 & – Bel 2 & – Bel 3 & . . .

and Boghossian writes that

> Since, however, there looks to be a potential infinity of such mediating background clusters of belief, a non-semantically specified optimality condition would consist in the specification of a situation in which it has somehow been guaranteed that none of this potential infinity of background clusters of belief is present. And it appears utterly incredible that there should be such a specification . . . The trouble is that proposition (*) is not finitely statable: there is no finite way to state what beliefs the [reductionist] must exclude before he may be assured of the desired concomitance of magpie beliefs and magpies. Literally any belief can frustrate the desired connection.[16]

Here it is clear that it is not so much the fact that the optimality conditions have to ensure the absence of *beliefs* which causes the trouble, but rather the fact that the range of items which have to be excluded is *potentially infinite*.

This does appear to be a damaging objection to sophisticated dispositionalism. But does it also apply to ultra-sophisticated dispositionalism? The Lewis-style analysis begins by writing out all of the platitudes about colour. But this presupposes that it is at least *in principle* possible to write out a *list* of the various platitudes, which presupposes in turn that the list of platitudes in

question is not infinitely long. It may indeed be the case in the colour example that the list of platitudes is finite, but Boghossian's point about mediating background clusters of belief would seem to suggest that this will not be so in the case of content. The analogue of the platitudes about colour in the ultra-sophisticated dispositional analysis of meaning is the list of clauses ruling out awkward background clusters of belief. Since this list of clauses is not finitely stateable, the Lewis-style analysis will not even be able to get started in the case of meaning, because there will be no way of circumscribing the list of relevant platitudes that the analysis takes as its starting point.

Thus, even if switching to ultra-sophisticated dispositionalism can help the dispositionalist avoid Boghossian's worry about vicious circularity, it does nothing to solve the additional worry about the unspecifiability of its optimal conditions. No form of dispositionalism about meaning – crude, sophisticated, or ultra-sophisticated – appears to constitute a successful response to KW's sceptical paradox.[17]

6.4 Fodor's "asymmetric dependency" account of meaning

The problem posed by KW is as follows: what constitutes the fact that Jones means *horses* by "horse", or *cows* by "cow"? In this section we look at the straight solution to KW's problem that has been proposed by Jerry Fodor.

Fodor is happy to follow KW in his claim that the fact that Jones means *horses* by "horse" cannot be an irreducible, *sui generis*, sort of fact (see §5.1):

> I suppose that sooner or later the physicists will complete the catalogue they've been compiling of the ultimate and irreducible properties of things. When they do, the likes of *spin*, *charm*, and *charge* will perhaps appear on their list. But *aboutness* surely won't; intentionality simply doesn't go that deep. It's hard to see, in face of this consideration, how one can be a realist about intentionality without also being, to some extent or other, a reductionist. If the semantic and the intentional are

> real properties of things, it must be in virtue of their identity with (or maybe their supervenience on?) properties that are themselves neither intentional not semantic. If aboutness is real, then it must really be something else.[18]

Recall that "semantic", "intentional", and "aboutness" are all phrases involving *meaning*. Fodor thus seeks to give an account of what it is for Jones to mean *horses* by "horse" which does not invoke any prior semantic or intentional materials.[19] One way to try to obtain an account of the sort Fodor is seeking would be to fill out (R) as follows:

(R) "X" means Y if and only if it is a law that Y's cause tokenings of "X".

Now obviously there are going to be lots of predicates which have a content ill-suited to being captured by (R), since they concern things which don't, or cannot, cause anything: predicates intuitively denoting non-existent things (perfect students, unicorns, etc.), or things which are non-spatiotemporal (numbers and other mathematical entities). But the idea here is to see how we get on with the simpler cases, such as "horse" and "cow", where there appears to be no such problem. If we can get (R) to work in these cases, then maybe we can somehow extend our account to include the more esoteric cases; while if we cannot get (R) to work in even these easy cases, we can rest assured that there is no chance it will work in the hard cases anyway.

The main problem for this construal of (R) is that it appears to leave no scope for a notion of *error*: it leaves no room for a notion of *mistaken* application of "X", and hence leaves no room for the idea of *correct and incorrect application* of "X". To see this, consider the case of "horse". (R) then claims that

(R) "Horse" means horses if and only if it is a law that horses cause tokenings of "horse".

Now one immediate problem is that if it is a law that horses cause tokenings of "horse", horses cannot *fail* to cause tokenings of horse. But this seems to preclude the possibility of misrepresentation: encountering a sheep and tokening "horse", so that the tokening of

"horse" thereby misrepresents, or represents falsely. (R) seems to imply that false representation is impossible: but if this is the case then it follows all applications of predicates are correct, so that the distinction between correct and incorrect application which is essential to meaning is lost.

Another problem is that there are going to be occasions when I do in fact misapply the predicate "horse": perhaps on dark nights, the cows in the field across from the pub look deceptively like horses, so that when I stagger out at closing time I invariably token "horse" on encountering them on the pastures opposite. Since the conditions under which this occurs are systematic and uniform – the same thing would happen again, given similar animals, similar conditions of illumination, and similar levels of alcohol in the bloodstream – it will be a law that horsey-looking-cows-on-dark-nights-etc. cause tokenings of "horse". It will then follow from (R) that "horse" means horsey-looking-cow-on-a-dark-night. Or actually, not quite: there will be some sets of background circumstances in which horses do cause tokenings of "horse" (good conditions of illumination etc.). So the class of things which sustain a lawlike connection with tokenings of "horse" will actually be *disjunctive*: it is a law that horses-in-good-conditions-of-illumination or horsey-looking-cows-on-dark-nights-etc cause tokenings of "horse". Then it follows from (R) that the predicate "horse" actually has a disjunctive meaning: according to (R) it means *horses-in-good-conditions-of-illumination or horsey-looking-cows-on-dark-nights*. Thus, applying "horse" to horsey-looking-cows-on-a-dark-night will not be a misapplication, since anything which is a horsey-looking-cow-on-a-dark-night is *ipso facto* a horse-in-good-conditions-of-illumination or a horsey-looking-cow-on-a-dark-night. Again, there seems to be no scope left for incorrect application of the predicate. This problem is known as the *disjunction problem*.

According to Fodor, this is the main problem facing naturalistic accounts of meaning. One way of trying to solve this problem would be to alter (R) along the following lines:

(R*) "X" means Y if and only if it is a law that Y's cause tokenings of "X" under optimal conditions.

Of course, the optimal conditions would have to be specified purely naturalistically, in non-semantic and non-intentional terms. Now it is plausible that the optimal conditions for tokening the predicate "horse" do not include conditions under which it is dark, etc., etc. This would mean that since horsey-looking-cows-on-a-dark-night do not cause tokenings of "horse" under optimal conditions, we avoid the disjunction problem. (R*) does not deliver the conclusion that "Horse" means horse-in-good-conditions-of-illumination or horsey-looking-cow-on-a-dark-night, so that tokenings of "horse" which occur on encountering a horsey-looking-cow on a dark night could after all count as false or incorrect.

The problem, of course, is the characterisation of a suitable set of optimal conditions in non-semantic and non-intentional terms. Note that (R*) is basically the same idea as lay behind sophisticated dispositionalism. To say that sugar is disposed to dissolve in water is basically to say that it is a law that if sugar is placed in water it will dissolve. So

> Jones means horse by "horse" iff Jones is disposed to apply "horse" to horses under ideal conditions

amounts to the same thing as

> Jones means horse by "horse" iff it is a law that horses cause Jones to token "horse" under optimal conditions.

This means that the arguments we developed in §6.2–6.3 against sophisticated dispositionalism also apply to any attempt to solve the disjunction problem which relies on something like (R*).[20]

Fodor's aim is therefore to provide a naturalistic solution to KW's sceptical problem which does not require the formulation of naturalistic optimal conditions in the manner required by (R*). To this end he develops his "asymmetric-dependency" account of meaning.[21]

If we are not allowed to appeal to naturalistically specified optimal conditions, how can we distinguish between *correct* tokenings of "horse" in response to the presence of horses, and *incorrect* tokenings of "horse" in response to the presence of

horsey-looking-cows-on-dark-nights? Fodor's idea is that the latter sort of tokening is incorrect because the causal relationship between tokenings of "horse" and horsey-looking-cows is *asymmetrically dependent* upon the causal relationship between tokenings of "horse" and horses.

What does this mean? Consider the fact that horsey-looking-cows sometimes cause tokenings of "horse". This is dependent upon the fact that horses sometimes causes tokenings of "horse", since it is plausible that *if it were not the case that horses sometimes cause tokenings of "horse", it would not be the case that horsey-looking-cows sometimes cause tokenings of "horse" either*. And the dependence is asymmetric since it does not hold the other way round: it is false that if it were not the case that horsey-looking-cows sometimes cause tokenings of "horse, it would not be the case that horses sometimes cause tokenings of "horse". *"Horse" means horses, and not horses or horsey-looking-cows, because the causal relation between horsey-looking-cows and tokenings of "horse" is asymmetrically dependent upon the causal relation between horses and tokenings of "horse"*. In general, the causal relation between Y's and tokenings of "X" can be said to constitute the fact that "X" means Y if for every Z, the causal relation between Z's and tokenings of "X" is asymmetrically dependent upon it.

How plausible is Fodor's account? It is difficult to avoid sympathising with the following worry, raised by Crane:

> [It is difficult] to see how asymmetric dependence goes any way towards *explaining* mental representation. I think that the conditions Fodor describes probably are true of mental representations. But I do not see how this helps us to understand how mental representation actually works. In effect, Fodor is saying: error is parasitic on true belief. But it's hard not to object that this is just what we knew already. The question rather is: what is error? Until we can give some account of error, then it does not really help us to say that it is parasitic on true belief.[22]

But the objection is hard to assess. KW asked us to specify some fact which constitutes an expression's meaning one thing rather than another, and it seems that Fodor has specified such a fact,

in purely causal and nomological terms. What is the additional demand for explanation, and how is it motivated? What sort of explanation is being asked for, and what are the constraints on its acceptability? Unfortunately, we cannot pursue these questions here, but must end this section by noting that Fodor's attempted solution is far from obtaining universal acceptance among philosophers of mind and language.[23]

6.5 McGinn on normativity and the ability conception of understanding

In his book *Wittgenstein on Meaning*, McGinn makes a number of suggestions as to how KW's sceptical paradox might best be responded to. One suggestion is that viewing understanding as consisting in the possession of an *ability* or a *capacity* might provide us with a plausible straight solution to the sceptical argument. A move to viewing linguistic understanding in this way is quite explicitly apparent in the work of the later Wittgenstein, as the following quotes illustrate:[24]

> The grammar of the word "knows" is evidently closely related to that of "can", "is able to". But also closely related to that of "understands". ("Mastery" of a technique)[25]
>
> To understand a language means to be master of a technique.[26]
>
> Only in the practice of a language can a word have meaning.[27]

In short, facts about what I mean are constituted by facts about which capacities I possess. This solution, it is claimed, avoids the sorts of pitfalls which allegedly vitiate dispositional responses to the sceptical argument. In particular, where dispositionalism has difficulty accounting for the *normative* aspects of meaning and understanding, the capacity response does not:

> One important way in which capacities and dispositions can come apart is through the possibility of *mistake*: two people could both have the capacity to classify red things on the

> basis of their appearance, but one of them fails to do so when presented with a red object because he believes his senses to be functioning abnormally – he is then *able* to do what he is not *disposed* to do. Or again, I have the capacity to dial telephone numbers, but in some circumstances I make mistakes: it is not a necessary condition of possessing an ability that one *always* exercise it correctly. [Likewise, a] disposition is not sufficient for having an ability: we can conceive of set-ups in which a person is disposed to ψ in conditions C but he has not the ability to ψ – he gets the disposition from some source other than the possession of the ability.[28]

McGinn thinks that the capacity/ability response can avoid the problems faced by dispositionalism because it can fulfil a condition whose satisfaction is sufficient for accommodating the normativity of meaning. The condition is that we provide: "(a) an account of what it is to mean something at a given time, and (b) an account of what it is to mean the same thing at two different times".[29] And the capacity suggestion easily satisfies these desiderata: "to mean addition by '+' at t is to associate with '+' the capacity to add at t, and to mean the same by '+' at t* is to associate with '+' the same capacity at t* as at t".[30]

Now this conception of what is required for an account of normativity is obviously suggested by some aspects of KW's presentation of the sceptical argument: the challenge, recall, was, to cite some fact which would establish that what you *now* mean by "+" is the *same* as what you meant in the past. The sceptic "puts the challenge in terms of a sceptical hypothesis about a change in my usage", and the challenge is to produce a fact which rules out the hypothesis that such a change has in fact occurred. The challenge is set up in this way to allow the sceptical problem to be intelligibly formulated: in order to be able to converse with me, the sceptic allows that I presently mean addition by "+" and proposes a sceptical hypothesis which suggests that I meant something incompatible in the past. As Kripke himself puts it "Only past usages are to be questioned. Otherwise we will be unable to formulate our problem".[31]

However, Boghossian points out that the notion of normativity McGinn is working with here cannot be the notion of normativity at

stake in KW's discussion. This is apparent because it places no substantial constraints on the choice of a meaning-constituting fact:

> The requirement defined by McGinn could hardly act as a substantive constraint on theories of meaning. . . Any theory of meaning that provided an account of what speakers mean by their expressions at arbitrary times – however crazy that theory may otherwise be – would satisfy McGinn's constraint.[32]

For instance, the dispositionalist response – which KW attacks precisely because he claims it cannot leave room for the normativity of meaning – easily satisfies McGinn's condition. Since Kripke's discussion does not suggest that we can have no criteria for the transtemporal identity of dispositions, we might just as easily say: to mean addition by "+" at t is to associate with "+" the disposition to add at t, and to mean the same thing by "+" at t* is to associate with "+" at t* the same disposition as at t.

Clearly something has gone wrong in McGinn's characterisation of normativity, and Boghossian succinctly explains where. Properly viewed, normativity does not concern the relation between *past* meaning and *present* meaning, *but rather between meaning something by an expression at a time and the use which is made of the expression at that time*. What I mean now by an expression determines what uses I now make of that expression are correct or incorrect, and this is the normative constraint which the sceptic is challenging us to account for. As Boghossian puts it:

> The normativity of meaning turns out to be, in other words, simply a new name for the familiar fact that, regardless of whether one thinks of meaning in truth-theoretic or assertion-theoretic terms, meaningful expressions possess conditions of correct use.[33]

The question as to how conditions of correct use can be so much as possible is what is at issue, not the question concerning the possibility of transtemporal identity conditions for meaning.

This leaves us with a problem however: given that McGinn's reading of the normativity requirement was naturally suggested by KW's presentation of the sceptical argument in terms of a

hypothesis about a change in usage, and given KW's claim that this is the only way in which the sceptical problem can be intelligibly formulated, we have to find some new way of framing the sceptical problem consonant with the proper understanding of the normativity requirement. Fortunately, this is not very difficult. As Boghossian puts it:

> Having a meaning is essentially a matter of possessing a correctness condition. And the sceptical challenge is to explain how anything could possess *that*.[34]

The challenge – to explain what it is for a linguistic expression to possess conditions of correct usage – can be set out without even mentioning the sceptic. Bringing in the sceptic is only a device – a dispensable device – for setting out the challenge. We should not take the dialogic setting in which Kripke introduces the challenge too seriously:

> The constitutive problem about meaning – how could there so much as be a correctness condition – can be stated quite forcefully without the actual provision of a convincing global reinterpretation of a person's words.[35]

Once we appreciate this, we can see that the sceptical challenge can be framed in a way which avoids McGinn's confusions about the nature of normativity.

Where do these considerations leave McGinn's capacity/ability solution to KW's sceptical problem? McGinn's claim that the capacity solution could avoid the worries faced by dispositionalism rests upon a mistaken conception of the normativity requirement, so to that extent McGinn's suggestion is severely weakened. Moreover, there are other worries which suggest that the capacity suggestion cannot constitute an adequate response to KW's sceptical challenge.

Wright says:

> If we ask, what capacity is constitutively associated with an understanding of "green"?, the natural answer is: the capacity to use the word correctly. And here "correctly" means, roughly: in ways which are appropriately sensitive to its meaning.[36]

The mere claim that understanding consists in the possession of a capacity thus appears to get the cart before the horse, at least so far as a response to KW's sceptic is concerned. For example, in the case of "green", a competent speaker possesses the capacity to suit his use with the expression to its meaning. But this – the meaning of green – is precisely what KW is challenging us to characterise! So we need to go further and say what the capacity in question is a capacity to do:

> Well, what then is the capacity, allegedly constitutive of an understanding of "green", a capacity, most fundamentally, to *do*? McGinn's discussion contains, so far as I have been able to see, no clear suggestion about how a proponent of his "straight solution" should respond to this question.[37]

This is perhaps a *little* unfair on McGinn, as he does attempt to characterise the relevant capacities in more detail:

> In Kripke's favourite example, the suggestion would be that it is the concept of *addition* that I exercise when I do computations involving "+" and not the concept of *quaddition*, because the capacity that gets brought to bear is the capacity to *add* and not to *quadd*, where the former capacity is conceived as a capacity to recognise what is the *sum* of pairs of numbers.[38]

McGinn's suggestion is thus that meaning addition by "+" consists in associating with "+" the capacity to *add*, and meaning quaddition by "+" consists in associating with "+" the ability to *quadd*. Unfortunately, this is a very weak response to KW's challenge, for KW can simply reformulate the challenge in terms of capacities: in virtue of what does Jones possess the capacity to *add*, rather than the capacity to *quadd*? All of the arguments of §5.1 can be brought back in to suggest that there is no plausible answer to *this* question.[39]

The problem McGinn's suggestion faces is thus as follows. Just as a speaker will have a range of dispositions associated with a given expression, he will also have a range of capacities associated with it. So McGinn must give us some account of which capacities are constitutive of understanding in a way which does not simply put the cart before the horse and which does not fall

prey to a simple reformulation of KW's original sceptical problem. And it seems that the provision of such an account is no easier than the dispositionalist's task of spelling out a suitable set of optimal conditions:

> The relevant capacities have to be singled out somehow, and the problem is structurally reminiscent of – indeed, in no way interestingly different from or more promising in outcome than – the problem, confronted by the dispositional response, of saying what puts a particular disposition in the idealized, meaning-constituting class which it is obliged to define.[40]

6.6 Wright's judgement-dependent conception of meaning

Wright's response attempts to meet KW's sceptical challenge by identifying and challenging an extremely important methodological assumption which underlies the presentation of that challenge. That assumption is this: that if semantic facts are to be legitimately countenanced then they must be *reducible* to some *other* class of facts which can be fully and adequately described *without the use of semantic and intentional notions (i.e. notions involving meaning)*. Wright's response questions this assumption: when faced with a demand for an account of the kind of fact that constitutes my meaning addition by "+", why is it not sufficient simply to reply "my meaning addition by '+'"? As Wright says

> [I]t cannot *always* be possible to justify a presumed genre of knowledge "from without" in the way the sceptic is here demanding.[41]

This means that the onus is on KW to provide us with a *positive* argument against non-reductionism about meaning (the view that facts about meaning are irreducible and *sui generis*). As we saw in §5.1, KW does gesture in the direction of such an argument. Recall that KW writes, of the claim that meaning such and such by an expression is an irreducible state, that:

> Such a move may in a sense be irrefutable, and taken in an appropriate way Wittgenstein may even accept it. But it seems desperate: it leaves the nature of the postulated primitive state – the primitive state of meaning addition by "plus" – completely mysterious. It is not supposed to be an introspectible state, yet we are supposedly aware of it with some fair degree of certainty whenever it occurs. For how else can each of us be confident that he *does*, at present, mean addition by "plus"?[42]

How can non-reductionism respond to this charge of desperation and mystification? Is it really mysterious and desperate to postulate the existence of a kind of state with the sorts of properties possessed by our intuitive conception of meaning? McGinn responds that, far from being mysterious and desperate, the properties in question are possessed by certain of our everyday psychological states, such as beliefs, thoughts, intentions, hopes, and so on. For example, introspection reveals that there is no state of consciousness, with its own affective phenomenology, which could *be* a given intention, and yet we are ordinarily taken to have *authoritative* and *non-inferential* access to the contents of our own intentions. Suppose I ask you whether you intend to travel to London next weekend. Your answer, whatever it is, will normally be given automatically: you will not have to *infer* that you have (or don't have) the intention to visit London from some other fact about you (say, from your behaviour). Moreover, your judgement that you intend (or don't intend) to visit London is authoritative in the sense that it is ordinarily taken to be the highest court to which one can appeal over the matter as to what you intend or do not intend to do. It is only *ordinarily* the highest court of appeal in such matters, because there are circumstances – for example, when there is independent evidence that you are self-deceived, or lying, or have misunderstood the question – in which it can be overturned. But the important point is that the onus is on someone who wants to overturn your judgement to bring such evidence forward: in the absence of this sort of evidence, you do not have to *justify* your judgement as to what you intend.[43] Intention, moreover, displays the same sort of "disposition-like theoreticity" as our intuitive conception of meaning: just as the ascription of an

intention to a person may be withdrawn if he behaves in certain ways in the future, the ascription of understanding to him may be withdrawn if he applies the expression in certain bizarre ways in future situations. How can the properties possessed by our intuitive conception of meaning be mysterious and desperate when they are possessed by a whole range of psychological states which figure constantly in our everyday mental lives and in our explanations of each other's actions? McGinn thinks these observations sufficient to rebut the sceptical argument at this point:

> How to give a philosophical *theory* of this kind of knowledge is of course a difficult and substantial question, but the lack of a theory of a phenomenon is not in itself a good reason to doubt the existence of the phenomenon. I therefore see no mystery-mongering in the claim that there are primitive non-experiental states which display a distinctive first-person epistemology.[44]

It is difficult not to agree with Wright's claim that this reply of McGinn's is, as it stands,

> about as flagrant an instance of philosophical stone kicking as one could wish for.[45]

In order to provide a satisfying response to the sceptic we may not need to provide a detailed philosophical *theory* of the nature of our first-person knowledge of meaning, but we surely do require at least a *sketch* of the epistemology of meaning, a sketch which would show how a state could possess both (a) a first-person epistemology of the sort described and nevertheless (b) "have to answer . . . to what one says and does in situations so far unconsidered". The feature (a) pulls us in the direction of a *Cartesian* conception of meaning. The first-person epistemology would be explained if we could view understanding as some qualitative introspectible state possessing its own distinctive phenomenology: the non-inferential nature of our knowledge of our intentions would be explained because when we possess an introspectible state with its own qualitative phenomenology we do not have infer this, and the first-person authority of intention would be explained because

we have special privileged access to our conscious, introspectible states. But, of course, KW has shown that that sort of view of meaning is a non-starter. The feature (b) pulls us in the direction of a *dispositional* conception of meaning. Just as an ascription of brittleness to a glass has to answer to how that glass would behave (crucially, whether it would shatter) in unactualised situations, so an ascription of understanding to a speaker has to answer to how that speaker would go on to apply the relevant expression (certain patterns of application would lead us to withdraw the original ascription of understanding, just as certain patterns of behaviour by the glass would lead us to withdraw the ascription of brittleness). Taking understanding to be constituted by a disposition would explain this. But, of course, the dispositional conception of meaning and understanding is also thrown into doubt by KW's arguments. These two features thus seem to pull in two different directions, and the same ambivalent pull permeates the psychological states whose first-person epistemology McGinn adverts to in his attempted response to KW. So, the upshot is that *a non-reductionist response to the sceptical challenge is not acceptable unless it can be shown how it accommodates both of these apparently incompatible properties of meaning – its first-person epistemology and "disposition-like theoreticity" – and in a way which avoids the pitfalls both of crude Cartesianism and the dispositional conception of meaning*.[46] McGinn makes no attempt at all to do this, and his non-reductionism is inadequate because of it.[47] Wright, however, does attempt to develop a form of non-reductionism which explicitly takes on this task. Wright argues that predicates such as "means addition by '+'" are *judgement-dependent*, and that viewing them as judgement-dependent allows us to explain *both* the first-person epistemology of meaning and its "disposition-like theoreticity". We must now explain Wright's conception of judgement-dependence.[48]

Suppose that we are considering a particular region of discourse. Consider the opinions formed by the practitioners of that discourse, formed under conditions which are, for that discourse, *cognitively ideal*. Call such opinions *best opinions*, and the cognitively ideal conditions the *C-conditions*. To take colour discourse as an example, if we consider the predicate "red", the C-conditions

would be "conditions of illumination like those which actually typically obtain at noon on a cloudy summer's day out of doors and out of shadow". We cannot find conditions *better* than these for determining whether an object is red.[49]

Suppose we find that the best opinions formed by the practitioners *covary* with the facts about the instantiation of a predicate "red", so that it is true that if the C-conditions obtain, a suitable subject will judge that x is red if and only if x is red. Wright suggests that there are two broad ways in which we might seek to explain this covariance. On the one hand, we might take best opinions to be playing at most a *tracking* role: best opinions about redness are just extremely good at *tracking* independently constituted facts about redness. In such a case, best opinion plays merely an *extension-reflecting* role, serving merely to *reflect* the independently determined extension of "red". On the other hand, we might try to explain the covariance of best opinion and fact by assigning to best opinion an altogether different sort of role. Rather than viewing best opinion as merely tracking the facts about the extension of "red", we can view them as themselves *determining* that extension. Best opinion, on this sort of view, does not serve merely to track independently constituted states of affairs which determine the extensions of colour predicates: rather, best opinion serves to *determine* those extensions and so to play an *extension-determining* role. When the covariance of best opinion and the facts about the instantiation of the central predicates of a region of discourse admits of this latter sort of explanation, the predicates of that region are said to be *judgement-dependent*; when it admits only of the former sort of explanation, the predicates are said to be *judgement-independent*.[50]

How do we determine whether colours, for example, are judgement-dependent? Wright's discussion proceeds by reference to what he terms *provisional equations*. These have the following form:

(PE) $(\forall x)$ (C $\rightarrow$ ((A suitable subject S judges that x is red) $\leftrightarrow$ x is red))

where "C" denotes the cognitively ideal C-conditions. Redness is then said to be judgement-dependent if and only if the provisional equation is *a priori* true, if the C-conditions can be specified *non-trivially* and in such a way that the question as to whether they are satisfied is *independent* of facts about redness, and if there is no better explanation as to why the covariance between best opinion and fact is *a priori* and substantial in this way.[51]

Wright now argues that in the case of self-ascriptions of intention best opinions can be viewed as playing an extension-determining role. The C-conditions are roughly as follows: *the subject S is not lying, is prey to no material self-deception, is not making a simple slip of the tongue, has an adequate grasp of the concepts requisite for the expression of the intention, and is adequately attentive to the question of the content of his intention*. Under these conditions a speaker's first-person avowal of an intention is authoritative, in the sense that there is ordinarily no higher authority to which we can appeal in order to overthrow his judgement that he so intends. Rejecting the subject's opinion requires showing that at least one of these conditions is not satisfied in the given instance. Equally, when these conditions obtain, the subject's opinions about the contents of his intentions, and the truth about the contents of the relevant intentions, match perfectly. And, crucially, when the C-conditions are specified in this way the provisional equation INTENTION is plausibly *a priori* and non-trivial:

INTENTION: $(\forall S)\ (C \rightarrow ((\text{Subject S judges that he intends to } \varphi) \leftrightarrow (\text{he intends to } \varphi)))$

There appear to be no worries about INTENTION satisfying the conditions required for intentions to be judgement-dependent: it is *a priori* because all that is required to know the truth of INTENTION is mastery of the concept of intention, the conditions don't seem to amount to a "whatever it takes" clause and so are not trivially specified, and the satisfaction of the C-conditions is logically independent of the truth of "S intends to φ". There

is no better explanation of this substantial *a priori* covariance between best opinion and fact unless "full-blown Cartesianism" is plausible – and as we saw above Wright agrees with KW that full-blown Cartesianism is not an option.[52]

If Wright is justified in his claim that self-ascriptions of intention are judgement-dependent the payoff is as follows. Firstly, a solution to the problem of the first-person epistemology of intention seems to be immediate: a subject's knowledge of the contents of his own intentions is non-inferential and authoritative because it is his *own* judgements about the contents of his own intentions that determine the extensions of the predicates which self-ascribe those intentions. And such knowledge is easily accomplished because it is easy enough to get oneself into the appropriate conditions. We also appear to have a nice solution to the problem of the relationship between a subject's possession of an intention and the – so far unactualised – behaviour that serves in part to determine whether he possesses that intention:

> Suppose a subject believes that he has formed an intention to continue the series – 2, 4, 6 . . . – in a certain way and he writes – 8, 10, 12. He need have given no thought yet to what he will write after "12", but when he comes to that point he confidently continues – 14, 16, 18 – and judges that he is following out his former intention, that he is, in Wittgenstein's phrase, "going on in the same way". Wright's idea is that just as his original judgement that he had that particular intention constitutively determined that he did indeed have that intention, so his later judgement that writing "14, 16, 18" implemented rather than frustrated that intention constitutes the fact that such behaviour does implement that particular intention. So best opinion . . . as to what current behaviour implements a former intention determines what behaviour implements that former intention.[53]

So Wright accounts for the fact that the ascription of an intention to someone has to answer to how they would behave in as yet unactualised situations: it explains the "disposition-like theoreticity" of intention.

We can thus explain the features which might incline us, respectively, either in the direction of a Cartesian account of intention, or in the direction of a purely dispositional construal, without succumbing to the problems faced by either of those types of theory. Suppose that Wright's account of intention is successful. What are the prospects for giving a similar account in the case of meaning and understanding? If meaning such and such by an expression can be taken to consist in the possession of a certain sort of intention then the application of Wright's judgement-dependent account of intention to the case of meaning will be straightforward. But even if this is not the case, we can still construct something like the story we constructed about intention for the case of meaning. As Wright says, responding to the sceptical argument via an account like his

> does not require construal of meaning as a kind of intention; it is enough that the concepts are relevantly similar – that both sustain authoritative first-person avowals, and that this circumstance is to be explained in terms of [judgement-dependence].[54]

How plausible is Wright's claim that self-ascriptions of meaning and intention are judgement-dependent? This question has received a lot of discussion in the literature, with many philosophers raising doubts about whether intention is plausibly viewed as judgement-dependent. We'll end this section by looking briefly at one objection to Wright's account, that proposed by Boghossian.[55]

Consider again the kind of provisional biconditional central to Wright's account:

> If C then (S believes that he means addition by "+" if and only if he means addition by "+").

The problem is this. We want to view S's best beliefs as determining the extension of predicates such as "S means . . . by '+'". But as Boghossian argues, we cannot view S's best *beliefs* (or *judgements*

or *opinions*) as playing an extension-determining role with respect to ascriptions of meaning without violating the condition that the question as to whether the C-conditions are satisfied is independent of prior facts about meaning. This is because we have to take for granted at least the *content* of the beliefs which figure in the provisional equation. Best opinions cannot constitutively determine the extension of predicates ascribing meaning or content, because we have to assume some *prior* determination of the content or meaning possessed by those best opinions themselves. It follows that meaning cannot be judgement-dependent, since the judgement-dependent account itself presupposes prior facts about meaning. As Boghossian puts it himself:

> It is already clear that there is a serious difficulty seeing how facts about mental content could conceivably satisfy the stated requirements on judgement-dependence. For it is inconceivable, given what *judgement-dependence* amounts to, that the biconditionals in the case of mental content should satisfy the requirements that their left-hand sides be free of any assumptions about mental content. For, at a minimum, *the content of the judgements* said to fix the facts about mental content have to be presupposed. And that means that any such biconditional will always presuppose a constitution of mental content quite independent of constitution by best judgement.[56]

No doubt, Wright has a reply to Boghossian's worry.[57] But at the very least, it shows that Wright's solution to the sceptical paradox, appealing though it is, faces considerable problems of its own. We'll leave the evaluation of Wright's account to the reader, and proceed in the following three sections to consider solutions arguably advocated by Wittgenstein himself.

6.7 Pettit's "ethocentric" account

In this section, I'll outline a sophisticated and ingenious response to KW's sceptical paradox, a response which has been developed in a series of works by Pettit. I'll then raise three challenges for the account.

6.7.1 Pettit's response to the sceptical paradox

Pettit identifies objective and subjective aspects to the notion of a rule. On the objective side, rules are normative constraints that operate over an open-ended and potentially infinite set of inputs. Take the simple rule governing addition. Given two numbers as inputs, this rule yields a number as output. Given 1 and 2 as inputs, it yields the number 3 as output, given 17 and 34 as inputs, it yields 51 as output, and so on. This rule is a normative constraint insofar as in operating as outlined, it yields verdicts about the correctness and incorrectness of answers to arithmetical queries: given the usual meanings of the relevant expressions, "3" is the correct answer to "1 + 2 = ?", "4" would be incorrect, "51" is the correct answer to "17 + 34 = ?", "57" would be incorrect, and so on. When thinking of a rule as an objective constraint in this way, we can think of it as a rule-in-extension – as an infinite set of ordered triples {(1, 1, 2), (1, 2, 3), . . ., (17, 34, 51), . . .} or as a rule-in-intension – an abstract object, such as a mathematical function, which determines the correct output for any given pair of inputs. (In the case of a one-place predicate, we could think of the rule-in-extension as a set of individuals (those to whom the predicate applies) and the rule-in-intension as a property or universal (in virtue of instantiating which the relevant individuals belong to the rule-in-extension).)

On the subjective side, Pettit highlights three aspects of the notion of a rule. Rules have to be identifiable independently of any particular application. For example, I have to be able to identify the rule for addition independently of giving the response "51" to the query "17 + 34 = ?", otherwise I could not view my grasp of the rule for addition as something with which I am attempting to conform in responding to arithmetical queries. The rule must also be directly but fallibly readable. It must be directly readable in the sense that it instructs me how I ought to answer arithmetical queries in a way which ultimately does not rest upon interpretation (upon the application of a further rule), and it must be fallibly readable in the sense that in any given case it must at least be possible for me to misapply the rule.

Overall, then, for Pettit rules are indefinitely normative constraints that are independently identifiable and directly and fallibly readable.

Pettit frames the challenge posed by KW's sceptic as consisting in an argument to the effect that nothing could simultaneously satisfy both the objective and subjective elements in the characterisation of a rule. The entities which plausibly satisfy the objective aspect – rules-in-extension and rules-in-intension – seem incapable of satisfying the subjective aspects. The rule-in-extension is an infinitely large set, and as Kripke puts it "the infinitely many cases of the [addition] table are not in my mind for my future self to consult";[58] for the rule-in-intension, there is no good story relating my mind to a rule-in-intension that singles out addition, and not quaddition, as the referent of "+".[59] Likewise, entities capable of satisfying the subjective elements are incapable of serving as indefinitely normative constraints: any finite set of examples, or finite state of my mind, will be consistent with the rule for quaddition being the normative constraint governing my responses to queries involving "+". Nor will dispositionalism yield a satisfactory answer to KW:

> [T]he dispositional theory does nothing to explain how in following a rule I am directly but fallibly guided by something that determines the right response in advance. A disposition may determine what I do but it cannot provide this sort of guidance.[60]

Thus, with nothing capable of satisfying both the objective and subjective elements, the notion of a rule, as Kripke puts it, "vanishes into thin air".

Against this, Pettit seeks an explanation of how a practice of following rules might have come about in a way that shows how the objective and subjective elements in the characterisation of a rule are not irreconcilable. In particular, Pettit seeks to explain how a finite set of examples might be legitimately regarded as representing a determinate rule to human rule-followers. Central to Pettit's story is the distinction between *instantiation* and *exemplification*. Instantiation is a two-place relation that holds between finite sets of examples on the one hand and rules on the other. For any finite set of examples there are an infinite number of rules which it can be taken to instantiate, but matters stand differently

with exemplification, which is a three-place relation which holds between sets of examples, rules, and agents: "Although any finite set of examples instantiates an indefinite number of rules, for a particular agent the set may exemplify just one rule".[61] Clearly, the number of ways I am inclined to "go on" from a particular set of examples will be less than the number of rules which that set of examples may be regarded as instantiating, but this will still fall far short of securing anything like the required degree of determinacy. Pettit's key idea is that we can achieve this by considering, not my extrapolative inclination neat, but rather my inclination to "go on" in *favourable conditions*, where favourable conditions are characterised as those which survive a process involving the elimination of interpersonal and intertemporal inconsistency of application. For example, suppose that now, in the daylight, I am inclined to apply the predicate "red" to the ripe tomato in front of me, whereas earlier, under the ultra-violet light, I was not. This is an inconsistency which I am disposed to resolve, and I attempt to do so by discounting at least one of the applications by deeming at least one of the circumstances in which the applications were made to be unfavourable. Likewise for discrepancies between the way I am inclined to apply the term and the way you are inclined to apply it. If these processes of resolving discrepancies result in a convergence on a set of conditions in which the term is agreed to be applicable, then we can view our applications as rule-governed and as genuinely expressive of grasp of a concept. Pettit dubs this an "ethocentric" account of rule-following, as the Greek word "ethos" may be adopted to connote the habit or practice of subjecting our extrapolative inclinations to this sort of process of rooting out discrepancies of application across persons and times.

Given that an inclination to use an expression in a certain way is just a disposition to use it, and given the emphasis on favourable conditions in the ethocentric account, it might seem that we have here simply another variant of the kind of sophisticated dispositionalism discussed in §6.2 above. And given this, one might wonder further whether the ethocentric account falls prey to the sorts of worry adumbrated in §6.2 of the sophisticated dispositionalist's invocation of ideal conditions. Pettit, however, explicitly stresses that his ethocentric account is not a version of sophisticated

dispositionalism. The sophisticated dispositionalist aspires to give us an account of the sort of fact in which Jones's following the rule for addition consists: in Pettit's terms, a "reductive analysis" of rule-following. Such an account may indeed fall prey to KW's objections. However, Pettit writes, he is not attempting to give such an analysis of rule-following but rather an *explanatory* account of how rule-following is possible, a genealogy of how rule-following might get going. As such, KW's worries fail to apply to the notion of favourable conditions as it appears in the ethocentric account, and the account's capacity to survive this attack effectively shows how it can square with the objective component of the notion of a rule, the idea that it is a normative constraint operating over an open-ended and indefinitely large number of inputs. Moreover, since no particular set of examples has to figure in the set giving rise to our original extrapolative inclinations, the independent identifiability of rules is secured. Likewise, we can secure direct readability, in a way which simultaneously makes clear how the account differs from reductive dispositionalism *vis-à-vis* the issue of normative guidance:

> If the rule is identified by inclination, then of course there is no difficulty about how it is directly readable. The inclination serves on our proposal, not just to identify the rule, but also to prompt the agent's responses: it has a dual function. The individual will read off the requirement of a rule in a new case by letting her inclination lead her, as with the simple rule, or by applying other rules whose requirements she ultimately reads off in that way. No mode of reading a rule could be more direct.[62]

Likewise, fallible readability is assured: since for all I know the conditions under which I am currently applying a term may subsequently be discarded by the ethocentric process of resolving discrepancies, "I will have to recognise in any instance of reading the rule that for all [I] know [I] may be forced *ex post* to judge that [I] got it wrong".[63]

Pettit's explanatory account of rule-following is undoubtedly sophisticated and far richer than the brief summary above can hope to capture. However, while I am unable to offer any decisive rebuttal of the ethocentric account I will now proceed to develop three challenges to it.

6.7.2 Circularity

Although Pettit describes the ethocentric view as an explanatory account of how it is possible for us to follow rules, as opposed to an analysis or an account of the fact which constitutes my following a rule, this does not mean that the account cannot be criticised on grounds relating to circularity. Just as a reductive analysis can fail if the analysans makes use of the analysandum, an explanation can fail if the explanans makes use of the explanandum: if we need to appeal to that which is to be explained in order to deploy that which aspires to do the explaining, the latter will fail to provide an adequate explanation of the former.[64] So Pettit's genealogy of rule-following must avoid this damaging sort of circularity.[65] What does this amount to in concrete terms? A key consideration here is the relationship between rule-following and linguistic meaning and understanding. We noted earlier that one can think of a rule-in-intension (in the case of a rule governing a predicate) as a property or a universal, and if we think of a predicate as having a property or a universal as its semantic value, the connection is immediate: the conditions for the correct application of the predicate will be determined by the relevant facts about which objects do or do not instantiate the associated property or universal. So, an explanatory account of how we might follow rules should not make use of facts about the semantic values of predicates or of expressions generally: to do so would be to rely on as an explanatory factor an instance of that which we are aspiring to explain.

I will now argue that the ethocentric account appears to do just this. The root of the problem is an assumption which, by Pettit's own lights, needs to be made in order for the ethocentric story about the resolution of disagreement to make sense. If I apply "red" to an object when I view it in conditions C and you refuse to apply "red" to the same object when you view it in conditions C*, according to the ethocentric account we will attempt to discount at least one of the applications and deem either C or C* to be unfavourable *vis-à-vis* the application of "red" – and all of this, crucially, on the assumption that "the term has the same semantic value on all sides", that "the word has the same semantic value in our mouths at different times and in the mouths of different people", and that "I take [you] to use the same word to express the

same belief [as me]".[66] An assumption along these lines is tantamount to an assumption about rule-following: in particular, that in the relevant exchanges the relevant word is governed by the same rule for different people or for the same person at different times. Without this assumption the ethocentric account is hamstrung, as we can see if we consider another example. If you apply "bank" to the branch of ANZ across the street and I do not (perhaps preferring instead to apply it to the grassy slope alongside the Water of Leith), it may not be appropriate for us to deem the conditions in which one of us made the relevant application as unfavourable. If (as seems likely) you and I attach different semantic values to the expression as used on the relevant occasions, the ethocentric story about the identification of favourable conditions will not even get underway. Thus, in order for the ethocentric account to get started in any particular case we need to make assumptions about which rules govern which expressions, and this disqualifies the ethocentric story from being a fully general explanation of how rule-following might be possible. Given that the ethocentric account is intended as a naturalistically respectable explanation of the possibility of rule-following and meaning, therefore, the challenge to Pettit is to provide a naturalistically respectable account of the possibility that two people (or two time slices of the same individual) might attach the same semantic value to a given expression.[67]

6.7.3 Normative guidance

We saw earlier that Pettit rejected reductive dispositionalism on the grounds that it could not capture the idea that rules provide *normative guidance* to those who follow them. The ethocentric account was deemed to do better on this score because it accorded a dual function to our inclinations, which were "not only to prompt the agent's responses, but also to make salient the rule she intends to follow: the rule that, given the inclination they engender, a certain set of examples can exemplify".[68] However, we might wonder whether this is persuasive. To be sure, *if* it is legitimate to think of an inclination – a disposition by another name – as making salient the rule which

an agent intends to follow, then we can think of the inclination as enabling the agent to be guided by the rule in addition to prompting her responses. Likewise, though, *if* it is legitimate to think of a reductively characterised disposition as making it the case that an agent is following a particular rule, then we can think of the fact that the agent has the relevant disposition as enabling that agent to be guided by the rule. But in neither case is the possibility of normative guidance properly ensured: the normative guidance constraint is intended to constrain *whether* it is legitimate to think of a reductively characterised disposition as constituting the fact that an agent is following a rule – and likewise whether it is legitimate to think of an inclination (again, a disposition by another name) as making salient the rule which an agent intends to follow. And just as it may be argued that the relationship between a disposition and one of its manifestations is not the right kind of relationship for a disposition to be viewed as constituting the fact that an agent is following a rule, it may be argued with equal force that the relationship between an inclination and the responses which it prompts is not the right kind of relationship to allow us to view those responses as guided by a rule which the inclination makes salient. So the ethocentric account only appears to satisfy the normative guidance requirement because Pettit has put the cart before the horse: a manoeuvre which would allow the reductive dispositionalist to reinstate his view. We thus leave Pettit with a second challenge: formulate the normative guidance requirement and show – in a non-question-begging way – that the ethocentric account satisfies it and that reductive dispositionalism fails to do so.[69]

6.7.3 The quasi-reality of rule-following?

Pettit represents himself as giving a "non-sceptical" solution to KW's sceptical paradox. It is not clear whether he means a straight solution in Kripke's sense, but if so there are some potential problems with the idea that the ethocentric account could lay claim to that appellation, even if it is successful. I'll suggest that the ethocentric account is in fact more akin to a form of Simon Blackburn's quasi-realism than realism about rule-following, and

that when we see this we can also see that the ethocentric account, as it stands, is damagingly incomplete.

In order to see this, I'll give a brief reminder of the general shape of Blackburn's quasi-realism, using quasi-realism about moral thought and talk as an example.[70] For Blackburn, the main aim of a metaethical theory is to give an account of the nature of moral judgement. A realist theory is one in which the starting point for the explanation is the notion of a moral state of affairs. Once moral states of affairs have been characterised, the realist can explain moral judgement as involving the expression of beliefs about the obtaining of those states of affairs. Vastly oversimplifying, the main difficulty faced by moral realism is that we face a dilemma when we attempt to characterise moral states of affairs. If we regard moral states of affairs as natural, as the sorts of states of affairs that are the subject matter of the natural and social sciences, then we run into the problem (highlighted in Moore's open question argument) that moral judgements appear to be intrinsically normative in a way that is apparently not shared by beliefs about the obtaining of natural states of affairs.[71] On the other hand, if we regard moral states of affairs as non-natural, we face metaphysical and epistemological problems of the sort highlighted in Mackie's "argument from queerness".[72]

The quasi-realist response to this is not, with the emotivist, to *deny* that moral judgements express beliefs or that there are such things as moral states of affairs. Rather, at the outset of our explanatory account of moral judgement, we do not help ourselves to the assumption that moral judgements express beliefs or that there are such things as moral states of affairs; instead, we aim to explain moral judgement on a more austere basis than the realist, but also, at the conclusion of the explanatory account, to recover philosophically hygenic notions of moral belief and moral states of affairs. The austere basis includes natural states of affairs, non-cognitive sentiments directed towards natural states of affairs, and beliefs about natural states of affairs – and crucially, does not include either moral states of affairs or beliefs about them. In the quasi-realist story, we begin with the assumption that moral judgements express non-cognitive sentiments, then – via the construction of a solution to the Frege–Geach problem and of an adequate notion of moral truth – we earn the right to the idea

that moral judgements express beliefs, some at least of which are true.[73] Because at the end of the story we have earned notions of moral truth and moral belief, Blackburn describes the position as "realist"; but because moral states of affairs do not figure among the explanantia, it is at most "quasi" realist.

We saw above that Pettit sees the ethocentric view of rule-following as eschewing any attempt at reductive analysis, as making no attempt to identify facts in which rule-following consists. Instead, the ethocentric view attempts to provide an explanatory account of how rule-following might be possible without appealing to any such facts. Blackburn describes the quasi-realist project in similar terms:

> [T]he problem is one of finding room for ethics, or placing ethics within the disenchanted, non-ethical order which we inhabit, and of which we are a part. "Finding room" means understanding how we think ethically, and why it offends against nothing in the rest of our world-view for us to do so. It does not necessarily mean "reducing" ethics to something else.[74]

Where Blackburn speaks of a "disenchanted, non-ethical order", Pettit speaks of "atoms in the void" and "a universe constructed out of the austere materials charted in the sciences".[75] And just as Pettit seeks an explanation of rule-following without identifying the fact that an agent is following a rule with any natural fact, Blackburn seeks to give a naturalistic explanation of moral judgements which does not involve the identification of moral states of affairs with any natural states of affairs.[76]

This suggests that the ethocentric account of rule-following should be regarded as on a par with Blackburn's quasi-realism, insofar as it aspires to give an account of our practice of regarding ourselves and others as following rules in which putative facts about rule-following do not figure as explanatory factors. Indeed, we can see some specific parallels between Pettit's account and the view about rule-following adumbrated by Blackburn himself.[77] Blackburn writes:

> I share Kripke's view that whatever dispositions we succeed in identifying they could at most give standards for selection of a function which we mean. They couldn't provide us with an

> account of what it is to be faithful to a previous rule. It is just that, unlike Kripke, I do not think dispositions are inadequate to the task of providing standards.[78]

Just as Blackburn holds that dispositions are adequate to the task of providing standards for selection of a function which we mean (or a rule which we follow), Pettit holds that our suitably normalised inclinations are adequate to the task of making salient the rule that we intend to follow. And just as Blackburn suggests that nonetheless we cannot give an account of what it is to be faithful to a rule in terms of those dispositions, Pettit does not attempt to identify the fact that we are following such and such a rule with facts about suitably normalised extrapolative inclinations.

In short, the ethocentric account should be regarded as a form of quasi-realism, rather than a form of realism, about rule-following. And this observation points us in the direction of an aspect in which the ethocentric account is incomplete. We saw that, for Blackburn, the main task in the moral case is that of explaining the nature of moral judgement. Likewise, in the case of rule-following, the task is to explain our practice of judging ourselves and others to be following rules. Pettit agrees that it would be a mistake to embark on this task by attempting to specify facts which constitute the fact that such and such an agent is following such and such a rule. However, whereas in the moral case (for example) Blackburn suggests that we start out making only the assumption that moral judgements express non-cognitive sentiments and work out from there – via solving the Frege–Geach problem and inter alia constructing a notion of moral truth – these constructive aspects of the quasi-realist project appear to be entirely lacking in Pettit's account. Our closing challenge to Pettit is, therefore: provide the missing "constructive" aspects of the quasi-realist account in the case of rule-following (what, initially, do we take judgements about rule-following to express and how do we get from that starting point to a notion of truth applicable to them?), or provide an explanation of how the ethocentric account of rule-following is not, after all, a form of quasi-realism.

6.8 Wittgenstein's dissolution of the sceptical paradox?

The sceptical paradox about sense which has been at the centre of this chapter was inspired, as we have seen, by Wittgenstein's later writing, especially certain sections of *Philosophical Investigations*. KW advances the sceptical problem, and attempts to neutralise its effect by developing a "sceptical solution", an account which tries to legitimise our practice of ascribing meaning and understanding in the absence of meaning-constituting facts. But what is KW's relationship to the actual, historical Wittgenstein? Does KW accurately reflect his views, and if not, what *are* his views on the topics we have been considering? The interpretation of Wittgenstein's later works is a matter of extreme controversy in contemporary philosophy, and any interpretation is likely to face serious objections from those wishing to defend their own favoured interpretations. But having said that, one must make a choice, and, arguably, the interpretation of Wittgenstein which sits best with a reading of the texts is that which has been proposed by John McDowell. It would be impossible to defend this exegetical claim in the present work, so I limit myself here to description of McDowell's Wittgenstein's views, and to remarks about their philosophical import.

Recall that Wright sees Wittgenstein as pointing to a problem which any non-reductionist conception of meaning and understanding must face: how is it possible for there to be states of mind, such as meaning and intention, which are non-inferentially and authoritatively known to their possessors, but which also display "disposition-like theoreticity"? However, although Wittgenstein's writings bring this question sharply into focus, "it is probably in vain to search Wittgenstein's own texts for a concrete positive suggestion"[79] about how to answer it. This is a reflection of the later Wittgenstein's philosophical method, which is conditioned by a mistrust of such "constitutive" questions. Wittgenstein thought that philosophy's job was not to give *explanations*: *a fortiori* it was not philosophy's job to explain how there could be states of mind with the features mentioned.

> Philosophy may in no way interfere with the actual use of language; it can in the end only describe it. For it cannot give it any foundation either. It leaves everything as it is.[80]
>
> Philosophy simply puts everything before us, and neither explains nor deduces anything. [81]

What, then, should we say about our practice of ascribing meaning and understanding, if we are not to attempt to explain how it is possible? Wittgenstein's view is that the job of philosophy is only to dispel possible misunderstandings of these practices:

> The results of philosophy are the uncovering of one or another piece of plain nonsense and of bumps that the understanding has got by running its head up against the limits of language.[82]
>
> What is your aim in philosophy? – To show the fly the way out of the fly-bottle.[83]

Once the false views of meaning and understanding have been destroyed, there is no further *constructive* contribution which philosophy can make. Philosophy's job is to provide *therapy*, to prevent us from slipping back into the bad old ways of thinking:

> The philosopher's treatment of a question is like the treatment of an illness.[84]

The philosopher should not try to provide a constructive *answer* to the question: rather, he should try to *cure* the temptation to think that there was a genuine question in the first place.[85] We can thus say nothing constructive or explanatory about our practices (or "language-games", as Wittgenstein puts it) of ascribing meaning and understanding:

> Our mistake is to look for an explanation where we ought to look at what happens as a "proto-phenomenon". That is, where we ought to have said: *this language-game is played*.[86]
>
> The question is not one of explaining a language-game by means of our experiences, but of noting a language game.[87]

Wittgenstein is thus a non-reductionist about meaning, but also a *philosophical quietist* who thinks that once we have pointed out the errors in other conceptions of meaning, there is nothing constructive left for us to say:

> If I have exhausted the justifications I have reached bedrock, and my spade is turned. Then I am inclined to say: "This is simply what I do".[88]

The philosopher cannot justify our practices of ascribing meaning and justification: all he can do is dispel misunderstandings of what those practices involve. After that, all he can say is "This game is played" or "This is what we do".

Wright is sceptical about this quietistic aspect of Wittgenstein's philosophy. *Should* we just accept that meaning, intention, and the like exhibit both first-person authority and "disposition-like theoreticity" and content ourselves with refuting mistaken conceptions of what meaning and intending consist in? Wright thinks that we should not give up too easily:

> It is an important methodological precept that we do not despair of giving answers to constitutive questions too soon; if the accomplishments of analysis in philosophy often seem meagre, that may be because it is difficult, not impossible.[89]

> I suspect that [Wittgenstein] did not succeed in representing to himself a sound theoretical basis for declining rather than . . . rising to the challenge posed by his own thought which I have tried to describe. In any case, we now confront a challenge: make out the constitutive answer which Wittgenstein . . . does not deliver, . . . or make out the necessary theoretical basis for the analytical quietism which, "officially", he himself adopted.[90]

As we saw in §6.6, Wright himself does try to deliver an answer to the constitutive question, via his "judgement-dependent" conception of meaning.

McDowell rejects this view of Wittgenstein, which represents him as raising some perfectly good philosophical questions but then declining to attempt to answer them:

> Wright takes Wittgenstein to have uncovered some good philosophical problems about meaning and understanding, [and as appealing to] an adventitiously negative view of philosophy's scope to justify not engaging with those tasks.[91]

Moreover, McDowell thinks that Wittgenstein himself would have been totally out of sympathy with something like Wright's "judgement-dependent" conception of meaning. There are two main reasons for this.

First, on Wright's "judgement-dependent" conception of meaning or intention, there is no inner state of meaning such and such or intending to so and so which our judgements about our meanings or intentions access or track. As Wright puts it:

> There is no essential inner epistemology of rule-following. To express the matter dangerously, we have nothing "in mind".[92]

McDowell thinks that Wittgenstein himself had no inclination to deny that having an intention is a matter of having something in mind:

> *Certainly* all these things happen in you. – And now all I ask is to understand the expression we use. – The picture is there. And I am not disputing its validity in any particular case. – Only I want to understand the application of the picture.[93]

Wittgenstein does not want to deny that having an intention is a matter of having something "in mind": he only wants to destroy *certain misleading conceptions* of *what it is* to have something "in mind".

Second, McDowell thinks that some of the implications of the judgement-dependent conception are actually destructive of our everyday conceptions of meaning and understanding. The judgement-dependent conception of intention implies that:

> There is nothing for an intention, conceived as determining subsequent conformity and non-conformity to it autonomously and independently of its author's judgements on the matter, to be.[94]

Likewise, the judgement-dependent conception of meaning implies that:

> The meaning, say, of an instruction for extending a numerical series [does not determine] what is correct at any point in the series in advance of anyone's working out the series to that point.[95]

McDowell thinks that in each case we have a dramatic departure from our ordinary, common-sense conceptions of meaning and intention. The idea

> that the meaning reaches forward in the series ahead of anyone who actually works the series out, and is so to speak already there waiting for such a person, ready to stand in judgement over her performance, at any point in the series that she reaches . . . is just part of the idea of meaning's normative reach.[96]

The move to a judgement-dependent conception of meaning cannot preserve the idea that the meaning of an expression lays down a normative constraint on future uses of that expression. And the departure from common sense is even clearer when we consider intention:

> Suppose I form the intention to type a period. If that is my intention, it is settled that only my typing a period will count as executing it. If that is indeed the intention which I form, nothing more than the intention itself is needed to determine what counts as conformity to it. Certainly it needs no help from my subsequent judgements. So there is something for my intention to type a period, conceived as determining what counts as conformity to it autonomously and independently of my judgements on the matter, to be: namely, precisely, my intention to type a period. An intention to type a period is exactly something that must be conceived in that way. This is common-sense.[97]

McDowell thinks that it is unlikely that Wittgenstein – who thought that "philosophy leaves everything as it is" – would accept an account of meaning and intention which implies the falsity of these common-sense views.

How, then, does Wittgenstein justify his philosophical quietism while at the same time leaving our everyday conceptions of meaning and intention intact? Intuitively, the fact that someone grasps a certain meaning is a fact about that person's mind. Suppose that one accepted a conception of the mind (call it the "master-thesis") according to which

> it is populated exclusively with items that, considered in themselves, do not sort things outside the mind, including specifically bits of behaviour, into those that are correct or incorrect in the light of those items.[98]

The master-thesis has it that, considered in themselves, mental states and acts just "stand there like a signpost".[99] A signpost, considered in itself, does not sort episodes of behaviour into those that constitute following the signpost correctly or incorrectly. So, if you accepted this conception of the mind, you would have to give some account of what *does* sort episodes of behaviour into those that constitute following the signpost correctly, or as acting in accord with the meaning which is grasped. A tempting answer would be that:

> What does sort behaviour into what counts as following the sign-post and what does not is not an inscribed board fixed to a post, considered in itself, but such an object *under a certain interpretation* – such an object interpreted as a sign-post pointing the way to a certain destination.[100]

But according to Wittgenstein this idea is hopeless. Recall the passage from *Philosophical Investigations* §198 which we quoted in §5.1:

> Any interpretation still hangs in the air along with what it interprets, and cannot give it any support. Interpretations by themselves do not determine meaning.

Suppose that meaning such and such by an expression was a matter of putting a certain interpretation on it. Then what is required for competence is that the *correct* interpretation is put on it.

But, according to the master-thesis, acts of mind – such as putting an interpretation on an expression – just "stand there like a signpost". So, we require that the interpretation of the expression *itself* be interpreted. And now we are off on a regress: the interpretation of the interpretation in its turn just "stands there like a signpost", and so will require interpretation, and so on, *ad infinitum*.

Now according to Kripke, at this point Wittgenstein accepts that there is no such thing as a fact about what anyone means and intends, and attempts to develop a sceptical solution to preserve meaning and intention in the light of this conclusion:

> This was our paradox: no course of action could be determined by a rule, because every course of action can be made out to accord with the rule. The answer was: if everything can be made out to accord with the rule, then it can also be made out to conflict with it. And so there would be neither accord nor conflict.[101]

But as many commentators have pointed out, the paragraph in *Philosophical Investigations* which follows this one shows that contrary to what Kripke claims, Wittgenstein does *not* accept the sceptical paradox that there is no such thing as a fact about meaning:

> It can be seen that there is a misunderstanding here from the mere fact that in the course of our argument we give one interpretation after another; as if each one contented us at least for a moment, until we thought of yet another standing behind it. What this shows is that there is a way of grasping a rule which is *not* an *interpretation*, but which is exhibited in what we call "obeying the rule" and "going against it" in actual cases.[102]

McDowell suggests that what this shows is that Wittgenstein does not accept the master-thesis plus the consequent need for interpretation, and then proceed to draw the conclusion that since interpretations do not determine meaning, there is no such thing as meaning an expression in one way rather than another; rather, *Wittgenstein blocks the route to the sceptical paradox by refusing to accept the master-thesis in the first place.*

Is Wittgenstein's refusal to accept the master-thesis justified? According to McDowell, the master-thesis is an extremely unintuitive piece of philosophical theory, and as such the onus is on its defenders to justify their adherence to it. Consider a mental occurrence, like the having of the thought *that people are talking about me in the next room*. Only a state of affairs in which people are talking about me in the next room will be in accord with this thought; if another state of affairs obtains, such as a state of affairs in which the next room is empty, this will not be in accord with that thought. The master-thesis then implies

> that whatever I have in my mind on this occasion, it cannot be something to whose very identity that normative link to the objective world is essential . . . that what a person has in mind, strictly speaking, is never, say, *that people are talking about her in the next room* but at most something that *can* be interpreted as having that content, although it need not.[103]

And McDowell finds this at worst an extraordinary idea, and at best highly counterintuitive and unmotivated:

> Once we realise that, the master-thesis should stand revealed as quite counterintuitive, not something on which a supposed need for constructive philosophy could be convincingly based.[104]

Wright, because he (perhaps unwittingly) accepts the master-thesis ends up denying that meaning and intending can be a matter of having something "in mind", and this leads him to a piece of "constructive philosophy", the judgement-dependent conception of meaning and intention. But once the master-thesis is revealed as an unintuitive and unmotivated philosophical invention, the need for such a piece of constructive philosophy simply falls away. Since there is no need for a judgement-dependent conception of meaning and intention, we can simply hold on to the common-sense idea that meaning "reaches ahead" and determines in advance, independently of anything we go on to say and do, what patterns of behaviour accord with it and which do not. Everything is left as it was:

> The question "How is it possible for meaning to reach ahead of any actual performance?" is just a specific form of the question "How is it possible for the concept of accord to be in place in the way that the idea of meaning requires it to be?" The Wittgensteinian response is not that these are good questions, calling for constructive philosophy to answer them. The Wittgensteinian response is to draw attention to the way of thinking that makes it look as if there are problems here.[105]

Questions like "How is meaning possible?" express a sense of spookiness, and Wittgenstein's point is that we should not indulge the sense of spookiness, but rather exorcise it. The question looks like an urgent one from the standpoint of a world-view that is inhospitable to meaning: a standpoint from which it looks like a task for philosophy to shoehorn the world into something as close as we can get to our previous conception of meaning. But philosophy's task is rather to dislodge the assumptions that make it look difficult to find a place for meaning in the world.[106]

Does Wittgenstein really succeed in exorcising the spookiness that a range of irreducible, *sui generis* facts about meaning and intention can make us feel? Or is Wright correct that the questions raised are good ones requiring constructive (as opposed to merely therapeutic) philosophy to answer them? And what of the call for an explanation of the first-person authority of ascriptions of meaning and intention? Can Wittgenstein exorcise the temptation we feel to make such a call? These are the questions that must be answered if we are to have a true estimate of Wittgenstein's contribution to the study of meaning.[107]

6.9 Ginsborg's "partial reductionism"

Recall from §6.2 that the sophisticated form of (reductive) dispositionalism holds something along the following lines:

> Jones means *addition* by "+" if and only if Jones is disposed to respond to arithmetical queries of the form "x + y =?" in ideal conditions with the *sum* of the numbers denoted by "x" and "y",

where the relevant ideal conditions are to be spelled out in purely non-semantic and non-intentional terms. In a recent set of papers, Ginsborg has argued that we can reply to KW's sceptic by replacing the above form of dispositionalism with the following:

> Jones means *addition* by "+" if and only if Jones is disposed to respond to arithmetical queries of the form "x + y =?" in ideal conditions with the *sum* of the numbers denoted by "x" and "y" and to regard those responses as appropriate,

where *taking-a-response-to-be-appropriate* is a primitive normative attitude, in the sense that we can think of Jones as having that attitude towards the relevant responses independently of thinking of him as attaching a particular (or indeed any) meaning to the "+" sign. Moreover, not only is Jones disposed to regard the relevant responses as appropriate, he is entitled to do so. In particular, not only is Jones disposed to answer "125" to "68 + 57 =?" and to regard that answer as appropriate, his doing so is legitimate.[108]

The position advocated by Ginsborg is in a sense an extension of reductive dispositionalism: it amounts to a form of what Ginsborg calls "partial reductionism". On the partially reductionist view semantic facts – like the fact that "+" means addition – are reducible to facts about our primitive normative attitudes. The reduction is "partial" in that although semantic facts are reduced to normative facts, there is no reduction of semantic facts to physical facts (or facts that can be characterised in non-semantic and non-intentional terms). Moreover, although the account yields an account of what our grasp of "+" as meaning *addition* consists in, it does not apply to our grasp of "appropriate" itself.[109]

Ginsborg holds that we need to extend reductive dispositionalism in this way, as the unextended form of reductive dispositionalism is inadequate: an automaton could be disposed to respond to arithmetical queries with the sum of the relevant numbers, but we would not want to say that the automaton genuinely understands the "+" sign to mean addition. The crucial ingredient which the automaton lacks is consciousness of the appropriateness of his answers: a disposition to regard his responses as appropriate.[110]

Because Ginsborg's position involves an extension of reductive dispositionalism in this way, she needs to defend reductive

dispositionalism against Kripke's "finitude" objection. I'll now suggest that Ginsborg's defence of reductive dispositionalism against KW's "finitude" objection is inadequate and that the attitude of taking-a-response-to-be-appropriate that figures in her account is not plausibly regarded as primitive in the relevant sense. Moreover, I'll argue that there is no clear motivation for adopting Ginsborg's partially reductive position in preference to a form of non-reductionism.

6.9.1 Ginsborg on the finitude problem

We saw in §6.2 that sophisticated dispositionalism attempts to respond to the "finitude" objection described in §5.1 by introducing the notion of ideal conditions. Given that due to the finite nature of his lifespan and cognitive capacities, and so on, Jones is not disposed to give the sum for numbers which are simply too large for him to handle, his actual dispositions are consistent with "+" denoting a function (such as skaddition) which yields the answer 5 for numbers that are too large for him to compute. The thought behind sophisticated dispositionalism was that we could get round this by taking what Jones means to be determined, not by his actual dispositions, but by his dispositions to respond in ideal conditions: since living long enough and having sufficient brain power are presumably components of the relevant ideal conditions, I am disposed to respond in ideal conditions with the sum rather than "5" even for numbers that are in fact too large for me to handle. In §6.2 we outlined an argument developed by Boghossian to the effect that it will not be possible for these ideal conditions to be outlined in non-semantic and non-intentional terms, and eventually concluded that no form of reductive dispositionalism yields a plausible response to KW's sceptical challenge.

Ginsborg argues that KW's "finitude" objection fails to damage reductive dispositionalism. Specifically, she argues that the "finitude" objection does not amount to a special problem faced by dispositional theories of meaning:

> The difficulties Kripke raises for the dispositional analysis of meaning are not specific to meaning, but simply difficulties for the idea of a disposition as such . . . [W]e can ascribe the

> disposition of brittleness to a glass even if we allow that the disposition could never be actualised on Alpha Centauri because the glass would decay before it got there. By the same token, it would seem that we should be able to ascribe a disposition even while allowing that this disposition could not manifest itself for numbers beyond a certain size.[111]

According to Ginsborg, we can defuse the "finitude" objection to reductive dispositionalism without relying on resources over and above those we would invoke in the case of an uncontroversial dispositional property like fragility (the existence of which is officially not under attack from KW's sceptic). We can set out Ginsborg's argument here in two steps:

> [a] When we call the glass fragile, we are typically content to understand its fragility as a matter of its having a tendency to break when struck, and it does not trouble us that we cannot give a precise and exhaustive characterization of the "ideal conditions" under which that disposition is actualized except by saying that they are the circumstances under which the glass in fact breaks when struck.
>
> [b] [T]he inability to characterize ideal conditions for the actualization of a disposition is no more problematic for supposedly meaning-constituting dispositions than it is for the kinds of dispositions which we invoke in everyday talk about objects. . . . [W]e can give a similarly circular specification of the disposition in virtue of which someone means addition by "plus", namely one on which the ideal conditions are those in which the person gives the sum.[112]

In response to Ginsborg, we can argue that neither [a] nor [b] is plausible. In response to [a], although it may be true that we don't have a "precise and exhaustive" characterisation of the ideal conditions for the manifestation of the dispositional property of fragility, it does appear to be relatively straightforward to give at least a rough but substantive account of those conditions. Consider my favourite whisky glass. In this case we can characterise the

ideal conditions as conditions in which the glass is dropped from a height of one metre in physical conditions of the sort that typically obtain in central Glasgow. This may be vague, but not sufficiently so to undermine the idea that there is a determinate fact of the matter as to the fragility of the glass. The importance of this will be apparent following our response to [b].

In response to [b], although Ginsborg speaks of "circularity" in [a] and [b], what she is describing is better thought of as *triviality*. Consider a biconditional which attempts to spell out what it is for an object to possess the property of F-ness:

> (F) x is F if and only if (if conditions C obtain then an e-type event involving x occurs).

A specification of the conditions C is trivial if they are specified such that their obtaining trivially ensures the occurrence of an e-type event (e.g. if the conditions "C" are specified as those conditions whose obtaining has "whatever it takes" to ensure the occurrence of an e-type event, or as in [a] the conditions for the manifestation of the fragility of a glass are characterised as "circumstances under which the glass in fact breaks when struck"). Trivially, if you strike a glass in circumstances in which the glass in fact breaks when struck, the glass will break. In general, triviality in the characterisation of "C" undermines (F) since the ensuing trivial truth of the right-hand side implies that anything whatsoever is F. Triviality, however, should be sharply distinguished from *circularity*: (F) would be circular if "F" were to be used in the specification of the conditions C: if, in other words, we adverted to something or other being F or not in the course of specifying the those conditions. Note that the main worry outlined in §6.2 about the ideal conditions that figure in a sophisticated dispositionalist account of meaning concerned circularity rather that triviality: the worry was that the specification of the ideal conditions would need to advert to absence of content bearing states like beliefs. And recall that although ultra-sophisticated dispositionalism

initially appeared to offer a way round the circularity problem, ultimately it failed due to the open-ended and potentially infinite nature of the set of relevant belief clusters.

Overall, then, Ginsborg's response to KW's finitude problem fails. In the case of fragility we can give a non-trivial and – crucially – non-circular specification of the conditions which are ideal for its manifestation, whereas in the case of meaning circularity in the specification of the ideal conditions constitutes an insurmountable problem.

6.9.2 How primitive is the primitive normative attitude?

As we noted above, Ginsborg holds not only that we take "125" to be the appropriate answer to "68 + 57 =?" but also that this attitude is *legitimate*. We'll now argue that Ginsborg's claim that the attitude is primitive – in the sense that we can think of someone taking it towards uses of an expression independently of thinking of them as attaching a meaning to that expression – is undermined by the role played in her account by reductively characterised ideal conditions.

Consider the expression "horse". According to Ginsborg, Jones means *horse* by "horse", since Jones is disposed to apply "horse" to horses in ideal conditions and to take those applications to be appropriate. Let's dub the ideal conditions "C". We know that often the conditions are less than ideal. For an example of less than ideal conditions, consider conditions C* in which it is a dark night and I am mildly inebriated. Suppose that in C* I'm disposed to apply "horse" to certain horsey-looking cows and to regard those applications as appropriate. Is it legitimate for me to take the application of "horse" to horsey-looking cows in these conditions to be appropriate? If so, "horse" will end up with a disjunctive extension which includes both horses and certain horsey-looking cows. So it can't be legitimate for me to regard my applications of "horse" in conditions C* as appropriate. But why not? The only answer would appear to be: because conditions C* are less than ideal for the application of "horse". And the reason C* are less than ideal

for the application of "horse" is that regarding them as ideal would generate an extension for "horse" that is out of line with the extension determined by its meaning: namely, the set consisting of all and only horses. So, the meaning of "horse" comes into play in the determination of which set of conditions – C or C* – is ideal for the application of "horse", hence the meaning of "horse" comes into play in determing which of our normative attitudes are legitimate. If this is the case, then, our normative attitude of taking a use of an expression to be appropriate is not independent of the meaning we attach to the expression.

Hence, Ginsborg's own account – in which reductively characterised ideal conditions play a substantive role in determining the extensions of our expressions – undermines the idea that we have a primitive entitlement to the normative attitudes we take towards certain uses of those expressions.

6.9.3 Partial reductionism versus non-reductionism

We saw in §6.5 and §6.6 that one of the main problems for non-reductionism was the problem of first-person knowledge of meaning. How is it possible for us to square the first-person epistemology of linguistic meaning with its "disposition-like theoreticity"? We saw that Wright tries to answer this question head on by constructing judgement-dependent accounts of intention and meaning, while McDowell suggests that the question is in fact based on confusion. Either way, a substantive philosophical task must be carried out by the non-reductionist. Now, given that Ginsborg's view identifies understanding with possession of a certain sort of disposition, this highlights the fact that Ginsborg too owes us a story regarding how the first-person epistemology of meaning is to be reconcliled with its disposition-like theoreticity: she can argue (with Wright) that the question can be answered via a piece of "constructive" philosophy, or (with McDowell) that no genuine question has been asked. However, and crucially, in addition to having to take on the philosophical obligations faced by the non-reductionist, by her own lights Ginsborg's position also has to take on some of the philosophical obligations of the

reductive dispositionalist – specifically, the obligation of defusing KW's finitude objection. This means that there appears to be no sound motivation for adopting the partially reductionist view in preference to one of the forms of non-reductionism. Indeed, given that the partially reductionist view faces *both* the central problem faced by reductive dispositionalism and the central problem faced by non-reductionism, far from steering a middle path between the two it ends up with a position that is worse than either. Indeed, Ginsborg's partial reductionism starts to look like the worst of both worlds.[113]

Chapter summary

In this chapter, we have been concerned with straight solutions to KW's sceptical paradox. We began by suggesting that the stakes are high, insofar as the sceptical paradox threatens the notion of mental content as well as the notion of linguistic meaning, thereby threatening our picture of ourselves as agents capable of intentional actions. We then looked at a number of straight solutions, beginning with sophisticated dispositionalism and ending with Ginsborg's recent account in terms of primitive normative attitudes.

Study questions

1. Give examples of five types of propositional attitudes and explain how each of them can figure in intentional action in virtue of their contents.
2. Could one be a sceptic about linguistic meaning without being a sceptic about mental content? Could one be a sceptic about mental content without being a sceptic about linguistic meaning?
3. Give at least three examples of your own to illustrate Boghossian's claim that "just about any stimulus can cause just about any belief, given a suitably mediating set of background assumptions". What sort of problem does this pose for sophisticated dispositionalism?

4. What does it mean to say that one causal relation is asymmetrically dependent on another? How might one use the idea in constructing a reductionist account of meaning and content?
5. Evaluate Wright's objection to McGinn's attempt to use the notions of capacity and ability to respond to KW's sceptical challenge.
6. Why is the combination of the intuitive first-person epistemology of meaning with the "disposition-like theoreticity" of meaning potentially problematic? Would a judgement-dependent account of meaning deal with the problem?
7. Is Pettit's "ethocentric" account of rule-following and meaning threatened by Boghossian's argument from belief holism (see question 3 above)?
8. What is involved in the claim that McDowell calls the "master-thesis"? If there are no good grounds for accepting it, does this dissolve KW's sceptical paradox?
9. Could there be a reason for adopting Ginsborg's "partial reductionism" in preference to a form of non-reductionism about meaning?

Further reading

A number of important papers on the topic of this chapter are collected in Miller and Wright (eds), *Rule-Following and Meaning*. Boghossian's literature-survey "The Rule-Following Considerations" is extremely useful in relating KW's arguments to themes in contemporary philosophy of language and mind. Three papers which attempt to defend versions of sophisticated dispositionalism are Forbes, "Scepticism and Semantic Knowledge"; Ginet, "The Dispositionalist Solution to Wittgenstein's Problem about Understanding a Rule"; and Heil and Martin, "Rules and Powers". See also Tennant, *The Taming of the True*, Chapter 4. For criticism of dispositionalism, see Boghossian, "The Rule-Following Considerations", §5, and "Naturalizing Content". For details of Lewis's account of reductive analysis, see his "How to Define Theoretical Terms" and "Psychophysical and Theoretical Identifications". For more on Boghossian's arguments against dispositionalism see McManus, "Boghossian, Miller and Lewis

on Dispositional Theories of Meaning"; and Miller, "Does 'Belief Holism' Show that Reductive Dispositionalism about Content Could Not Be True?". Horwich defends a non-standard version of dispositionalism in his *Meaning*, Chapters 4 and 10, and also in "Meaning, Use and Truth"; see also *Wittgenstein's Metaphilosophy*, Chapter 5. For a critique of Horwich, see Miller, "Horwich, Meaning and Kripke's Wittgenstein".

For Fodor's views, see "Meaning and the World Order", Chapter 4 of his *Psychosemantics*; and "A Theory of Content (I and II)" in his *A Theory of Content and Other Essays*. For discussion of Fodor, see Crane, *The Mechanical Mind*, Chapters 4 and 5; Zalabardo, "A Problem for Information Theoretic Semantics"; and Boghossian, "Naturalizing Content".

McGinn's views on Wittgenstein are contained in his book *Wittgenstein on Meaning*. This contains some extremely useful material but should be read in conjunction with Wright, "Critical Notice" and Boghossian, "Review".

Wright's "judgement-dependent" account of meaning and intention is outlined in his "On Making Up One's Mind: Wittgenstein on Intention" and "Wittgenstein's Rule-Following Considerations and the Central Project of Theoretical Linguistics". Other useful exposition is provided in his "Moral Values, Projection, and Secondary Qualities" and *Truth and Objectivity*, appendix to Chapter 3. For criticism of Wright, see Holton, "Intention-Detecting"; Miller and Divers, "Best Opinion, Intention-Detecting, and Analytic Functionalism"; Miller, "An Objection to Wright's Treatment of Intention", "Another Objection to Wright's Treatment of Intention", and "Primary Qualities, Secondary Qualities and the Truth about Intention"; Johnston, "Objectivity Disfigured", appendix on "Two Distinctions"; Sullivan, "Problems for a Construction of Meaning and Intention"; Edwards, "Secondary Qualities and the A Priori" and "Best Opinion and Intentional States"; and Blackburn, "Circles, Finks, Smells, and Biconditionals".

For Pettit's attempt to solve the rule-following problem using an analogue of judgement-dependence, see his "The Reality of Rule-Following" and *The Common Mind*. Other relevant papers by Pettit include "Realism and Response-Dependence" and

"A Theory of Normal and Ideal Conditions". For a useful exchange on Pettit's attempted solution, see Summerfield, "On Taking the Rabbit of Rule-Following out of the Hat of Representation" and Pettit, "Affirming the Reality of Rule-Following".

Wittgenstein's key texts on rule-following are *Philosophical Investigations*, esp. §§1382–1342, and *Remarks on the Foundations of Mathematics*, esp. §6. A good book that traces the connections between Wittgenstein's views and contemporary philosophy of mind and language is Thornton, *Wittgenstein on Language and Thought*. McDowell's interpretation of Wittgenstein is most accessible in "Meaning and Intention in Wittgenstein's Later Philosophy". Other important papers are McDowell, "Intentionality and Interiority in Wittgenstein", "Wittgenstein on Following a Rule", "Are Meaning, Understanding, etc. Definite States?", and "How Not to Read *Philosophical Investigations*: Brandom's Wittgenstein". For a critique of McDowell's interpretation of Wittgenstein, see Wilson, "Semantic Realism and Kripke's Wittgenstein". The debate between Wright and McDowell on the import of Wittgenstein's rule-following considerations is pursued in a number of papers, including: Wright, "Self-Knowledge: The Wittgensteinian Legacy"; and McDowell, "Reply to Crispin Wright". See also McDowell, *Mind and World*; Wright's critical study, "Human Nature?"; Miller, "Rule-Following, Response-Dependence, and McDowell's Debate with Anti-Realism"; Ebbs, *Rule-Following and Realism*, Chapter 5; De Gaynesford, *McDowell*, Chapter 4; and Thornton, *John McDowell*, Chapter 1.

Ginsborg's papers on this topic include "Inside and Outside Language: Stroud's Non-Reductionism about Meaning", "Primitive Normativity and Skepticism about Rules", and "Meaning, Understanding, and Normativity". For a detailed critique see Miller, "Rule-Following, Meaning and Primitive Normativity".

For more general discussions of Wittgenstein, see Malcolm, *Nothing is Hidden*; and Hanfling, *Wittgenstein's Later Philosophy*. For a very readable and philosophically stimulating biography, see Monk, *Ludwig Wittgenstein: The Duty of Genius*.

For additional suggestions for further reading, see Kusch's excellent annotated bibliography in *Oxford Bibliographies Online*.

Notes

1 The term "propositional attitude" was coined by Bertrand Russell. It is relatively easy to see the reason for this term. A belief consists in taking a certain sort of attitude (the "belief-attitude") to a proposition or thought; so having the belief that Scotland will win the next World Cup consists in taking the "belief-attitude" to the proposition *that Scotland will win the next World Cup*. Likewise, having a desire to win the National Lottery consists in taking the "desire-attitude" to the proposition that you win the National Lottery, and so on. The "aboutness" of propositional attitudes is sometimes described as "intentionality". See Crane, *Mechanical Mind*, pp.31–40 for a discussion.

2 See Dummett, *Nature and Future of Philosophy*, Chapter 13.

3 Note that this only follows if the explanation in question is a *genuine* one. Obviously, if the notion of responsibility is only *apparently* explained by the notion of free will, the conclusion that the notion of responsibility is incoherent will not force the conclusion that the notion of free will is incoherent. But this is not relevant to the question at hand: we are asking what follows from KW's argument against linguistic meaning given the assumption that the Gricean explanation of the relationship between thought and language is a *genuine* one.

4 Boghossian, "Rule-Following Considerations", p.144. McGinn disagrees with this claim of Boghossian's, and argues that KW's sceptical paradox cannot even be formulated at the level of mental content. It turns out that this is based on McGinn's mistaken conception of the normativity of meaning (see §6.5 and Boghossian, "Rule-Following Considerations", pp.146–149).

5 Boghossian, "Status of Content", p.171.

6 Given the conclusion of this section – that KW's arguments are of a completely general nature, applying both to linguistic items and to mental states – I won't be fastidious about distinguishing between mental content and linguistic meaning in the remainder of the chapter.

7 Boghossian, "Rule-Following Considerations", p.165.

8 Ibid., p.177.

9 Ibid.

10 Ibid.

11 Ibid.

12 See Lewis, "Radical Interpretation", for this sort of story.

13 See the Further reading section at the end of the chapter for references to Lewis. An extremely clear exposition of Lewis's method can be found in Chapter 2 of Smith, *The Moral Problem*. I follow Smith's exposition in the text.
14 So, if the platitudes are "the property of redness causes us, under certain conditions, to have experiences of the property of redness" and "the property of redness is more similar to the property of orangeness than the property of blueness" and . . . , T(x, y, z) will be "(x causes us, under certain conditions, to have experiences of x) & (x is more similar to y than z) & . . .".
15 The right-hand side of this definition is known as the Ramsey Sentence for "red", after the British philosopher Frank Ramsey (1903–1930), who originated the Lewis style of reductive analysis.
16 Boghossian, "Naturalizing Content", pp.90–91.
17 For more on Boghossian's arguments, see the papers by McManus and Miller mentioned in the Further reading section.
18 Fodor, *Psychosemantics*, p.97.
19 Note also that Fodor thinks that the problem should be discussed, not with respect to public language expressions such as "horse" or "cow", but with respect to the expressions of a mental language, a language of thought, which he calls *mentalese*. In what follows, we will ignore this complication; for an account of Fodor's hypothesis that we think in mentalese, see Fodor, *Language of Thought*, or for a quicker exposition, Crane, *Mechanical Mind*, Chapter 4.
20 Fodor argues convincingly against a range of suggestions as to how the optimal conditions should be cashed out: see his "A Theory of Content (part I)" in *A Theory of Content and Other Essays*.
21 In fact, Fodor only claims to offer sufficient (rather than necessary *and* sufficient) conditions for meaning such-and-such in his asymmetric-dependency account, but we can gloss over this for present purposes.
22 Crane, *Mechanical Mind*, pp.180–181. Crane speaks of "mental representation" here, since he's discussing Fodor's views with respect to the expressions of mentalese (see note 19). For our purposes, we can just read the talk of "mental representation" here as talk of "meaning".
23 Boghossian raises the worry that Fodor's asymmetric-dependency account is after all just another version of the "optimal conditions" solution to KW's sceptical problem. If this is the case, the worries that we developed in §§6.2–6.3 will apply even to

Fodor's own attempted solution. See Boghossian, "Naturalizing Content", pp.71–73.

24 For more on this aspect of Wittgenstein's views, see McGinn, *Wittgenstein on Meaning*, Chapter 1, esp. pp.28–42.

25 Wittgenstein, *Philosophical Investigations*, §150.

26 Ibid., §199.

27 Wittgenstein, *Remarks on the Foundations of Mathematics*, p.344.

28 McGinn, *Wittgenstein on Meaning*, p.173.

29 Ibid., p.74.

30 Ibid.

31 Kripke, *Wittgenstein on Rules and Private Language*, p.14.

32 Boghossian, "Rule-Following Considerations", p.147.

33 Ibid. p.148.

34 Ibid., p.149.

35 Ibid., p.150 n.14.

36 Wright, "Critical Notice of McGinn", p.111.

37 Ibid.

38 McGinn, *Wittgenstein on Meaning*, pp.169–170.

39 McGinn actually considers this rejoinder very briefly (ibid., p.174), and dismisses it as depending upon an "unargued reductionism" about capacities or abilities. But as we'll see in §6.6, though the dismissal of unargued reductionism might be in order, it cannot by itself do the desired work against KW's sceptic.

40 Wright, "Critical Notice of McGinn", p.111. Of course, this is not to say that the Wittgensteinian point – that we should view understanding as akin to the possession of an ability is *false*: it is just to say that even if true, it cannot *by itself* constitute an adequate response to KW's sceptical paradox. It seems to me that the objections against the capacity/ability account also apply to the "activity based model of understanding" suggested by McCulloch in *The Mind and Its World*, Chapter 4.

41 Wright, "Kripke's Account", p.110.

42 Kripke, *Wittgenstein on Rules and Private Language*, p.51.

43 Of course, you may have to justify *why* you have it, but this is different from having to justify *that* you have it.

44 McGinn, *Wittgenstein on Meaning*, pp.160–161.

45 Wright, "Critical Notice of McGinn", p.113. "Philosophical stone kicking", after Dr Johnson's infamous attempt to refute Berkeley's idealism by kicking a stone and saying "I refute him thus". Wright's claim is that just as Johnson completely misses Berkeley's point, McGinn likewise misses Kripke's.

46 Wright argues that this problem is brought sharply into focus by Wittgenstein himself in some sections of *Philosophical Investigations*. See Wright, "On Making Up One's Mind", for the relevant sections.

47 The same goes for McGinn's non-reductionism about abilities and capacities (see §6.5).

48 The exposition I give of Wright's account of judgement-dependence is necessarily concise: the reader should study carefully the papers by Wright mentioned in Further reading, for detailed expositions and attempts to apply the conception in particular cases (such as discourse about shape, colour, and morals).

49 Think about it: it would be no use taking it into a lab and illuminating it, or whatever, as this is more likely to *distort* the appearance of the colour it actually possesses. Conditions in which its physical microstructure is examined won't be better either. Suppose we find an object which looks red to suitable subjects in the ideal conditions specified, but with a microstructure different from that typically found in red things. What would we say: that the object isn't *really* red, or that we have discovered that there can be more than one physical basis for redness? It seems to me that we would say the latter (constrast this with a natural kind, such as gold). Note also that I'm simplifying somewhat: we need to say a bit more about what constitutes the "suitability" of a subject. See Wright, "Wittgenstein's Rule-Following Considerations s", §3.

50 In a sense, Wright is attempting to construct a modern version of the distinction between primary and secondary qualities. Secondary qualities, like *redness*, will turn out to be judgement-dependent, while *primary* qualities, such as squareness, will turn out to be judgement-independent. So, in a way, Wright is attempting to respond to KW's sceptical paradox by viewing *means addition by "+"* as a sort of secondary quality.

51 The idea here is clear: redness is judgement-dependent, since there is a strong connection (*a priori*, non-trivial, etc.) between best opinion about redness and the facts about redness. Note that a specification of the C-conditions is trivial if it formulates the conditions in such a way that they amount to the claim that the subject in question has "whatever it takes" to form the right opinion concerning e.g. redness. Plainly, the fact there is a *trivially* specified set of conditions under which best opinions covary with the facts tells us nothing interesting about redness: the covariance between best opinion and fact would be *a priori* for any

discourse, given such a trivial specification of the C-conditions. Also, the question as to whether the C-conditions are satisfied has to be logically independent of facts about redness, for the following reason. We want to claim that the facts about redness are *determined* by our best opinions about redness; but we'll lose this idea that the facts are dependent upon best opinion if the question as to whether a given opinion is best presupposes some fact or other about redness. This would make best opinion dependent upon the facts about redness, rather than vice versa. As mentioned in note 50, Wright argues that the conditions are satisfied in the case of colour discourse, but that they cannot be collectively satisfied in the case of discourse about shape.

52 Although Wright thinks that there are some worries about the non-triviality of the C-conditions – due to the presence of the clause ruling out self-deception – he also thinks that these can be satisfactorily dealt with. See Wright, "Wittgenstein's Rule-Following Considerations", §4. For an argument to the contrary, see Miller, "Another Objection to Wright's Treatment" and "Primary Qualities".

53 Edwards, "Best Opinion and Intentional States", p.24.

54 Wright, "Wittgenstein's Rule-Following Considerations", p.206.

55 For details of the other objections, see the Further reading section of this chapter.

56 Boghossian, "Rule-Following Considerations", p.185.

57 See Wright, *Truth and Objectivity*, pp.138–139.

58 Kripke, *Wittgenstein on Rules and Private Language*, p.22.

59 Here Pettit has in mind Kripke's brief discussion of Fregean platonism at the very end of Chapter 2 of *Wittgenstein on Rules and Private Language*.

60 Pettit, "Reality of Rule-Following", p.34.

61 Ibid., p.36.

62 Ibid., p.42.

63 Ibid., p.41.

64 Note that Pettit refers to the ethocentric genealogy as "conceptual explanation"; Pettit, "Affirming the Reality", p.435.

65 Note that the damaging kind of circularity is distinct from the circularity in the biconditionals which, according to Pettit, provide a "summary" of the ethocentric account and which may well be harmless. See ibid.

66 Pettit, "A Theory of Normal and Ideal Conditions", p.144.

67 It may be that Pettit faces a dilemma here. If a naturalistic account of sameness of semantic value cannot be given, then

the ethocentric account cannot get going. On the other hand, if a naturalistic account of sameness of semantic value can be given, this may yield resources for an account of how an expression can have a semantic value which will render the ethocentric account redundant.

68 Pettit, "Reality of Rule-Following", pp.36–37.

69 I've made the point here against the presentation in Pettit, "Reality of Rule-Following", but the same point can be made with respect to the presentation in *The Common Mind*. Pettit writes (p.89):

> My theory agrees with most other accounts in positing an extrapolative disposition but it gives the disposition a distinctive role. First, it invokes the disposition to explain how some instances can present a rule to an agent: to explain how the instances can exemplify the rule. Second, it invokes the disposition to explain how the agent can be directly prompted in new cases. Third, it connects the presentation and the prompting aspects of the disposition in order to make sense of what it is for a rule to speak directly to a subject, telling the subject how to go on. Because the inclination serves the presentational role, its significance in the role of prompting the agent's responses in new cases is transformed from what it might otherwise be. The prompting does not involve an unintelligible leaning: an exogenously inspired urge to go this way rather than that. It appears as a leaning that the rule itself elicits in the agent, a response that it forces on any subject that understands it.

But, again, the question is *whether* an extrapolative disposition *can* indeed play the presentational role. In assuming that it is unproblematic to think of an inclination – a disposition – as having a presentational role that can straightforwardly connect with its prompting role in this way, Pettit effectively begs the question about the issue concerning normative guidance.

70 For a fuller exposition, see Miller, *Contemporary Metaethics*, Chapter 4.

71 See ibid., Chapter 2 for more on the open question argument.

72 See ibid., Chapter 6.

73 The Frege–Geach problem is the problem of showing how – independently of the assumption that they express beliefs – moral statements can figure in "unasserted" contexts such as the antecedents of conditionals. See §9.2 for an outline. The view of

the general nature of the quasi-realist project relied on in the current section owes much to McDowell, "Projectivism and Truth in Ethics".

74 Blackburn, *Ruling Passions*, p.49.
75 Pettit, *Common Mind*, p.341.
76 Pettit himself briefly notes a parallel with Blackburn in Pettit, "Affirming the Reality", p.435 n.6.
77 The relevant paper is Blackburn, "Individual Strikes Back" in which I take Blackburn to be gesturing in the direction of a quasi-realist account of judgements about rule-following.
78 Ibid., p.37.
79 Wright, "Critical Notice of McGinn", p.127.
80 Wittgenstein, *Philosophical Investigations*, §124.
81 Ibid., §126.
82 Ibid., §119.
83 Ibid., §309.
84 Ibid., §254.
85 Note that Wittgenstein's conception of the role of philosophy is even more radical than that of the logical positivists (see §3.4). According to the logical positivists, the questions of metaphysics are *pseudo-questions*; but even after these have been dispensed with, there is scope for constructive philosophical work in the form of *conceptual analysis*. According to the later Wittgenstein, philosophy doesn't even get as far as conceptual analysis: conceptual analysis only seems necessary because of our acceptance of a range of problematic assumptions, and once these assumptions have been dislodged, the need even for conceptual analysis falls away.
86 Ibid., §654.
87 Ibid., §655.
88 Ibid., §217,
89 Wright, "Wittgenstein's Rule-Following Considerations", p.191. See also Wright, "Study Note on Wittgenstein".
90 Wright, "Critical Notice of McGinn", p.128.
91 McDowell, "Meaning and Intention in Wittgenstein", p.51.
92 Wright, "Critical Notice of McGinn", p.127.
93 Wittgenstein, *Philosophical Investigations*, §423.
94 Wright, quoted by McDowell in "Intentionality and Interiority", p.163.
95 McDowell, "Meaning and Intention in Wittgenstein", p.48.
96 Ibid.

97 McDowell, "Intentionality and Interiority", pp.163–164. For a reply, see Wright, *Rails to Infinity*, p.339.
98 McDowell, "Meaning and Intention in Wittgenstein", p.41.
99 See Wittgenstein, *Philosophical Investigations*, §85.
100 McDowell, "Meaning and Intention in Wittgenstein", p.41,
101 Wittgenstein, *Philosophical Investigations*, §201.
102 Ibid.
103 McDowell, "Meaning and Intention in Wittgenstein", p.46.
104 Ibid.
105 Ibid., p.49.
106 McDowell, *Mind and World*, postscript to Lecture 5.
107 Needless to say, we cannot even begin to answer these extremely difficult questions here. The debate about Wittgenstein's contribution to the philosophy of meaning is one of the most hotly contested taking place in contemporary philosophy of language and mind, and we have only looked at some aspects of it in this section. See Further reading for details.
108 So whereas Kripke demands that our entitlement to regard "125" as the appropriate answer to the query flows from a story about the fact that makes it the case that "+" means addition, for Ginsborg the fact that makes it the case that "+" means addition flows from facts about normative attitudes to which we are primitively entitled.
109 You might think that this latter limitation is damaging, insofar as "appropriate" itself cannot be defended from KW's attack. What makes it the case that "appropriate" means *appropriate*? In response, Ginsborg suggests that our grasp of the concept *appropriate* "does not seem to depend on a capacity to 'go on' from the use of the corresponding word in particular cases"; Ginsborg,"Primitive Normativity and Skepticism about Rules", pp.251–252.
110 See Ginsborg, "Inside and Outside Language", p.168.
111 Ibid., p.158 (she cites Blackburn, "The Individual Strikes Back").
112 [a] and [b] are taken from Ginsborg, "Inside and Outside Language", pp.159–160.
113 The three objections made against Ginsborg in §6.9 draw on lines of argument developed in Miller, "Rule-Following, Meaning and Primitive Normativity".

Chapter 7

Sense, intention, and speech-acts

Grice's programme

In the previous three chapters we have considered various versions of scepticism about meaning, and we have suggested a number of lines of response to such scepticism. Our discussion of attempts to save sense in the light of semantic scepticism has been far from exhaustive. Rather than discuss further responses to scepticism about sense, we return, in the final three chapters, to questions about the nature of sense and the relationship of issues in the theory of meaning to metaphysical issues in philosophy in general.

7.1 Homeric struggles: Two approaches to sense

In his inaugural lecture, delivered in Oxford in 1969, P. F. Strawson began by asking a number of questions:

> What is it for anything to have a meaning at all, in the way, or in the sense, in which words or sentences or signals have meaning? What is it for a particular sentence to have the meaning or meanings it does have?[1]

These questions have been central to the present book. Recall that in §2.7 we distinguished, very roughly, between *sentence-meaning* and *speaker's-meaning*. Sentence-meaning concerns the strict and literal meaning of a given sentence-type. So the following would be a statement of the sentence-meaning of "Jones makes an excellent cup of tea":

(1) "Jones makes an excellent cup of tea" means that Jones makes an excellent cup of tea.

Recall that on the Fregean account of meaning developed in Chapter 2 this could be cashed out as:

(2) "Jones makes an excellent cup of tea" is true if and only if Jones makes an excellent cup of tea.

Now we said that speaker's-meaning is, roughly, a matter of the information someone uttering a particular token of an expression intends to convey by means of the utterance of that expression. In the example in §2.7, the speaker's-meaning of my utterance of a token of "Jones makes an excellent cup of tea" is that Jones is an uninteresting philosopher.

This raises the question of the relationship between sentence-meaning and speaker's-meaning. Do we explain sentence-meaning in terms of speaker's-meaning, or vice versa? In the Fregean account outlined in Chapter 2, questions of sentence-meaning are dealt with by semantics, and questions about speaker's-meaning are dealt with by pragmatics. The central question which pragmatics must answer, on the Fregean picture, is: *given that a type of sentence has a particular sentence-meaning*, what determines the speaker's-meaning of utterances of tokens of that type? So according to the Fregean picture, the explanation of sentence-meaning – in terms of truth-conditions – comes first, with the explanation of speaker's-meaning coming afterwards. In his inaugural lecture, Strawson notes that there is a struggle between philosophers who align themselves with Frege, and seek to explain speaker's-meaning in terms of sentence-meaning and truth-conditions, and philosophers who argue that this gets matters upside down: according to these latter philosophers, sentence-meaning must be explained in terms of speaker's-meaning rather than vice versa:

> A struggle on what seems to be such a central issue in philosophy should have something of a Homeric quality; and a Homeric struggle calls for gods and heroes. I can at least, though, tentatively, name some living captains and benevolent shades: on the one side, say, Grice, Austin, and the later Wittgenstein; on the other, Chomsky, Frege, and the earlier Wittgenstein.[2]

Strawson terms Grice, Austin, and the later Wittgenstein "theorists of communication-intention": speaker's-meaning, a matter of what speakers *intend to communicate* in uttering particular linguistic tokens or in performing particular speech-acts, is the fundamental notion, and the notion of sentence-meaning is explained (in part) in terms of it. Strawson, aligning himself with these theorists, sums their viewpoint up thus:

> As theorists, we know nothing of human *language* unless we understand human *speech*.[3]

Or in the words of Searle, another communication-intention theorist:

> It is not, as has generally been supposed, the symbol or word or sentence, or even the token of the symbol or word or sentence, which is the unit of linguistic communication, but rather it is the *production* of the token in the performance of the speech-act that constitutes the basic unit of linguistic communication.[4]

Frege and the others, on the other hand, might be called "truth-conditional theorists": sentence-meaning, the possession of truth-conditions, is the fundamental notion, and it is only after explaining this that we are able to proceed to the explanation of speaker's-meaning. Language is explained first, speech afterwards.

The debate between the theorists of communication-intention and the truth-conditional theorists is of such fundamental interest because it promises an answer to the question about the relationship between *language* and *thought*. We have already touched on this in §6.1. If the theorists of communication-intention are right, sentence-meaning, or linguistic-meaning, can be explained in terms of the contents of speakers' mental states: in this sense, the

content of mental states will be *explanatorily prior* to linguistic meaning. In this chapter, we concentrate on one of the most famous theorists of communication-intention, Paul Grice (1913–1988).

7.2 Grice on speaker's-meaning and sentence-meaning

In his short but crucially important article "Meaning", Grice begins by distinguishing between two senses in which the expressions "means", "means something", and "means that", may be taken. The first of these is the *natural* sense, exemplified by:

(3) "Those spots mean measles"
(4) "Those spots didn't mean anything to me, but to the doctor they meant measles"
(5) "The recent budget means that we shall have a hard year".

This type of meaning is sometimes referred to as *indicator-meaning*, since the idea is that spots *indicate* the presence of measles and so on. Grice points out that one feature of natural meaning is that if x means that *p* (where x is an object or objects, *p* a proposition) in the sense of natural meaning, it follows that *p*. Where natural meaning is concerned, it makes no sense to say, for example, "Those spots meant measles, but he hadn't got measles" or "The recent budget means that we shall have a hard year, but we shan't have".

This sense of meaning contrasts with *non-natural* meaning, exemplified by:

(6) "Those three rings of the bell (of the bus) mean that the bus is full"
(7) "That remark, 'Smith couldn't get on without his trouble and strife', meant that Smith found his wife indispensable".

Unlike the natural sense of meaning, there is no entailment from x means that *p* to *p*: it makes perfect sense to say that "Those three rings of the bell mean that the bus is full, but the conductor is mistaken and it isn't actually full" or "That remark . . . meant

that Smith found his wife indispensable, but in fact he deserted her seven years ago". (Why does Grice call these latter cases cases of *non-natural* meaning? It is not a matter of *convention* that spots mean measles, whereas it is in some sense a matter of convention that three rings mean that the bus is full: in other countries, for example, the convention might be different (three rings of the bell might mean that the bus is stopping).)

Grice is concerned purely with the non-natural sense of meaning, and he uses the abbreviation "means$_{NN}$" to distinguish this sense. Grice proceeds in two steps. *First, he aims to explain speaker's-meaning$_{NN}$ in terms of the intentions of the utterer. Second, he aims to explain sentence-meaning$_{NN}$ in terms of speaker's-meaning, so that ultimately we have an account of sentence-meaning$_{NN}$ in terms of utterers' intentions*. So, in the first stage he is concerned with locutions such as

(8) Speaker A meant$_{NN}$ something by sentence x (on a particular occasion of use).

Given an account of this sense of "means$_{NN}$" in terms of A's intentions, Grice goes on to attempt to explain sentence-meaning, by explaining "means$_{NN}$" as it appears in locutions such as

(9) Sentence x means$_{NN}$ (timeless) something (that so-and-so).

Let's work towards Grice's account of (8) and speaker's-meaning. Concentrate on the case where the sentence x is a declarative, or "informative", sentence, such as "Jones makes an excellent cup of tea" (as opposed to "non-informative" sentences or imperatives, such as "Close the window"). We are looking for a set of necessary and sufficient conditions for the truth of

(10) Miller meant$_{NN}$ that Jones makes an excellent cup of tea by his utterance of "Jones makes an excellent cup of tea".

Suppose that I direct this utterance to my colleague Divers, so that Divers constitutes the "audience" for this remark. A first stab at a sufficient condition for (10) might be:

(i) Miller intends his utterance of the sentence to induce the belief that Jones makes an excellent cup of tea.

But as Grice points out, the intention to induce a belief in the audience is not on its own sufficient for a case of meaning$_{NN}$:

> I might leave B's handkerchief near the scene of a murder in order to induce the detective to believe that B was the murderer; but we should not want to say that the handkerchief (or my leaving it there) meant$_{NN}$ anything or that I had meant$_{NN}$ by leaving it that B was the murderer.[5]

The problem here is that I do not intend the detective to *recognise the intention* behind my leaving the handkerchief near the scene of the murder (this would be self-defeating in this case). So we can add, to the condition (i) that Miller intends his utterance of the sentence to induce the belief that Jones makes an excellent cup of tea in Divers, the condition

(ii) Miller also intends Divers to recognise the intention behind his utterance.

Do we now have a sufficient condition for the truth of (10)? Grice thinks not, and that we need to add a further condition. He brings this out by considering the following two cases:

(11) I show Mr X a photograph of Mr Y displaying undue familiarity to Mrs X
(12) I draw a picture of Mr Y behaving in this manner and show it to Mr X.

Grice suggests that whereas we would want to say that by drawing the picture in the case of (12) I mean$_{NN}$ that Mr Y had been

unduly familiar with Mrs X, we would not want to say this in the case of my producing the photograph. Why not? Because my intention that Mr X form the belief that Mr Y had been unduly familiar with Mrs X, and my intention that Mr X recognise this intention of mine, do not play any part in the explanation of why Mr X forms this belief: even if I had not had these intentions – if Mr X had stumbled on the photograph by accident, for example – Mr X would still have formed this belief. This is not so in the example of the drawing. So, in addition to my intention that Mr X form this belief, and my intention that Mr X recognise this intention of mine, *I must also intend that Mr X's recognition of this intention of mine plays a part in the explanation of why Mr X forms the belief*. Thus, the following three conditions are jointly sufficient (and individually necessary) for the truth of (10):

(i) Miller intends his utterance of "Jones makes an excellent cup of tea" to induce the belief that Jones makes an excellent cup of tea in Divers
(ii) Miller intends Divers to recognise the intention behind his utterance of "Jones makes an excellent cup of tea"
(iii) Divers's recognition of Miller's intention (in (ii)) plays a part in the explanation of why Divers forms the belief that Jones makes an excellent cup of tea.

This account can be generalised to deal with "non-informative" cases. Suppose we are at a department meeting, and we are trying to decide who is the best person to take on tea-making duties. I want my colleagues to choose Jones. Then the meaning$_{NN}$ of my utterance of "Jones makes an excellent cup of tea" is Jones's being selected as tea-maker if: I intend to induce my colleagues to pick Jones by means of my utterance, and I intend to induce them to pick Jones *via* their recognition of this intention.

Grice sums up his account of speaker's-meaning$_{NN}$ as follows:

> "A meant$_{NN}$ something by x" is (roughly) equivalent to "A intended the utterance of x to produce some effect in an audience by means of the recognition of this intention"; and we may

> add that to ask what A meant is to ask for a specification of the intended effect.[6]

Having accounted for speaker's-meaning, Grice goes on to consider sentence-meaning. We are looking for a necessary and sufficient condition for the truth of

(13) The sentence "Jones makes an excellent cup of tea" means$_{NN}$ (timeless) that Jones makes an excellent cup of tea.

Grice suggests that (13) is true if and only if tokens of "Jones makes an excellent cup of tea" are *regularly* or *conventionally* associated with the speaker's-meaning$_{NN}$ that Jones makes an excellent cup of tea.[7] The conventional way for a speaker A to mean$_{NN}$ that Jones makes an excellent cup of tea is to utter a token of the type "Jones makes an excellent cup of tea". Of course, there are other uses of this token: a speaker can use it to mean$_{NN}$ that Jones be picked as tea-maker, or to mean$_{NN}$ that Jones is an uninteresting philosopher. But these uses are *non-conventional*, the exception rather than the rule, as it were.

Thus, the notion of sentence-meaning is explained in terms of the notion of speaker's-meaning and the notion of convention, where the notion of speaker's-meaning is explained in terms of utterers' intentions. This is why Grice's account is sometimes known as a "convention plus intention" account of sentence-meaning. Note that Grice's analysis of sentence- or linguistic-meaning is intended to be a *reductive analysis*: the notion of speaker's-meaning is defined in terms of utterers' intentions, in a way that requires no use of the notion of sentence-meaning; and the notion of sentence-meaning is defined in terms of speaker's-meaning and convention, in a way which requires no use of the notion of sentence-meaning. The meaning of language is non-circularly analysed in terms of mental content.

7.3 Searle's modifications: Illocutionary and perlocutionary intentions

Needless to say, many counterexamples to Grice's analysis have been suggested, both by philosophers opposed to his analysis and

by philosophers wishing to refine it. This has led to an extremely complicated series of epicycles, which we shall not attempt to review here. Rather, we look only at two suggested counterexamples and one suggested refinement.

The first counterexample is due to Paul Ziff:

> On being inducted into the army, George is compelled to take a test designed to establish sanity. George is known to be an irritable academic. The test he is being given would be appropriate for morons. One of the questions asked is: "What would you say if you were asked to identify yourself?". George replied to the officer asking the question by uttering (14):
>
> (14) Ugh ugh blugh blugh ugh blug blug.[8]

George intends the officer to be offended, and, moreover, he intends to induce the officer's being offended by means of the officer's recognition of this intention. So according to Grice's analysis, George means_{NN} something by (14). But according to Ziff, this is implausible. George does not mean *anything* by (14) (that is the whole point of his uttering it in this situation).

The second counterexample is due to John Searle:

> Suppose that I am an American soldier in WW2 and that I am captured by Italian troops. And suppose also that I wish to get these troops to believe that I am a German officer in order to get them to release me. What I would like to do is to tell them in German or Italian that I am a German officer. But let us suppose that I don't know enough German or Italian to do that. So I attempt to put on a show of telling them that I am a German officer by reciting those few bits of German that I know, trusting that they don't know enough German to see through my plan. Let us suppose I know only one line of German, which I remember from a poem I had to memorize in a high school German course. Therefore I, a captured American, address my Italian captors with the following sentence: "Kennst du das Land, wo die Zitronen blühen?"[9]

Here, the soldier intends to induce the belief that he is a German officer in the Italians, and he intends that they form this belief in virtue of their recognition of this intention. So, according to Grice's account the following is true:

(15) The soldier means$_{NN}$ that he is a German officer by "Kennst du das Land, wo die Zitronen blühen?".

But according to Searle this is implausible:

> In this case it seems plainly false that when I utter the German sentence what I mean is "I am a German officer", or even "Ich bin ein deutscher Offizier", because what the words mean is, "Knowest thou the land where the lemon trees bloom?". Of course, I want my captors to be deceived into thinking that what I mean is "I am a German officer", but part of what is involved in the deception is getting them to think that that is what the words which I utter mean in German.[10]

The diagnosis of why Grice's account is susceptible to counterexamples like the two we have considered is, according to Searle, that it suffers from two defects:

> First of all, it fails to distinguish the different kinds of effects – perlocutionary versus illocutionary – that one may intend to produce in one's hearers, and it further fails to show the way in which these different kinds of effects are related to the notion of meaning. A second defect is that it fails to account for the extent to which meaning is a matter of rules or conventions. That is, [Grice's] account of meaning does not show the connection between one's meaning something by what one says and what that which one says actually means in the language.[11]

In order to explain how Searle attempts to amend Grice's account so that it is freed of these defects, we must first explain the distinction between *illocutionary* and *perlocutionary* effects.

Note that in our final quote from Grice in §7.2, he said that to ask what A meant by x is to ask for a specification of the effect of

his utterance of x. Of course, not all effects of the utterance of x are relevant to its speaker's-meaning, and we looked at Grice's attempt to narrow down the effects to those that *are* relevant to speaker's-meaning: the effects must be intended, induced in an audience, and the upshot of the audience's recognition of the relevant intention. *Searle suggests that we narrow down even further the class of effects relevant to speaker's-meaning. In addition to the conditions just mentioned, the effects must be illocutionary rather than perlocutionary*. An illocutionary effect is one which is in part the content of an illocutionary intention, *where an illocutionary intention is an intention that is fulfilled simply on the basis of the audience's recognition of the presence of the intention*.

Examples of illocutionary effects include: Jones's is given a report, Jones is given a warning, Jones is given a promise, Jones is given a proposal, Jones is given an order, Jones is thanked. In each case, all that is required for the occurrence of the effect is Jones's recognition that the speaker has the relevant intention: if Jones recognises my intention to warn him in uttering "I warn you", then he is thereby warned; if Jones recognises my intention to thank him in uttering "I thank you", then he is thereby thanked; if Jones recognises my intention to issue him with an order in uttering "I order you", then he is thereby ordered.

A perlocutionary effect, on the other hand, is one which is in part the content of a perlocutionary intention, where a perlocutionary intention is an intention that is *not* fulfilled simply on the basis of the audience's recognition of the presence of the intention. Examples of perlocutionary effects include: Jones forms a certain belief, Jones is amused, Jones is impressed, Jones is embarrassed. In each case, Jones's recognition that the speaker has the relevant intention does not guarantee the occurrence of the effect: if Jones recognises my intention to induce him to form a certain belief by means of an utterance, he may nevertheless fail to form the belief in question; if Jones recognises my intention to amuse him by means of an utterance, he may nevertheless fail to be amused; if Jones recognises my intention to impress him by means of an utterance, he may nevertheless be unimpressed; and if Jones recognises my intention to embarrass him by means of an utterance, he may fail to be embarrassed.

Searle thus amends Grice's account by requiring:

(iv) that the speaker's intentions be illocutionary and
(v) that he intends to respect the conventions governing the use of the relevant expression in the language concerned:

In the performance of an illocutionary act the speaker intends to produce a certain effect by means of getting the hearer to recognise his intention to produce that effect, and furthermore, if he is using words literally, he intends this recognition to be achieved in virtue of the fact that the rules for using the expressions associate the expressions with the production of that effect.[12]

How does this modification deal with the two counterexamples we described earlier? The first counterexample is taken care of, since the speaker's intention to offend the officer is not illocutionary: the officer's recognition that the speaker intends to offend him by means of the utterance does not guarantee that he will actually take offence. Moreover, "Ugh ugh . . ." is not conventionally associated with the hearer's being offended. So the modified account can accommodate the thought that the speaker's utterance of (14) does not mean_{NN} anything. In the second counterexample, the intention is not illocutionary either, since someone's recognising that you intend to induce a belief in them by means of a given utterance does not guarantee that they will form this belief. Moreover, the American soldier does not intend the Italians' recognition of his intention to be achieved in virtue of the fact that in German the effect of inducing the belief that one is German is conventionally associated with the utterance of "Kennst du das Land, wo die Zitronen blühen?". In German this sentence is conventionally associated with an altogether different sort of effect (that which consists in a certain question being answered). So, the modified account can accommodate the thought that the utterance of "Kennst du das Land, wo die Zitronen blühen?" does not mean_{NN} "I am a German officer".

It is worth noting that the second part of Searle's modification – dealing with the rules which conventionally govern the use of the expressions concerned – is necessary. Because the intention is to induce the belief in the Italians that one is a German officer, it might seem as if the first condition – requiring that the speaker's intentions be illocutionary – is enough to deal with this example. But consider the following example. The intention to announce that so and so is an illocutionary intention: mere recognition on the part of the audience that one intends to announce so and so is enough to guarantee that one *has* announced so and so. So suppose that the example is as before, except that the American soldier intends to *announce* to the Italians that he is German by means of uttering "Kennst du das Land, wo die Zitronen blühen?". Then condition (iv) will not kick in. But condition (v) will kick in, since in German the sentence "Kennst du das Land, wo die Zitronen blühen?" is not conventionally associated with the illocutionary act of announcing that one is a German officer.[13]

However, this shows that there is a problem with Searle's modified version of Grice's analysis. The problem is that the analysis now appears to be circular. Recall that Grice's central idea is that we can non-circularly define a notion of speaker's-meaning in terms of speaker's intentions, and then non-circularly define the notion of sentence-meaning in terms of speaker's-meaning and convention. The result is supposed to be a non-circular analysis of sentence-meaning. *But the second condition introduced in Searle's modification seems to presuppose the notion of sentence-meaning: that is, the speaker's-meaning conventionally associated with tokens of a sentence-type.* This renders the account of sentence-meaning circular: we define this in terms of speaker's-meaning, the definition of which presupposes an account of sentence-meaning. Is there any way Searle can respond to this charge of circularity? In the next section we go on to look at further problems faced by Gricean analyses.

7.4 Objections to Gricean analyses

In this section, we'll look at two objections to Gricean accounts which have been developed by Mark Platts. We'll see that although the first of these objections fails, the second objection does pose a

very serious problem for Gricean accounts. We'll then develop a couple of further worries for Grice. In the next section we'll look at Simon Blackburn's response to Platts's objection, and argue that Blackburn's response fails to solve the problem.

Platts's first objection also concerns an alleged circularity in Gricean accounts. It starts with the observation that to find out what intentions a person has typically involves knowing what the expressions of their language mean:

> Utterers' intentions are not recognised by unfailing intuition, nor do Acts of God figure large. It is perhaps possible that very simple intentions be detected quasi-behaviouristically; but for intentions of any fair degree of complexity, this is simply implausible, the behavioural guide being too *inexact*. Any explanation of how such intentions are recognised will inevitably rely upon the audience's recognition of the literal meaning of the sentence; that meaning is the route to the speaker's intentions, the reverse journey usually being impossible.[14]

My meaning$_{NN}$ that Jones makes an excellent cup of tea by my utterance of "Jones makes an excellent cup of tea" requires that I have certain intentions and that I intend that my audience should recognise those intentions. So my audience's recognising that I mean$_{NN}$ that Jones makes an excellent cup of tea requires that they recognise that I have these intentions. But this recognition will usually have to proceed via their understanding some verbal expression of these intentions, which in turn requires their knowledge of the meanings of those expressions. Speaker's-meaning$_{NN}$ cannot be elucidated in terms of speakers' intentions, because knowledge of speaker's intentions *presupposes* knowledge of speaker's-meanings$_{NN}$.

But this objection actually misses the mark. To see this, we have to appreciate the nature of the Gricean claim that mental content is prior to linguistic meaning. As Martin Davies has pointed out

> The kind of priority that concerns us here is priority in the order of philosophical analysis or elucidation. To say that the notion of X is *analytically prior* to the notion of Y is to say that Y can be analysed or elucidated in terms of X, while the analysis or elucidation of X itself does not have to advert to Y.[15]

Grice is claiming that the notion of speakers' intending such and such is *analytically prior* to the notion of sentence-meaning. In addition to analytic priority there is another notion of priority, epistemic priority:

> To say that X is *epistemologically prior* to Y is to say that it is possible to find out about X without having to proceed via knowledge about Y, whereas finding out about Y has to go by way of finding out about X.[16]

But analytic priority is one thing, epistemological priority another: the two notions are logically independent. So it does not follow from the fact that X is analytically prior to Y, that X is epistemologically prior to Y. For example, although the details are controversial and complicated, most philosophers would accept that the notion of personal identity should be analysed in terms of the notions of physical and psychological continuity: but it doesn't follow from this analysis that when I meet you in the street I need to advert to these complex facts about continuity in order to know that you are the same person I was talking to yesterday (I do that by *looking* at you). So in particular, Grice, who is committed to the claim that speaker's intentions are analytically prior to sentence-meaning, is not committed to the claim that it is possible to find out about speaker's intentions without having to proceed via knowledge of facts about sentence-meaning. Platts's first objection thus fails to touch the claim of analytic priority which is definitive of Gricean accounts.

Platts's second objection is more damaging:

> On Grice's theory, sentence-meaning is defined in terms of the intentions with which the sentence is uttered, along, perhaps, with the response standardly secured in an audience by that utterance. Now, as an account of the meanings of sentences in natural languages this will not do for a simple reason: the majority of such sentences, natural languages containing a denumerable infinity of sentences, will never be uttered. They will therefore not be uttered with any intentions, nor will their utterance induce any response in an audience. What, then, can Grice say about these unuttered sentences?[17]

Given that these sentences have never actually been uttered with any intentions, it seems that Grice's only option is to try to account for their meaning in terms of the intentions with which they *would* be uttered, and the responses they would elicit, were that ever to happen. That is, Grice can attempt to account for the meanings of unuttered sentences in terms of *hypothetical intentions*. But not *all* hypothetical intentions can be relevant to meaning; so, Platts asks, what *constraints* are there on the hypothetical intentions which constitute the meaning of a particular sentence? If there are *no* such constraints, "the meanings of unuttered sentences will be left completely indeterminate: they *could* mean anything, so they *do* mean nothing".[18] On the other hand, if there are constraints, what are they? Platts argues that the only answer available to Grice at this point is that the hypothetical intentions are constrained by the meaning of the sentence, thus rendering the analysis circular:

> Generally, the constraint upon the hypothetical intentions with which a sentence can be uttered, and upon the audience's responses to such an utterance, is precisely the meaning of the sentence . . . If this is correct, the attempt to define the meanings of unuttered sentences in terms of hypothetical intentions and responses is hopeless: for it presupposes a prior notion of sentence-meaning.[19]

This difficulty is a reflection of the problems Gricean analyses face in attempting to account for the *compositionality* of meaning: that the meaning of a given sentence, for example, is determined by the meanings of its constituent words and their mode of combination. Consider how the Fregean position developed in Chapters 1 and 2 can deal with the meanings of unuttered sentences by means of theses 2 and 11:

Thesis 2: The semantic value of a complex expression is determined by the semantic values of its parts.

Thesis 11: The sense of a complex expression is determined by the senses of its constituents.

Theses 2 and 11 – the principles of compositionality for semantic value and sense, respectively – tell us how the meaning of a complex sentence is determined, regardless of whether that sentence has been uttered or not. The problem for the Gricean position is that it is difficult to see how it can contain analogues of these theses. The Gricean would in some sense have to mimic the simple Fregean semantic theory that we sketched in §1.7, and correlate intentions (or more generally intentional states) of speakers with the axioms of that theory. The idea would then be that the intentions corresponding to complex unuttered sentences could be generated systematically from the intentions corresponding to the theory's axioms. The intentions corresponding to whole sentences – dealt with by theorems in the semantic theory – would be determined by the intentions corresponding to words and modes of combination – dealt with by the theory's axioms.

There are two main problems with this idea. First of all, there are good reasons for thinking that if there are states of speakers which correspond to the axioms of a semantic theory for their language, they cannot be intentions (or propositional attitudes of any sort). This is because there is a constraint on propositional attitudes, such as intentions, which the states of speakers corresponding to semantic axioms do not satisfy. This point has been well made by Gareth Evans and Crispin Wright. Evans and Wright argue that there is a necessary condition which all genuine intentional states must satisfy to qualify as such, and that the states of speakers which correspond to semantic axioms do not satisfy this condition. In order to motivate this condition they ask us to contrast, on the one hand, the belief that a man might have to the effect that a certain substance is poisonous, with the disposition that a rat might have to avoid a similarly contaminated substance. Can we describe the rat as having a genuine *belief* that the substance is poisonous? Evans and Wright suggest not: for whereas in the case of the man the belief is, to use Evans's phrase, "at the service of many distinct projects", and can interact with others of his beliefs and desires to produce new beliefs and desires, none of this obtains in the case of the rat's disposition. In the case of the man, for instance, the belief could be at the service of projects such as killing an adversary, retaining good health, getting out of an obligation by taking a small dose, to name but a few. None of

this is possible in the case of the rat: the putative "belief" is harnessed to the single "project" of avoidance of the substance. This is supposed to be a reflection of the fact that propositional attitudes and intentional states, such as beliefs, desires, and intentions, come in articulated systems or holistic networks. And it is because genuine beliefs come in such networks and can thus interact with other beliefs, that they can indeed be at the service of many distinct projects. For example, the man's belief that the substance is poisonous can be at the service of the project of getting out of a particular obligation because that belief can, together with the beliefs that a small amount of the substance causes only a mild illness and that a mild illness will release him from the obligation, lead to the belief that taking a small amount of the substance will enable him to avoid fulfilling the obligation.

A crude version of the constraint suggested by the Evans–Wright discussion might therefore run as follows:

> A state P of an agent W is a genuine propositional attitude or intentional state only if P can interact with others of W's propositional attitudes and intentional states to produce new propositional attitudes – thus putting P at the service of many distinct projects of the agent.

Evans and Wright both claim that states of speakers corresponding to semantic axioms violate the above constraint. Far from being at the service of many distinct projects the speaker's possession of such a state is, says Evans,

> Exclusively manifested in speaking and understanding a language; the information is not even potentially at the service of any other project of the agent, nor can it interact with any other beliefs of the agent (whether genuine beliefs or other "tacit" beliefs) to yield further beliefs.[20]

While Wright puts it like this:

> The (implicit) knowledge of a meaning-theoretic *axiom* would seem to be harnessed to the single project of forming beliefs about the content of sentences which contain the expression, or exemplify the mode of construction, which it concerns

and he asks the following (rhetorical) question:

> What is supposed to be the role of *desire*? What is the (implicit?) desire which explains why the subject puts his axiomatic beliefs to just this use, and what are the different uses to which they might be put if his desires were different?[21]

No plausible answers to these questions suggest themselves, so the conclusion is that states of speakers corresponding to semantic axioms cannot plausibly be viewed as propositional attitudes. *A fortiori*, they cannot be viewed as intentions, and this surely bodes ill for the Gricean attempt to accommodate compositionality.[22]

Even if the Gricean can deal with this worry, there is a further problem. Suppose that speakers could have intentions, or intentional states, corresponding to the axioms of a semantic theory. What would the content of such intentions and intentional states be? What intention, for example, would correspond to the axiom dealing with the proper name "Plato"? The only plausible answer seems to be something along the following lines: *the speaker intends to use the word "Plato" to refer to Plato and only to Plato*. But from a Gricean point of view this is hopeless: the content of the intention explicitly concerns the *semantic properties of a linguistic expression*, so that Grice's attempt to give a non-circular analysis of the semantic notions applicable to linguistic expressions is scuppered.[23]

7.5 Response to Blackburn

Blackburn has attempted to respond to Platts's dilemma concerning "hypothetical intentions" which we outlined in the previous section. He argues that Platts's objection is confused:

> It is certainly true that although a speaker will at any time only have understood a certain set of sentences which he has been exposed to, he is equipped to go on and understand new ones, and will standardly do so in just the way other speakers would as well. That is fortunate. But it is not some mysterious thing, the meaning of the new sentence which "constrains" the

> speakers and explains this identity in psychology . . . nothing constrains a group . . . except their training and the way they find it natural to take that training.[24]

Blackburn thus thinks that Grice can steer through the horns of Platts's dilemma by viewing the hypothetical intentions in question as being constrained by training and the ways speakers naturally react to such training.

But this surely gets matters back to front. Training no doubt constrains the verbal behaviour of speakers, but it can only do the work necessary to avoid Platts's dilemma if it is training relevant to the grasp of sentence-meaning. Just as Platts asked after the constraints on hypothetical intentions, we can go further and ask after the constraints on training. Platts's dilemma simply re-emerges further down the line for Blackburn. If a speaker's training with a given sentence is unconstrained, then "the meanings of that sentence will be left completely indeterminate: it *could* mean anything, so it *does* mean nothing". On the other hand, if there is a constraint, what is it? As before, the only available answer seems to be the literal meaning of the sentence concerned, and the circularity of this is again evident.

Blackburn's response to Platts thus fails. But in addition to the response to Platts, Blackburn also has a positive suggestion about how the Gricean can accommodate compositionality. The passage in which he develops this suggestion is worth quoting at length:

> One of the merits of a convention-belief approach is that it concentrates upon the total act of communication – the whole desire or belief communicated by a whole sentence. The presence of a word is subsidiary – a word is something whose presence is a meaning-determining feature of a sentence. We can illustrate this in the figure of the radical interpeter. His initial hypothesis is that some native utterance communicates some whole judgement or command – that p. From a number of such hypotheses he can start to extract the features which recur and whose presence seems to be determining the interpretation to be given to any sentence; once this is done he can predict the way in which new sentences will be taken. A convention-belief approach will

> see the second stage as one of correlating recurrent *features* of utterances with recurrent *features* of beliefs which natives seem to be displaying. The presence of the word "fish" may indicate, for instance, that the speaker is displaying a belief about fish. Once there is regularity in the features of sentences which indicate features of beliefs, there is the possibility of using those features in new combinations to display new beliefs, and we have the elasticity of language. Here, then, we have the place of system according to one natural view of language.[25]

Blackburn's idea seems to be that we can reach an interpretation of a body of native beliefs and utterances and then "project" a structure into the language, a structure which determines the meanings of complex sentences. But it is difficult to see what comfort this can give the Gricean. The discernment of structure acts as a *constraint* on the process of interpretation: as such, the process of interpreting whole sentences *presupposes* rather than *delivers* compositionality. The fact that structure constrains interpretation is well made by Elizabeth Fricker:

> Insofar as interpretation is determinate, the primary feature ensuring this is the fact that the assignment of a meaning to one sentence constrains and is constrained by the assignments made to other sentences . . . Some principle for discerning structure in sentences is essential to interpreting a language, and the correctness of an interpretation cannot be considered to be independent of facts about their structure.[26]

This bears some fleshing out. Suppose that we are out to interpret the speech of some previously unmet tribe. We observe one of the natives pointing to a nearby mountain and uttering "Lo, monadi kel guro". What does he mean? We will no doubt initially take the native to be saying something about the nearby peak, and let's suppose, for the sake of argument, that this is in fact what he is trying to do. He might have been saying any number of things about it, maybe that it is icy, that it would be exciting to climb it, that it would command a good view of the neighbouring terrain and so on. What enables us to decide which of these interpretations is the correct one is the contribution the predicate "guro" makes to

the meanings of other sentences in which it appears. For example, suppose at some later time after his utterance of the above sentence we observe the native picking up a glass of milk containing ice cubes and uttering "Lo pazo kel guro"; then the most plausible interpretation of the earlier sentence will be that the mountain in view is icy. There may be alternative interpretations – we might interpret him as having said that the mountain is white – but the method whereby such interpretations are eliminated should be clear enough (we can check to see what he says about a glass of coca-cola filled with ice cubes, for example). Now the crucial point is that we can only employ this method of narrowing down the possible interpretations on the assumption that the predicate "guro" makes the same systematic contribution to the meanings of the sentences in which it figures. And to say that the language contains expressions which make *systematic* and *recurring* such contributions is just to say that the language is semantically structured. The interpretation of one sentence constrains, and is constrained by, the interpretation of others via the presence of such expressions, and so via the presence of semantic structure.

Interpretation thus *presupposes* a principle for discerning structure: it cannot by itself deliver such a principle. Interpretation *assumes* that expressions make systematic and recurring contributions to the meanings of the sentences in which they figure: it therefore cannot explain the possibility of such a system. So it is difficult to see how Blackburn's story really can afford the Gricean an explanation of the place of system in language.

Our conclusion is thus that Blackburn's attempt to respond to the central problems facing Gricean accounts of meaning is a failure. Given this, the prospects for a communication-intention account of sentence-meaning appear bleak. In the next chapter we leave communication-intention accounts of meaning, and return to work in the alternative, truth-conditional tradition. Before that, though, some unfinished business from Chapter 2.

7.6 Strawson on referring revisited

In §2.12 we left hanging the question whether Strawson's main objection to Russell's theory of descriptions is convincing.

The result of the previous two sections, that the prospects for a communication-intention account of sentence-meaning appear bleak, allows us to find against Strawson.

As noted in §7.1, Strawson allies himself with the communication-intention theorists when he asserts "As theorists, we know nothing of human *language* unless we understand human *speech*". This plays a large part in his attack on Russell. Recall from §2.12 Strawson's view that:

> The sentence [(49)] is certainly significant; but this does not mean that any particular use of it is true or false. We use it truly or falsely when we use it to talk about someone; when, in using the expression, "The king of France", we are in fact mentioning someone. The fact that the sentence and the expression, respectively, are significant just is the fact that the sentence *could* be used, in certain circumstances, to say something true or false, that the expression *could* be used, in certain circumstances to mention a particular person.[27]

Strawson is here clearly suggesting the kind of communication-intention view criticised in the previous two sections. This is even clearer when we consider other remarks that he makes in "On Referring":

> To give the meaning of an expression (in the sense in which I am using the word) is to give *general directions* for its use to refer to or mention particular objects or persons; to give the meaning of a sentence is to give *general directions* for its use in making true or false assertions . . . [T]o talk about the meaning of an expression or sentence is . . . to talk about . . . the rules, habits, conventions governing its correct use, on all occasions, to refer or assert.[28]

> [It] is people who mean, not expressions. People use expressions to refer to particular things. But the meaning of an expression is not the set of things or the single thing it may correctly be used to refer to*: the meaning is the set of rules, habits, conventions for its use in referring*.[29]

> The question whether the sentence is significant or not is the question whether there exist . . . language habits, conventions or rules that the sentence logically could be used to talk about something.[30]

Given that these claims play a crucial role in Strawson's critique of Russell, we can conclude that Strawson's main objection to Russell's theory of descriptions misses its mark.[31]

Although Strawson's objection fails, it may point us in the direction of a more compelling difficulty for Russell. We saw in §7.4 that the Fregean can deal with the problem about the meanings of unuttered sentences that so vexes communication-intention views via the two theses:

Thesis 2: The semantic value of a complex expression is determined by the semantic values of its parts.

Thesis 11: The sense of a complex expression is determined by the senses of its constituents.

Russell, however, in rejecting Frege's notion of sense, must reject thereby thesis 11, regarding the compositionality of sense. Is thesis 2, asserting the compositionality of semantic value, strong enough on its own to allow the Russellian to deal with the problem of the meanings of unuttered sentences? Our findings in §2.10 suggest that the Russellian may here face insuperable difficulties of his own. The resolution of this issue, however, must here be left for another occasion.

Chapter summary

We outlined Grice's distinction between natural and non-natural meaning. We explained Grice's attempt to account for (non-natural) sentence-meaning in terms of (non-natural) speaker's-meaning,

then outlined how he proposed to account for speaker's-meaning in terms of the intentions that speakers have to induce beliefs in an audience via their recognition of those intentions. We outlined some counterexamples to Grice's analysis developed by Ziff and Searle, and described how Searle proposes to modify Grice's account in order to avoid them. This involved distinguishing between illocutionary and perlocutionary effects and intentions. We saw that it was unlikely to result in the kind of reductive or non-circular analysis Grice aspired to. We looked at a number of other worries for Grice's account, including the difficulties it faces in dealing with compositionality.

Study questions

1. Construct three examples of your own to illustrate the distinction between natural and non-natural meaning.
2. Use examples of your own to explain the need for conditions (ii) and (iii) in Grice's account of speaker's-meaning.
3. How might Searle respond to the challenge that his alterations (in §7.3) to Grice's account of speaker's-meaning lead to a damaging circularity in the analysis?
4. Construct an example of your own to show that analytic priority does not imply epistemological priority.
5. How might a Gricean account of language deal with the compositionality of meaning?
6. How would you modify conditions (i)–(iii) in Grice's analysis to accommodate Searle's idea that the relevant intentions should be illocutionary?

Further reading

Grice's classic paper "Meaning", as well as many other relevant articles, can be found in his collected papers *Studies in the Ways of Words*. Strawson's views are set out in his "Meaning and Truth". Searle's modification of Grice is contained in his "What Is a Speech-Act?". See also Searle, *Speech-Acts*. Stephen Schiffer's book

Meaning is an attempt to develop a systematic Gricean approach to language, although Schiffer has more recently come to the conclusion that Grice's programme is a failure (see his *Remnants of Meaning*). See also Davies, *Meaning, Quantification, and Necessity*, Chapter 1. For criticism of Grice, see Ziff, "On H. P. Grice's Account of Meaning"; and Platts, *Ways of Meaning*, Chapter 3. Blackburn's response to Platts can be found in Chapter 3 of *Spreading the Word*. The Evans–Wright argument, used in §7.4, can be found in Evans, "Semantic Theory and Tacit Knowledge" and Wright, "Theories of Meaning and Speakers' Knowledge". For further discussion of the Evans–Wright argument, see Miller, "Tacit Knowledge".

A full treatment of the issues dealt with in this chapter would include discussion of the contribution of J. L. Austin and also of Grice's notion of "conversational implicature". See Longworth, "Semantics and Pragmatics".

For further discussion see also Bach, "Speech-Acts and Pragmatics"; Borg, "Intention-Based Semantics"; and Schiffer, "Convention and Intention in the Theory of Meaning".

Notes

1 Strawson, "Meaning and Truth", p.171.
2 Ibid., p.172.
3 Ibid., p.189.
4 Searle, "What Is a Speech-Act?", pp.221–222.
5 Grice, "Meaning", p.217.
6 Ibid., p.220.
7 We cannot go into the question as to the nature of convention here; but most Griceans take as their starting point the account developed in Lewis, *Convention*.
8 Ziff, "On H. P. Grice's Account of Meaning", p.2.
9 Searle, "What is a Speech-Act?", pp.229–230.
10 Ibid., p.230.
11 Ibid., p.229.
12 Ibid., pp.230–231. Note that with the adoption of (iv), Searle must now amend (i), since as originally stated it concerned a perlocutionary intention (the intention to induce a particular belief in

one's audience). Searle's idea is that we should model speaker's-meaning on the case of intending to make a promise rather than on the intention to induce a belief. See ibid., §5.

13 We might now start to wonder whether condition (iv) is redundant. Are there any cases it rules out that are not also ruled out by (v)?

14 Platts, *Ways of Meaning*, p.91.

15 Davies, "Philosophy of Language", p.97.

16 Ibid.

17 Platts, *Ways of Meaning*, p.89.

18 Ibid.

19 Ibid., p.90.

20 Evans, "Semantic Theory and Tacit Knowledge", p.133.

21 Wright, "Theories of Meaning and Speaker's Knowledge", pp.227–228.

22 In the last two paragraphs I've drawn on Miller, "Tacit Knowledge", §2.

23 This difficulty would also afflict any attempt to view the states corresponding to semantic axioms as subdoxastic states, non-propositional attitude states which nevertheless have informational content. The informational content in question would still have to be explicitly semantic. It is an interesting question whether the two difficulties we have developed can be dealt with by harnessing the Gricean account to Evans's account of the nature of the states corresponding to semantic axioms. On Evans's account, these states are neither propositional attitude states nor subdoxastic states, so that neither of our worries would apply. For a discussion of Evans's account, see Miller, "Tacit Knowledge".

24 Blackburn, *Spreading the Word*, p.128.

25 Ibid., pp.129–230.

26 Fricker, "Semantic Structure and Speaker's Understanding", p.59.

27 Strawson, "On Referring", p.330.

28 Ibid., p.327.

29 Ibid., p.328 (emphasis added).

30 Ibid., p.329.

31 It is perhaps worth noting in passing that one of the objections Donnellan raises against Russell and Strawson is that they "make a common assumption about the question of how definite descriptions function: that we can ask how a definite description functions in some sentence independently of a particular occasion upon which it is used" (Donellan, "Reference and Definite Descriptions",

p.282). Although this seems like a fair objection against Russell, the remarks from Strawson that we have just considered seem to suggest that its applicability to Strawson is not straightforward. Donnellan continues his point about Strawson: Although he can sum up his position by saying '"Mentioning" or "referring" is not something an expression does: it is something that someone can use an expression to do', he means to deny the radical view that a 'genuine' referring expression *has* a referent, functions to refer, independent of the context of some use of the expression. The denial of this view, however, does not entail that definite descriptions cannot be identified as referring expressions unless the sentence is being used". If Donnellan is right about this, might it provide Strawson with a means of replying to the objection made against him in the text?

Chapter 8

Sense and truth

Tarski and Davidson

In §1.7 and §2.5 we introduced the idea of a systematic semantic theory in the formal sense: a theory which delivers, for each well-formed declarative sentence of a particular language, a theorem which gives the meaning or sense of that sentence. We saw that for Frege the sense of a sentence can be given by stating its truth-condition:

> Every [sentence] expresses a sense, a thought. That is, by our stipulations, it is determined under what conditions the [sentence] refers to the True. The sense of this [sentence], the thought, is the thought that these conditions are fulfilled.[1]

This is an idea which has been taken up by the influential American philosopher, Donald Davidson (1917–2003):

> To give truth-conditions is a way of giving the meaning of a sentence.[2]

In this chapter, we explain Davidson's idea that a systematic theory of meaning for a natural language can be provided by a theory which generates, for each declarative sentence, a theorem which states its truth-conditions.

8.1 Davidson and Frege

Davidson makes it clear that when he speaks of a theory of meaning, he is primarily interested in the notion which Frege termed sense:

> Frege held that an adequate account of language requires us to attend to three features of sentences: reference [semantic value], sense, and force. Elsewhere I have argued that a theory of truth patterned after a Tarski-type truth-definition tells us all we need to know about sense. Counting truth in the domain of reference [semantic value], as Frege did, the study of sense thus comes down to the study of reference [semantic value].[3]

As we'll see, Davidson says that the study of sense comes down to the study of semantic value, because he thinks that we can – and must – provide a theory of meaning for natural languages which does not involve the invocation of senses as objects or entities. But what about the Fregean semantic theory which we sketched in §1.7? We saw that this theory works by assigning only semantic values to expressions, and displays not only how the semantic value of a sentence is a function of the semantic values of its constituents, but also how specifications of sentences' *truth-conditions* can be generated purely on the basis of assignments of semantic values. It might seem, then, that so far as a systematic theory of meaning in the formal sense is concerned, Frege has already shown us how to bring the study of sense down to the study of semantic value or reference. In fact, however, Davidson thinks that Frege's model for a theory of semantic value (and thereby for a systematic theory of sense) is unsatisfactory, because it provides no useful or explanatory account of how sentence-meaning can be a function of word-meaning. Suppose we were to ask for the meaning of the sentence "Theaetetus flies":

> A Fregean answer might go something like this: given the meaning of "Theaetetus" as argument, the meaning of "flies" yields the meaning of "Theaetetus flies" as value. The vacuity of this answer is obvious. We wanted to know what the meaning of "Theaetetus flies" is; it is no progress to be told that it is the meaning of "Theaetetus flies". This much we knew before any theory was in sight. In the bogus account just given, talk of the structure of the sentence and the meanings of words was idle, for it played no role in producing the description of the meaning of the sentence.[4]

Frege's semantic theory works by assigning "unsaturated" or "incomplete" entities – namely, functions – to predicates and quantifiers as their semantic values, and "saturated" or "complete" entities – The True or The False – to sentences as their semantic values. Davidson finds this type of theory completely unexplanatory, and he writes that "it seems to label a difficulty rather than solve it".[5] Davidson's task, therefore, is to find a model for the construction of systematic theories of meaning which does not work by postulating incomplete entities as the semantic values of predicates or quantifiers. Davidson argues that such a model is provided by a theory of truth along the lines of the definition of truth constructed by the Polish logician, Alfred Tarski (1901–1983). We'll now attempt to get clear on Davidson's motivation for this claim, and then we'll explain the workings of Tarskian theories of truth.

8.2 Davidson's adequacy conditions for theories of meaning

Davidson's strategy, in broad outline, is relatively simple. First of all, he lays down a number of general conditions which any adequate theory of meaning for a natural language must satisfy. It then emerges that one way of providing a theory which meets each of these conditions would be to adapt a theory of truth, in the style of Tarski, for the natural language concerned. We'll concentrate on three general conditions of adequacy of theories of meaning:

(A) **The extensional adequacy condition**: The theory of meaning must generate a theorem which "gives the meaning" of each sentence of the language under consideration.

(B) **The compositionality condition**: The theory should be *compositional*. A theory of meaning is compositional if and only if (a) it has only finitely many axioms, and (b) each of the meaning-giving theorems is generated from the axiomatic base in such a way that the semantic *structure* of the sentence concerned is thereby exhibited.

What motivation is there for seeking compositional semantic theories in preference to their more readily available non-compositional counterparts? Why should the construction of a semantic theory be constrained by the requirement that it reflect the semantic structure of the language concerned? Davidson's answer is that the construction of specifically compositional theories will take us some way towards providing answers to each of the following three questions:[6]

(a) How is it possible, given the finitude of their capacities, for speakers of a natural language to understand a potential infinity of sentences?

(b) How is it possible to understand utterances of previously unencountered sentences?

(c) How is it possible for a natural language to be learnable? (I.e. how is it possible for explicit training with only a relatively small number of sentences to secure competence with a possibly very large set of sentences outwith that initial set?)

Briefly, the axioms of a theory of meaning deal with the semantic properties of the words or subsentential expressions of a language, while the theorems, which are generated from these axioms, deal

with the semantic properties of the sentences of the language. A compositional theory of meaning will therefore show how the meanings of sentences are systematically dependent upon the semantic properties of their constituent words and the way they are put together. Since there are only finitely many words, dealt with by only finitely many axioms in a compositional theory of meaning, we will have an insight into how (a), (b), and (c) can be answered: speakers with finite capacities will be able to understand a potential infinity of sentences because on the basis of a finite set of axioms we will be able to generate a potential infinity of meaning-giving theorems; we understand unfamiliar sentences via understanding the familiar words out of which they are composed and their mode of composition, and this will be modelled on the way in which a meaning-giving theorem for an unfamiliar sentence can be derived from axioms dealing with familiar words; and the way in which meaning-giving theorems can be derived from axioms for familiar words and modes of combination also provides insight into the learnability of natural languages.[7] The third general condition of adequacy of theories of meaning is:

> **(C)** **The interpretation condition**: The theory of meaning must allow us to correctly interpret the speakers of that language, in accordance with constitutive principles governing interpretation (such as the principle of charity).

We'll explain this condition in §8.7.

In short, a theory of meaning must be extensionally adequate and compositional, and it must enable us to interpret the speech of its speakers.[8] Before working towards the sort of theory of meaning which Davidson favours, we'll look at two types of theory which he rejects.

8.3 Intensional and extensional theories of meaning

What sort of theorem "gives the meaning" of a sentence? Davidson first of all considers a theory of meaning which yields, for each sentence of the language, a theorem of the form

(1) s means m

where "s" is replaced by a name of a sentence, and "m" is replaced by a term which refers to or denotes the meaning of that sentence. Meanings are thus conceived as entities, and the job of the theory of meaning is to pair sentences with such entities. Now just as Davidson rejected Frege's semantic theory on the grounds that the assignment of entities to predicates and quantifiers as their semantic values was completely unexplanatory, he also rejects theories of meaning which assign entities to expressions as their meanings or senses:

> The one thing meanings do not seem to do is oil the wheels of a theory of meaning – at least as long as we require of such a theory that it non-trivially give the meaning of every sentence in the language. My objection to meanings in the theory of meaning is not that they are abstract or that their identity conditions are obscure, but that they have no demonstrated use.[9]

We can perhaps view Davidson as applying a version of *Ockham's Razor* – "that entities are not to be multiplied beyond necessity" – to meanings: Davidson will show that everything we want from a theory of meaning can be achieved without bringing in meanings as entities, so that Ockham's Razor tells us that we should not postulate the existence of such entities.

But what about a theory of meaning which attempts to satisfy the extensional adequacy condition by generating theorems of the form

(2) s means that p

where "s" is replaced by a name of a sentence and "p" is replaced by some sentence which gives its meaning. Such a theory does not require the postulation of meanings as entities, and would appear to satisfy the extensional adequacy condition in the most direct way possible. But Davidson rejects the idea that (2) is the correct form for a meaning-giving theorem of a theory of meaning. Davidson's reason for rejecting theories of this form has been well spelled-out by Mark Platts. Platts notes that ". . . means that . . ."

creates what is known as an *intensional context*, a context in which the substitution of expressions having the same semantic value need not preserve the semantic value (truth-value) of the original sentence. (An intensional context is thus a context in which thesis 3 (see §1.4) is violated.) To see the intensionality of ". . . means that . . ." consider the true sentence

(3) "Snow is white" means that snow is white.

Now, the sentence "Snow is white" has the same semantic value (truth-value) as the sentence "Grass is green". But when we substitute the latter for the former in (3), we obtain the false sentence

(4) "Snow is white" means that grass is green.

The problem now is that

> The only systematic account of intensional contexts that survives even the most cursory glance explains the intensionality by reference to the notion of *meaning*.[10]

The reason that the move from (3) to (4) is invalid is precisely that the sentences "Snow is white" and "Grass is green" *do not have the same meaning*. This contrasts with the move from

(5) "Jones is an oculist" means that Jones is an oculist

to

(6) "Jones is an oculist" means that Jones is an eye-doctor

which is valid, because the sentences "Jones is an oculist" and "Jones is an eye-doctor" do in fact have the same meaning.

This shows that a theory of meaning which yielded theorems of the form of (2) would presuppose the very notion which it is the purpose of a theory of meaning to elucidate. As Platts puts it

> The general explanation of intensional contexts requires reference to meanings; and explanation of the properties of any given intensional construction requires reference to the meanings

> of the particular expressions occurring within it. If this is correct, there is no point to the employment of intensional idioms within the systematic, axiomatic component of a theory of meaning.[11]

A theorem of the form of (2) creates a context the general explanation of which requires the use of the notion of meaning; a theory of meaning which generates theorems of the form of (2) therefore cannot provide a genuinely explanatory account of meaning.

In addition, the logical machinery required in a theory of meaning which is itself couched in a language which contains intensional constructions will be immensely complicated. As we'll see, the logical machinery required in a theory of meaning which is couched in a purely extensional language can be provided by the familiar and relatively simple Fregean logic which we described in Chapter 1.

The upshot of these considerations is that a theory of meaning should be extensional in the sense that it does not postulate intensional entities, such as meanings or senses, and that it should be couched in a language which itself contains no intensional constructions. But how is it possible to give a theory of meaning in purely extensional terms – in other, words, to show that the study of sense comes down to the study of semantic value?

8.4 Extensional adequacy and Tarski's Convention (T)

Davidson's answer to this question is that a theory of meaning is extensionally adequate if it yields, for each sentence of the language under consideration, a theorem of the form

(7) s is true if and only if p

where "s" is replaced by a name of the sentence and "p" by the sentence in question itself. Thus, a meaning-giving theorem for the sentence "snow is white" would be

(8) "Snow is white" is true if and only if snow is white.

This is a case where we are trying to give a theory of meaning for English in English itself. The language *for* which we are trying to give a theory of meaning is called the *object-language*, and the language *in* which that theory is given is called the *meta-language*. Thus, in the case of a theory yielding (8) the object-language is English, and the meta-language is English also. But of course, there is no reason why the object- and meta-languages have to be identical. For example, we could give a theory of meaning for French (the object-language) in English (the meta-language). Such a theory would yield not (8), but rather

(9) "La neige est blanche" is true if and only if snow is white.

Likewise, a theory of meaning for German (the object-language) could be given in English as meta-language. Such a theory would yield

(10) "Schnee ist weiss" is true if and only if snow is white.

Obviously, in cases such as these, where the object- and meta-languages are different, we cannot require that "p" in (7) be replaced by the sentence concerned itself, because the sentence will not belong to the meta-language in which the theory is couched. For example, in the case where the object-language is German and the meta-language is English, we cannot require that "p" be replaced by "Schnee ist weiss" itself because this sentence does not belong to the meta-language, English. What we require in this sort of case is rather that "p" be replaced by a sentence of the meta-language which *translates* it.

In general, then, Davidson requires that the theory of meaning yield a theorem of the form (7) for each sentence of the object-language, where "s" is replaced by a name of the relevant sentence, and "p" by that sentence itself (in the case where the object-language is contained within the meta-language) or by a translation of that sentence into the meta-language (in the case where the object-language is not contained within the meta-language).

That is simply a statement of Davidson's condition of extensional adequacy. But how did Davidson reach it? And how is the connection with Tarskian theories of truth forged? Platts suggests that Davidson's argument for taking the condition of extensional adequacy in this way is contained in the following passage:

> Anxiety that we are enmeshed in the intensional springs from using the words "means that" as a filling between description of sentence and sentence, but it may be that the success of our venture depends not on the filling but on what it fills. The theory will have done its work if it provides, for every sentence s in the language under study, a matching sentence (to replace "p") that, in some way yet to be made clear, "gives the meaning" of s. One obvious candidate for matching sentence is just s itself, if the object-language is contained in the meta-language; otherwise a translation of s in the meta-language. As a final bold step, let us try treating the position occupied by "p" extensionally: to implement this, sweep away the obscure "means that", provide the sentence that replaces "p" with a proper sentential connective, and supply the description that replaces "s" with its own predicate.[12]

Following Platts, we can break the argument down into six separate stages.

Stage 1: The condition of extensional adequacy, in its most general form, states that for each sentence of the language under consideration, the theory of meaning must generate a theorem which in some sense "gives its meaning".

Stage 2: Requiring that the theory of meaning generate a theorem of the form of either (1) or (2) is no good as a way of cashing out extensional adequacy: such theories would be obscure and unexplanatory.

Stage 3: Since we are aiming for a theorem which spells out an equivalence in point of meaning, we could try replacing "means that" with the biconditional "if and only if", familiar from Fregean logic. This would produce something of the form

(11) s if and only if p.

Stage 4: But (11) is not syntactically well-formed, for the expression on the left-hand side is not a sentence, but rather a *name* of a sentence. Remember that when you enclose a sentence in quotation marks, as in "Snow is white", what you get is a name of the enclosed sentence, not another sentence. When we have

(12) "Snow is white" if and only if snow is white

we actually have something like

(13) Peter if and only if snow is white

which is simply a piece of ungrammatical nonsense.

Stage 5: One way to render (12) grammatical would be to make the left-hand side into a sentence by concatenating the name with a predicate. What we require for extensional adequacy is thus that for each sentence it yield a theorem of the form

(14) s is X if and only if p

where "s" is replaced by a name of that sentence, and "p" is replaced by that sentence itself (if the object-language is contained within the meta-language), or by a translation of that sentence into the meta-language (if the object-language and the meta-language are distinct).

Stage 6: In his paper "The Semantic Conception of Truth", Tarski wrote:

> We shall call any equivalence of the form
>
> (T) s is true if and only if p
>
> (with "p" replaced by any sentence of the language to which the word "true" refers, and "s" replaced by a name of this sentence) an "equivalence of the form (T)" . . . Now at last we are able

> to put into a precise form the conditions under which we will consider the usage and the definition of the term "true" as adequate from the material point of view: we wish to use the term "true" in such a way that all equivalencies of the form (T) can be asserted, and we shall call a definition of truth "adequate" if all these equivalencies follow from it.[13]

Tarski is here laying down the following as a condition of adequacy on any proposed definition of "is true": any acceptable definition of truth should have as a consequence all instances of the (T) schema

(T) s is true if and only if p

where "s" is replaced by a name of any sentence of the language for which truth is being defined (the object-language), and "p" is replaced by that sentence or a translation thereof, depending on whether or not the object-language is contained in the meta-language.

This is Tarski's condition of *material adequacy* on definitions of "is true" (sometimes known as Convention (T)).[14] Given this condition, we know that the predicate "X" in (14) will at least be co-extensive with the truth-predicate for the object-language. So we might as well take "X" in (14) to be the truth-predicate for the object-language. The condition of extensional adequacy on theories of meaning is thus that for each sentence of the object-language, they yield a theorem of the form

(7) s is true if and only if p

where "s" is replaced by a name of the relevant sentence, and "p" by that sentence itself (in the case where the object-language is contained within the meta-language) or by a translation of that sentence into the meta-language (in the case where the object-language is not contained within the meta-language).

Interpreting the extensional adequacy condition in this way has a number of interesting consequences. A theorem of the form of (7) gives the *truth-conditions* of the sentence named on the left-hand side. We have thus arrived, by an independent route, at Frege's

conclusion that the sense of a sentence can be given by stating its truth-condition. Also, given this interpretation, Davidson's condition of extensional adequacy on theories of meaning is identical to Tarski's Convention (T), his condition of material adequacy on theories of truth. *Importantly, Tarski provided a formal account which shows how a theorem of the form of (7) can be generated for each sentence of a formal language – basically, the language of Frege's predicate logic – in a way which respects the compositionality constraint (B).* What this suggests is that we can use systematic truth-theories of the sort developed by Tarski as models for compositional theories of meaning for natural languages. This is exactly what Davidson proposes to attempt: given any stretch of natural language, we first of all attempt to formalise it into Frege's logical language, and once it is in this form we construct a Tarski-style theory which shows how the truth-conditions of sentences are systematically dependent upon the semantic properties of their parts.

Before examining further Davidson's proposal to use a Tarski-style truth-theory as a model for a theory of meaning, we must pause to see how Tarski-style truth-theories actually work.

8.5 Tarskian truth-theories

We'll first of all sketch a truth-theory for a very simple form of the language of propositional logic, a language whose syntax is given by the rules stated in §1.2. In such a language (call it L), we distinguish between *atomic sentences*, represented by individual sentential constants, such as P, Q, R, and so on, and *complex sentences*, such as P & Q, P v Q, –P, and so on, which result from combining the sentential connectives and the atomic sentences (there will be infinitely many of these). In order to keep things simple, we'll suppose that this language has only three atomic sentences, "P", "Q", and "R", and we'll ignore the conditional and biconditional connectives. What we are after is a theory which shows how the truth-conditions of complex sentences can be derived systematically from the truth-conditions of the atomic sentences and clauses governing the connectives. This theory can actually be stated quite simply:

(1a) "P" is true if and only if snow is white
(1b) "Q" is true if and only if grass is green
(1c) "R" is true if and only if penguins waddle
(2a) For any sentences A, B, "A & B" is true if and only if "A" is true and "B" is true
(2b) For any sentences A, B, "A v B" is true if and only if "A" is true or "B" is true
(2c) For any sentence A, "–A" is true if and only if "A" is not true.[15]

Given these axioms, we can generate, for any complex sentence of the language, a theorem which specifies its truth-condition. Take "P & –Q", for example. By step (2a)

"P & –Q" is true if and only if "P" is true and "–Q" is true.

By step (2c)

"P & –Q" is true if and only if "P" is true and "Q" is not true.

By steps (1a) and (1b)

"P & –Q" is true if and only if snow is white and grass is not green

Thus, we have derived a truth-condition for "P & –Q" from axioms spelling out the semantic properties of its parts. Note that the truth-condition of the complex sentence is derived from the truth-conditions of its constituents "P", "–Q", with the truth-condition for "–Q" in turn derived from the truth-condition for "Q".

Now, suppose we wanted to use the above truth-theory as a model for a theory of meaning for a fragment of a natural language, such as English. The first thing we would have to do would

be translate the sentences of this fragment of English into the language L. Having done that, we could then go on to apply the truth-theory. But as we saw in Chapter 1, the capacity of a simple propositional language such as L to translate sentences of natural language is very limited. In order to capture the validity of even quite simple argument forms, we need the more complicated language of *predicate* logic. So can we give a truth-definition for the language of predicate logic? Immediately, there appears to be a problem with simply copying the truth-theory for L. The truth-theory for L derives the truth-conditions of a complex sentence of L from the truth-conditions of its constituents, but in many cases the constituents of the complex sentences of the language of predicate logic won't *even have* truth-conditions. For example, consider

(15) $(\forall x)(Fx \,\&\, Gx)$.

Suppose we wanted to spell out the truth-conditions of (15) in a way which reflects the contributions made by the semantic properties of its constituents. The constituents of (15) are the open sentences "Fx", "Gx", and because the variables in these sentences are unbound, "Fx" and "Gx" are neither true nor false. The complex sentences of the language of predicate logic are composed out of open sentences; open sentences are neither true nor false, and therefore do not have truth-conditions. So giving a truth-theory for the language of predicate logic cannot consist in showing how the truth-conditions of complex sentences are a function of the truth-conditions of their constituents.

What we require is some analogue, for open sentences, of the notion of truth as applied to closed sentences (sentences, such as "$(\forall x)Fx$", in which the variables are bound by quantifiers). Tarski suggests that the analogue of truth for open sentences is *satisfaction*: just as closed sentences are either true or false, open sentences are either satisfied or not by objects. For example, the open sentence "Fx", where we take "F" to represent the predicate ". . . is a philosopher", is satisfied by Socrates and Plato, but not by Clint Eastwood or Hillary Clinton. Tarski's suggestion regarding the construction of a truth-theory for the language of predicate logic is therefore as follows: first of all define the notion of satisfaction, and then define the

notion of truth in terms of this notion of satisfaction. This will allow us to derive the truth-conditions of complex sentences of predicate logic in terms of the semantic properties of their constituents.

For technical reasons, which we need not go into here, Tarski takes satisfaction to be a relation not between open sentences and individual objects, but rather between open sentences and *infinite ordered sequences of objects*. An example of an ordered sequence of objects would be <Socrates, Plato, London, Donald Trump, . . .>. Because we are dealing with *ordered* sequences, this would count as different from <Plato, Socrates, London, Donald Trump, . . .>. The convention is that an open sentence with *n* variables is satisfied by a sequence just in case it is satisfied by the first *n* members of that sequence: the subsequent members of the sequence are simply ignored. Thus, the open sentence Fx above would be satisfied by both <Socrates, Plato, London, Donald Trump, . . .> and <Plato, Socrates, London, Donald Trump, . . .> (since there is only one variable in Fx, the crucial question is whether the first member of these sequences satisfies F, which they do in this case). Likewise, the open sentence Gx_1x_2, where "G" is ". . . wrote about . . .", would be satisfied by <Plato, Socrates, London, Donald Trump, . . .> but not by <Socrates, Plato, London, Donald Trump, . . .> (since Plato wrote about Socrates, but not vice versa).

We'll now show how Tarski's account of truth in terms of satisfaction works for an extremely simple predicate language L*. This language contains a one-place predicate "F", translating ". . . is a philosopher", a one-place predicate "G", translating ". . . is Greek", variables, the universal quantifier, and the existential quantifier. We first of all give the definition of satisfaction for L*.[16]

Definition of satisfaction for L*: We take X, Y, and so on to range over infinite ordered sequences of objects, and A, B, and so on to range over open sentences of L*. X_i denotes the ith member of the sequence X (so, for example, for X = <Socrates, Plato, Rod Stewart, Adele, . . .>, X_2 = Plato, X_3 = Rod Stewart, and so on).

(1) ∀X: X satisfies "Fx_1" if and only if X_1 is a philosopher
(2) ∀X: X satisfies "Gx_1" if and only if X_1 is Greek
(3) ∀X, A: X satisfies "–A" if and only if X does not satisfy "A"
(4) ∀X, A, B: X satisfies "A & B" if and only if X satisfies "A" and X satisfies "B".
(5) ∀X, A: X satisfies "$(\exists x_1)A$" if and only if there is a sequence Y, differing from X in at most the first place, such that Y satisfies "A".
(6) ∀X, A: X satisfies "$(\forall x_1)A$" if and only if every sequence Y, differing from X in at most the first place, is such that Y satisfies "A".

Clauses (1)–(4) are pretty obvious, but (5) and (6) are far from obvious. We'll explain why Tarski spells them out as he does very shortly. Before doing so, though, we now define truth for closed sentences in terms of the notion of satisfaction just defined (again, we'll see why Tarski does it in this way in due course):

Definition of "true" for L*: A closed sentence of L* is true if and only if it is satisfied by *all* sequences.

The main reason for spelling out clauses (5) and (6) and the definition of truth for closed sentences in this way is that they work: given these definitions we can generate the intuitively correct truth-conditions for the closed sentences of L*. We'll illustrate this by means of a few examples. First, though, note that given our definition of satisfaction, a closed sentence is satisfied by all sequences if and only if it is satisfied by at least one sequence. The left-to-right side of this is trivial, but what about the right-to-left side? Why is it the case that if a closed sentence is satisfied by at least one sequence, it is satisfied by all sequences? Suppose that there is one sequence X which satisfies "$(\exists x_1)Fx_1$". By (5), this means that there is some sequence X*, differing from X in at most the first place, such that X* satisfies the open sentence "Fx_1". Since X* satisfies "Fx_1", X_1 will be a philosopher (Kant, say, for the sake of argument). Now choose an arbitrary sequence X**: <London, Sydney, Stan Laurel, Alfred Hitchock, . . .>. Does X** satisfy

"$(\exists x_1)Fx_1$"? Is there a sequence, differing from X** in at most the first place, such that that sequence satisfies "Fx_1"? The answer is that there is: simply take the first member of X** and replace it with the first member of X* (Kant): the resulting sequence satisfies "Fx_1", since Kant is a philosopher, so by (5), X** also satisfies "$(\exists x_1)Fx_1$". So, on the assumption that there is one sequence which satisfies the closed sentence "$(\exists x_1)Fx_1$" it follows that an arbitrary sequence satisfies "$(\exists x_1)Fx_1$": in other words, if a closed sentence is satisfied by one sequence, then it is satisfied by all sequences.

Example 1: "$(\forall x_1)Gx_1$"
Intuitively, this sentence is true if and only if everything is Greek. We can derive this truth-condition as follows:

> "$(\forall x_1)(Gx_1)$" is true if and only if it is satisfied by all sequences (from the definition of "true")
>
> "$(\forall x_1)(Gx_1)$" is true if and only if it is satisfied by at least one sequence (from above)
>
> "$(\forall x_1)(Gx_1)$" is satisfied by a sequence X if and only if every sequence X*, differing from X in at most the first place, satisfies "Gx_1" (from (6)).

Every such sequence X* satisfies "Gx_1" if and only if everything is Greek (if everything is Greek, every choice for X^*_1 will be Greek, so that every X* will satisfy "Gx_1"; conversely, if there is something which is not Greek, choosing it as X^*_1 will result in a sequence X* which does not satisfy "Gx_1").

Putting all of these together gives:

> "$(\forall x_1)(Gx_1)$" is true if and only if everything is Greek.

Example 2: "$(\exists x_1)–Fx_1$"
Intuitively, this sentence is true if and only if there is something which is not a philosopher. We can derive this truth-condition as follows:

> "$(\exists x_1)$–Fx_1" is true if and only if it is satisfied by all sequences (from the definition of "true")
>
> "$(\exists x_1)$–Fx_1" is true if and only if it is satisfied by at least one sequence (from above)
>
> "$(\exists x_1)$–Fx_1" is satisfied by a sequence X if and only if there is some sequence X*, differing from X in at most the first place, which satisfies "–Fx_1" (from (5)).
>
> There is such a sequence X* if and only if there is a sequence X* such that X^*_1 does not satisfy "Fx_1" (from (3)).
>
> There is a sequence X* such that X^*_1 does not satisfy "Fx_1" if and only if there is at least one object which is not a philosopher (if there is such a person, say Plato, we can take X^*_1 to be this person; if there is no such person, there will be no appropriate X^*_1).

Putting all these together gives:

> "$(\exists x_1)$–Fx_1" is true if and only if there is at least one thing which is not a philosopher.

Example 3: "$(\exists x_1)(Fx_1 \ \& \ Gx_1)$"
Intuitively, this sentence is true if and only if there is something which is both Greek and a philosopher. We can derive this truth-condition as follows:

> "$(\exists x_1)(Fx_1 \ \& \ Gx_1)$" is true if and only if it is satisfied by all sequences (from the definition of "true")
>
> "$(\exists x_1)(Fx_1 \ \& \ Gx_1)$" is true if and only if it is satisfied by at least one sequence (from above)
>
> "$(\exists x_1)(Fx_1 \ \& \ Gx_1)$" is satisfied by a sequence X if and only if there is some sequence X*, differing from X in at most the first place, which satisfies "$Fx_1 \ \& \ Gx_1$" (from (5)).
>
> There is such a sequence X* if and only if there is a sequence X* such that X^*_1 satisfies both "Fx_1" and "Gx_1" (from (4)).
>
> There is a sequence X* such that X^*_1 satisfies both "Fx_1" and "Gx_1" if and only if there is at least one thing which is both

> Greek and a philosopher (if there is such a person, say Plato, we can take X^*_1 to be this person; if there is no such person, there will be no appropriate X^*_1).

Putting all of these together we get:

> "$(\exists x_1)(Fx_1 \ \& \ Gx_1)$" is true if and only if there is something which is both Greek and a philosopher.

In each case, we have shown how the truth-conditions of a closed complex sentence depend systematically on the semantic properties of the sentence's constituents, even though those constituents are not themselves capable of truth or falsity. This shows that if we can translate the sentences of a fragment of natural language into a language like L* – the language of predicate logic – we can provide a Tarskian truth-theory for that fragment. Given the argument of the previous section, it would follow that we can provide a systematic theory of meaning for that fragment of language. Theories of meaning in this style do not contain intensional constructions, and

> They make no use of meanings as entities; no objects are introduced to correspond to predicates or sentences; and from a finite set of axioms it is possible to prove, for each sentence of the language to be interpreted, a theorem that states the truth-conditions of that sentence.[17]

We now return to Davidson's idea that Tarskian theories of truth can serve as theories of meaning for natural languages.

8.6 Truth and translation: Two problems for Davidson

Platts argues[18] that there is a problem with the six-stage Davidsonian argument rehearsed in §8.4, stemming from Davidson's use of the notion of *translation*. Tarski spelled out his condition of material adequacy in the course of attempting to give

a definition of "true": so there is no problem about his stipulating that in the instances of the (T) schema the sentence replacing "p" should be a translation of the sentence named by "s". However, Davidson, unlike Tarski, cannot help himself to the notion of translation in spelling out his condition of extensional adequacy. Translation means *correct translation*, and correct translation is essentially *meaning-preserving* translation: the reason that "Snow is white" is a correct translation of "Schnee ist weiss" is that the former preserves the *meaning* of the latter. It is legitimate for Tarski to presuppose this notion because his project is that of defining truth: but Davidson's project is that of providing a theory of meaning, so if he wants to avoid a damaging circularity in his theory he will have to avoid using the notion of a meaning-preserving translation in the adequacy conditions on a theory of meaning (deciding whether the adequacy condition was satisfied would itself presuppose that we had an adequate theory of meaning). But if Davidson drops the condition that "p" be replaced by a translation of the sentence named by "s", the six-stage argument appears to falter at stage 6: there will be no guarantee that the predicate "X" will be co-extensive for "is true", and so no reason for thinking that a theory of truth can serve as a theory of meaning. In effect, Davidson is faced with a dilemma. If he uses the notion of correct translation in the condition of extensional adequacy, then his account simply presupposes the notion – meaning – which it is attempting to elucidate; if he does not use the notion of correct translation, the route to forging the connection between theories of meaning and Tarskian truth-theories is blocked.

Another problem which Davidson faces is this. Davidson's theory of meaning issues in theorems of the form

(7) s is true if and only if p.

In (7) the "if and only if" is taken to be the *material biconditional* of propositional logic. Indeed, Davidson takes this to be an advantage of his theory of meaning: it can be framed using the familiar materials of basic, truth-functional logic. But this leaves us with a problem. Recall that the truth-table for the material biconditional is as follows:

P	Q	P ↔ Q
T	T	T
T	F	F
F	T	F
F	F	T

In other words, a material biconditional is true, just in case the two sentences flanking it have the same truth-value. Now, given this, it certainly follows that

(16) "Snow is white" is true if and only if snow is white

is true: "'Snow is white' is true" and "Snow is white" have the same truth-value (true), so the overall biconditional is true also. But so are

(17) "2 + 2 = 4" is true if and only if Auckland is the most populous city in New Zealand
(18) "Fish swim" is true if and only if dogs bark

for the same reason: in each case the sentence on the left-hand side of the biconditional has the same truth-value as the sentence on the right. But surely we would not conclude from this that the sentence "2 + 2 = 4" means *that Auckland is the most populous city in New Zealand* or that the sentence "Fish swim" means *that dogs bark*? Again, Davidson appears to be faced by a dilemma: either he strengthens "if and only if" so that it is stronger than the material biconditional, in which case his aim to give a theory of meaning in a purely extensional language is endangered; or he is committed to totally implausible claims about what certain sentences mean.

How can Davidson respond to these worries? In the next section, we'll see that the answer to this question turns on a constraint on theories of meaning that we have so far paid little attention to: the interpretation condition.

8.7 Radical interpretation and the principle of charity

In fact, Davidson is perfectly aware of the need to say something in response to the two problems we raised in the previous section. The first problem is addressed in the following quote:

> In Tarski's work, T-sentences are taken to be true because the right branch of the biconditional is assumed to be a translation of the sentence truth-conditions for which are being given. But we cannot assume in advance that correct translation can be recognised without pre-empting the point of radical interpretation; in empirical applications, we must abandon the assumption. What I propose is to reverse the direction of explanation: assuming translation, Tarski was able to define truth; the present idea is to take truth as basic and to extract an account of translation or interpretation.[19]

The key here is obviously the third adequacy constraint on theories of meaning, the interpretation condition (C): the theory of meaning must allow us to *correctly interpret* the speakers of that language, in accordance with constitutive principles governing interpretation. Davidson's idea, very roughly, is as follows. We lay down a set of constitutive constraints on the interpretation or translation of the utterances of the speakers of the language for which we are attempting to provide a theory of meaning. These constitutive constraints can be spelled out *without* presupposing the notion of correct translation. If a theory of meaning satisfies these constraints, the sentence on the right-hand side of one of its theorems will effectively provide a translation of the sentence named on its left-hand side. The notion of translation is not *presupposed*; rather, the notion of correct translation is *yielded* by the constitutive constraints on interpretation. The upshot of this is that Davidson's condition of extensional adequacy can be spelled out without using the notion of translation, as follows: for each sentence of the object-language, a theorem of the form

(7) s is true if and only if p

"gives the meaning" of the sentence named by "s" if and only if it is delivered by a theory which satisfies the constitutive constraints on interpretation. Thus, we need to explain Davidson's views on interpretation. What is interpretation and what are the constitutive constraints which govern it?

Davidson is attempting to spell out what is involved in constructing systematic theories of meaning for natural languages. Two distinctive features of natural languages which we have thus far ignored are that (a) they are spoken, and (b) they contain indexical expressions, such as "I", "here", and "now". The truth-values of sentences containing indexical expressions change according to context: for example, "It is raining here now", as uttered by AM in Canberra on 19 May 2017 would be true, whereas the same sentence, uttered by AM in Glasgow on 23 June 2016 would be false. Davidson's way of coping with this, in a way which takes (a) into account as well, is to view truth as a relation between a sentence, a speaker, and a time. The theorems of a theory of meaning will thus have the following form:

(19) "It is raining here now" as uttered by AM at time t is true if and only if it is raining in the vicinity of AM at time t.

If the theory of meaning which issues in (19) is correct, it will allow us to *interpret* AM's utterance of the sentence "It is raining here now" as an act of saying *that it is raining in the vicinity of AM at time t*:

> We interpret a bit of linguistic behaviour when we say what a speaker's words mean on an occasion of use. The task may be seen as one of redescription. We know that the words "Es schneit" have been uttered on a particular occasion and we want to redescribe this uttering as an act of saying that it is snowing.[20]

Call a theory of meaning which allows us to do this correctly "interpretative". Now the problem facing us can be stated as follows: how can we say what it is for a theory of meaning to be "interpretative", other than saying that it issues in correct translations or interpretations of speakers' utterances? *Davidson's answer to this*

is that a theory of meaning will be interpretative if it satisfies a number of constitutive constraints on interpretation, principally, the principle of charity.

Suppose that Kurt utters the sentence "Es regnet" and that we wish to interpret his utterance: we wish to describe his uttering that sentence as an act of saying that-such-and-such. What sort of evidence are we allowed to go on in attempting to construct such an interpretation? The evidence allowed by Davidson concerns what Kurt *holds true*: the idea is that we use

(20) Kurt holds-true "Es regnet" if and only if it is raining in Kurt's vicinity

as evidence for

(21) "Es regnet", as uttered by Kurt, is true if and only if it is raining in Kurt's vicinity.

Two points should be made about the selection of facts concerning what speakers hold-true as evidence for interpretation. First, such evidence is semantic, since holding a sentence true is believing that the sentence is true. Since the evidence allowed is semantic, Davidson is *not* attempting to give an account of how interpretation can proceed on a purely non-semantic basis (in this respect he differs from Quine; see §4.7). Second, although the evidence is semantic, it is still *thinner* than the notion of an utterance's being an act of saying that such and such: one can know that a speaker holds a sentence true without knowing what that sentence means. So even though we start from semantic evidence, we can still make progress in our attempt to understand other, richer, semantic notions:

> A good place to begin is with the attitude of holding a sentence true, of accepting it as true. This is, of course, a belief, but it is a single attitude applicable to all sentences, and so does not ask us to make finely discriminated distinctions among beliefs. It is an attitude an interpreter may plausibly be taken to be able to identify before he can interpret, since he may know that a person intends to express a truth in uttering a sentence without having any idea *what* truth.[21]

Davidson's plan is to impose constraints on the selection of evidence in the form of (20) such that (21) will allow us to interpret Kurt's utterance of "Es regnet" as an act of saying *that it is raining*.

We need such constraints because on its own (20) does *not* provide evidence for (21). This is because of the interdependence of belief and meaning:

> A speaker who holds a sentence to be true on an occasion does so in part because of what he means, or would mean, by an utterance of that sentence, and in part because of what he believes. If all we have to go on is the fact of honest utterance, we cannot infer the belief without knowing the meaning, and have no chance of inferring the meaning without the belief.[22]

Suppose that it is a rainy afternoon in Berlin. Suppose also that, for some strange reason, Kurt does *not* believe that it is raining (perhaps he is under the influence of a drug, or suffering from some other sort of perceptual malfunction). Then we cannot take the fact that Kurt holds the sentence "Es regnet" true as evidence for (20). Likewise, suppose that "Es regnet" actually means something different, say, it is Tuesday. Then we could not take the fact that Kurt holds this true as evidence for the claim that he believes that it is raining. In order to move from a claim about what sentences Kurt holds true to a claim about what he believes to be the case, we need to know the meanings of those sentences; and in order to move from a claim about what sentences Kurt holds true to a claim about what he means, we need to know what he believes to be the case. We are thus faced with a problem: we want to move from evidence concerning what Kurt holds-true to a claim about what he means; in order to do this we need to know what he believes; but in order to get to what he believes from what he holds-true, we need to know what he means. How can we break out of this circle?

Davidson's answer is that we break out of the circle via an application of the *principle of charity*: we assign to Kurt the beliefs that *we* (the interpreters) would have in the circumstances in question, and then go on to move from these beliefs, together with the facts about what sentences Kurt holds-true, to conclusions concerning

what those sentences mean. In other words, we break out of the circle by assuming that in the circumstances in question Kurt has the beliefs which, by our lights, are true:

> [We] solve the problem of the interdependence of belief and meaning by holding belief constant as far as possible while solving for meaning. This is accomplished by assigning truth-conditions to [native] sentences that make native speakers right when plausibly possible, according, of course, to our own view of what is right.[23]

To take the simple example above. If we were to find ourselves in Berlin in the middle of a downpour, among the beliefs we would have would be the belief that it is raining. So, by the principle of charity, we assume that Kurt has this belief. We can then move from (20) to (21).

Let's sum up the story so far. In order to avoid the problem about presupposing the notion of correct translation, we impose the interpretation condition as a condition of adequacy on theories of meaning: the theory must allow us to *correctly interpret* the speakers of that language. In order to correctly interpret the speakers of the language, we must take as evidence facts about which sentences they hold true. But this evidence only counts as evidence if we make assumptions about what they believe. The principle of charity licenses such assumptions: we ascribe to the natives, by and large, the beliefs that we would deem to be appropriate in the relevant circumstances. Thus, for extensional adequacy, we do not require that in theorems of the form of

(7) s is true if and only if p

the sentence replacing "p" translates the sentence named by "s". The work done by this circular specification of extensional adequacy is now done by the non-circular requirement that the theorem be generated by a theory which satisfies the constitutive constraints on interpretation, principally, the principle of charity. This is how Davidson proposes to take truth as basic and to extract an account of translation and meaning: we can after all forge a connection between theories of meaning and Tarskian truth-theories without begging the question by invoking a notion

of correct translation. The second horn of the dilemma posed for Davidson in §8.6 is thus neutralised.[24]

At this stage, a few clarificatory notes are in order:

Note 1: It might seem as if the principle of charity used above leaves it impossible to interpret a speaker as ever having a belief which is, by our own lights, false. But this was only a very crude version of the principle. Properly construed, the principle of charity is a *holistic* constraint applying, not to individual beliefs, but rather to *systems* of belief: we must interpret a speaker so that *most* of the beliefs in his system are, by our lights, true. This leaves room for us to ascribe *some* false beliefs to him. In fact, there is an even more sophisticated version of the principle according to which we interpret speakers not as necessarily having beliefs which are true by our own lights, but as having beliefs which are *intelligible* by our own lights. This allows us to attribute false beliefs to a speaker, so long as the source of his error is, by our own lights, *explicable* or *intelligible*. Davidson sums this up in the slogan "The aim of interpretation is not agreement but understanding".[25]

Note 2: According to Davidson, the principle of charity is not merely a useful heuristic principle which we can rely on to facilitate the practical process of interpretation. The adoption of the principle is rather *constitutive* of the whole process of interpretation:

> The methodological advice to interpret in a way that optimizes agreement [or minimizes inexplicable disagreement, in the more sophisticated version] should not be conceived as resting on a charitable assumption about human intelligence that might turn out to be false. If we cannot find a way to interpret the utterances and other behaviour of a creature as revealing a set of beliefs largely consistent and true by our own standards [or not inexplicably inconsistent and wrong by our own standards], we have no reason to count that creature as rational, as having beliefs, or as saying anything.[26]

Note 3: The necessity of applying the principle of charity stems from the constitutive interdependence of belief and meaning. This shows where Davidson stands on the question raised in §6.1 concerning the relationship between linguistic meaning and mental

content. Davidson holds a "no-priority" view: linguistic meaning – the meaning of sentences – and mental content – the content of propositional attitudes such as belief – must be explained *together*, or not at all.

Note 4: Although we have written as if there was only one constitutive constraint governing interpretation, the principle of charity, this principle is only one of a *range* of such constitutive principles. As Hookway puts it

> Interpretation rests upon a number of normative standards. We are constrained to look for true beliefs, to look for rationally coherent bodies of belief, to avoid ascribing inexplicable ignorance, to look for reasonable desires, to look for coherent patterns of preferences, and so on.[27]

Note 5: Davidson writes that it is unlikely that the constraints on theories of meaning – extensional adequacy, compositionality, and interpretation – will constrain the choice of a correct theory of meaning to within uniqueness. There will probably be different theories of meaning satisfying all of the relevant constraints:

> It is not likely that only one theory will be found satisfactory. The resulting indeterminacy of interpretation is the semantic counterpart of Quine's indeterminacy of translation. On my approach, the degree of indeterminacy will, I think, be less than Quine contemplates . . . But in any case the question of indeterminacy is not central [to my concerns]. Indeterminacy of meaning or translation does not represent a failure to capture significant distinctions; it marks the fact that certain apparent distinctions are not significant. If there is indeterminacy, it is because when all the evidence is in, alternative ways of stating the facts remain open.[28]

Davidson writes that the amount of indeterminacy left open will be less than that envisaged by Quine because his use of the principle of charity is more wide-ranging than Quine's (who takes it to apply only to the translation of the logical constants) and because "the uniqueness of quantificational structure is apparently ensured if Convention T is satisfied".[29] Presumably, the fact that Davidson allows a richer sort of evidence than Quine to count in the process

of interpretation will also contribute to the reduction of indeterminacy. But what is the source of Davidson's confidence that any residual indeterminacy will be "insignificant"? He writes:

> A rough comparison may help give the idea. A theory of measurement for temperature leads to the assignment to objects of numbers that measure their temperature. Such theories put formal constraints on the assignments, and also must be tied empirically to qualitatively observable phenomena. The numbers assigned are not uniquely determined by the constraints. But the *pattern* of assignments is significant. (Fahrenheit and Centigrade temperature are linear transformations of each other; the assignment of numbers is unique up to a linear transformation.) In much the same way, I suggest that what is invariant as between different acceptable theories of truth is meaning.[30]

And what is invariant between different acceptable theories of truth is the role they assign to the sentences of the language under consideration:[31]

> The meaning (interpretation) of a sentence is given by assigning the sentence a semantic location in the pattern of sentences that comprise the language. Different theories of truth may assign different truth conditions to the same sentence . . . while the theories are (nearly enough) in agreement on the roles of the sentences of the language.[32]

Whereas in Quine's case that which is invariant between different translation manuals – stimulus meaning – does not amount to anything even resembling our intuitive concept of meaning, in Davidson's account that which is invariant between different acceptable theories of truth is conceived as approximating to our intuitive conception of meaning.[33]

8.8 Holism and T-theorems

Thus far we have been considering only Davidson's response to the first of the worries raised in §8.6. But what about the second worry

that was raised? How can Davidson rule out "rogue" T-theorems such as

(18) "Fish swim" is true if and only if dogs bark

from the theory of meaning? Recall that the dilemma Davidson faced was as follows: either he strengthens "if and only if" as it appears in the T-theorems so that it is *stronger* than the material biconditional, in which case his aim to give a theory of meaning in a purely extensional language is endangered; or he is committed to totally implausible claims about what certain sentences mean (e.g. that the sentence "Fish swim" means *that dogs bark*).

Davidson's strategy in responding to this dilemma is similar to that discussed in the previous section. He attempts to neutralise the second horn of the dilemma by showing that we can stick to an extensional theory of meaning, whose theorems are couched in terms of only material equivalencies, but invoke the holistic constraints on interpretation to rule out theories of meaning containing "rogue" T-theorems.

Let's go back to our rainy afternoon in Berlin, and Kurt's utterance of "Es regnet". Suppose that there is also an aeroplane passing overhead at the time of the utterance. Then the following material biconditional will be true, since both its left- and right-hand sides have the same truth value:

(22) "Es regnet", as uttered by Kurt, is true if and only if there is an aeroplane passing overhead.

Now given that this is true, how can we rule it out from the theory of meaning we are attempting to construct for Kurt's language? Can we say that although it is true, it nevertheless corresponds to a false statement about what Kurt holds true? The relevant statement would be

(23) "Es regnet" is held-true by Kurt if and only if there is an aeroplane passing overhead.

But this sentence (call it an H-sentence) is actually true as well: at the time of utterance it is true that Kurt holds-true the sentence

"Es regnet" and it is also true that there is an aeroplane passing overhead.

What this shows is that what supports the claim that (21) gives the meaning of "Es regnet" is not the truth of the individual H-sentence (20): *rather, it is the fact that (21) is a theorem of a theory of meaning none of whose T-theorems corresponds to false H-sentences*. The idea is then that we can say that although (22) is true, it does not give the meaning of "Es regnet", because a theory of meaning containing it would at some point generate a theorem corresponding to a *false* H-sentence. The constraints on interpretation are holistic in the sense that they apply to the theory as a whole, not to individual theorems considered in isolation. As Hookway puts it

> The empirical test supports the theory as a whole; we accept the statement about what "Es regnet" means because the theorem follows from an empirically adequate theory of meaning. Statements of truth-conditions cannot be verified singly: utterances and actions always reflect the influence of many of an agent's beliefs and desires – there are no straightforward links between belief and behaviour.[34]

So why is it likely that a theory of meaning containing (22) would at some point generate a theorem corresponding to a *false* H-sentence? Well, given that "regen" is being treated as a subsentential expression which coincides in extension with "aeroplane", the theory is likely to include a clause like the following in its axiomatic base:

(24) (∀x)(x satisfies "Regen" if and only if x is an aeroplane)

Now such a theory is also likely to generate a T-theorem like

(25) "Das ist Regen", as uttered by Kurt, is true if and only if the demonstrated object is an aeroplane.

But what H-sentence corresponds to this? It will be

(26) Kurt holds-true "Das ist Regen" if and only if the demonstrated object is an aeroplane.

And this is plausibly false. It is easy to find circumstances in which Kurt holds-true "Das ist Regen" but in which the demonstrated object is not an aeroplane, and circumstances in which the demonstrated object is an aeroplane but in which Kurt does not hold-true "Das ist Regen". The left-hand side of (26) is false, while the right-hand side is true: so by the truth-table for the material biconditional, (26) itself is false. Since the theory of meaning containing (22) in this way leads to a T-theorem which corresponds to a false H-sentence, we are justified in rejecting this theory of meaning, even though (22) itself is true. This contrasts with the theory containing (20), which, because it does not contain (24), has no such implication.

This shows how the holistic constraints on interpretation, and the holistic process of confirming a theory of meaning, allow us to rule out theories of meaning containing "rogue" T-theorems without requiring anything more than material equivalence in our statements of truth-conditions: the second horn of the dilemma is thus neutralised.

8.9 Conclusion: Theories of meaning and natural language

As we have seen, Davidson thinks that a theory of meaning for a natural language can be provided by a theory of truth in the Tarskian style which satisfies the conditions of extensional adequacy, compositionality, and interpretation. Given a fragment of natural language, what we require in order for a theory of meaning of this sort to be constructed is that the sentences of this language be couched in a form which renders them amenable to Tarskian treatment. In effect, what this requires is that the sentences be translatable into the language of Frege's symbolic logic. Now many of the sentences of natural language do easily translate into this formal notation: any student who has taken an elementary logic course at university will be familiar with this process (see also §1.1). But there are many regions of natural language which resist straightforward translation into Frege's logical language. Both Frege and Tarski are pessimistic about providing systematic semantic theories or theories of truth for natural languages,

and limit themselves to constructing theories for formal languages and the language of mathematics: according to them, natural languages are simply too amorphous and paradox-ridden for this to be a viable project. Davidson, however, is more optimistic about the prospect of constructing a systematic theory of meaning for a natural language. He ends one of his most important papers, "Truth and Meaning", with the following comment:

> Since I think there is no alternative, I have taken an optimistic and programmatic view of the possibilities for a formal characterisation of a truth-predicate for a natural language. But it must be allowed that a staggering list of difficulties and conundrums remains. To name a few: we do not know the logical form of counterfactual or subjunctive sentences; nor of sentences about probabilities and about causal relations; we have no good idea what the logical role of adverbs is, nor the role of attributive adjectives; we have no theory of mass terms like "fire", "water", "snow", nor for sentences about belief, perception, and intention, nor for verbs of action that imply purpose. And finally, there are all the sentences that seem not to have truth-values at all: the imperatives, optatives, interrogatives, and a host more. A comprehensive theory of meaning for a natural language must cope successfully with each of these problems.[35]

Since the publication of "Truth and Meaning", philosophers of language – including Davidson himself – have expended much effort in attempting to show how the various locutions he mentions in this passage can be brought within the compass of a systematic theory of meaning. We have not had space to look at any of the details of these attempts in this chapter, but hopefully we have given the reader some sense of what it is that these philosophers of language have been aiming at.[36]

Chapter summary

This chapter outlined Donald Davidson's idea that a Tarskian theory of truth can serve as a theory of meaning for a natural language. We saw that Davidson claimed that Frege had no good

story about how word-meaning determines sentence-meaning and that he also rejected theories of meaning that treat meanings as entities or that otherwise help themselves to intensional notions. We sketched how a Tarskian theory of truth would work in the case of a simple fragment of language that contained universal and existential quantifiers, and in order to see how such a theory might be used to interpret the speech of a group of language users, we explained how Davidson's principle of charity functions as a holistic constraint in the process of radical interpretation.

Study questions

1. How convincing is Davidson's claim that Frege's account of how sentence-meaning is a function of word-meaning is completely unexplanatory?
2. What is wrong with a theory of meaning whose theorems are of the form "s means that p"?
3. Show that a (closed) universally quantified sentence is satisfied by at least one sequence if and only if it is satisfied by all sequences.
4. Show how the truth-condition of "$(\forall x_1)(Fx_1 \ \& \ Gx_1)$" is a function of the semantic properties of its parts.
5. Show how the truth-conditions of the closed complex sentences in examples 1–3 in §8.5 are dependent on the semantic properties of their constituents, using the Fregean theory sketched in §1.7.
6. Given that the biconditional "'Snow is white' is true if and only if diamonds are hard" is true, is Davidson committed to the implausible claim that "Snow is white" means that diamonds are hard?
7. Does the principle of charity imply that we cannot interpret a speaker as having a false belief?
8. Explain why a plausible theory of meaning is unlikely to contain (17) and (18) from §8.6.

Further reading

Davidson's main papers on truth, meaning, and radical interpretation are collected in his *Inquiries into Truth and Interpretation*.

The best overall introductions to his philosophy are Gluer, *Donald Davidson: A Short Introduction*; Evnine, *Donald Davidson*; and Joseph, *Donald Davidson*. A more substantial introduction is Lepore and Ludwig, *Donald Davidson: Meaning, Truth, Language and Reality*. These are especially good on situating Davidson's philosophy of language in the more general context which includes his philosophy of mind and metaphysics. See also Ramberg, *Donald Davidson's Philosophy of Language*; Platts, *Ways of Meaning*; Blackburn, *Spreading the Word*, Chapter 8; and Hookway, *Quine*, Chapter 10. For a useful collection of papers, see Lepore (ed.), *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*. Davidson's philosophy of language cannot be studied adequately without a knowledge of his philosophy of mind: for this, see Davidson's collected papers, *Essays on Actions and Events*, and the collections Lepore and McLaughlin (eds), *Actions and Events: Perspectives on the Philosophy of Donald Davidson*; and Hintikka and Vermazen (eds), *Essays on Davidson: Actions and Events*. Tarski's results on truth are set out in his "The Concept of Truth in Formalized Languages". This is a difficult and highly technical paper, and Tarski's work is probably best approached via the more informal "The Semantic Conception of Truth" and "Truth and Proof". There is a good – but terse – presentation of Tarski's theory of truth in Platts, *Ways of Meaning*, Chapter 1; see also the presentations in Quine, *Philosophy of Logic*, and Haack, *Philosophy of Logics*. For a comprehensive discussion of interpretations of Tarski, see Milne, "Tarski on Truth and its Definition". For high-level, article-length overviews, see Wiggins, "Meaning and Truth Conditions: From Frege's Grand Design to Davidson's"; Higginbotham, "Truth and Reference as the Basis of Meaning"; and Segal, "Truth and Meaning". See also Hahn (ed.), *The Philosophy of Donald Davidson*; Ludwig (ed.), *Donald Davidson*; Zeglen (ed.), *Donald Davidson: Truth, Meaning and Knowledge*; and Lepore and Ludwig (eds.), *A Companion to Donald Davidson*.

Two contributions by Davidson to the philosophy of language not mentioned above are his attack on the idea that the notion of convention has a substantive role to play in the theory of meaning and his account of first-person knowledge. For the former, see his "Communication and Convention" (in his *Inquiries*) and his "A Nice Derangement of Epitaphs" (in his *Truth, Language and History*).

For some commentary see Miller, "Postscript to Jane Heal's 'Radical Interpretation'". For the latter, see Davidson's "First Person Authority" (in his *Subjective, Objective, Intersubjective*). For a critique see Wright, "The Problem of Self-Knowledge (II)".

An issue which has only recently begun to be discussed in the literature is the nature of Davidson's view of KW's sceptical paradox about rule-following; see Verheggen, "Davidson's Treatment of Wittgenstein's Rule-Following Paradox". For an extended discussion of Davidson's notion of "triangulation" and its role in securing determinacy of meaning, see Myers and Verheggen, *Donald Davidson's Triangulation Argument: A Philosophical Inquiry*.

Notes

1 *Grundgesetze*, I.32 (Beaney, *The Frege Reader*, p.221).
2 Davidson, *Inquiries into Truth and Interpretation*, p.24.
3 Ibid., p.109.
4 Ibid., p.20.
5 Ibid., p.17.
6 See ibid., pp.8–9, 18, 56.
7 This is the argument used against Quine in §4.3: a theory of meaning which has an infinite number of semantic primitives, and which is therefore not finitely axiomatisable, will be in principle unlearnable. There are some worries – which we cannot go into here – about whether Davidson really can answer (a), (b), and (c) by imposing the compositionality constraint on theories of meaning. For a discussion, see Miller, "Tacit Knowledge". (a), (b), and (c) have also figured prominently in the work of Chomsky: see Collins, *Chomsky* for an overview.
8 In fact, there are some further adequacy conditions which Davidson imposes, but we'll simply ignore these in what follows.
9 Davidson, *Inquiries*, pp.20–21.
10 Platts, *Ways of Meaning*, p.53.
11 Ibid.
12 Davidson, *Inquiries*, pp.22–23.
13 Tarski, "Semantic Conception of Truth", p.344.
14 Tarski also imposes some conditions of "formal adequacy", designed to help avoid the semantic paradoxes. We won't consider these further conditions here. For an account of them, see the papers by Tarski listed under Further reading.

15 In actual fact, we would need to complicate matters a little here. As they stand, these clauses include bound variables within quotation marks: in order to avoid this, we should really use the spelling function and e.g. write A^"&"^B instead of "A & B" (see Platts, *Ways of Meaning*, Chapter 1).
16 For the generalisation to the case of languages containing predicates of more than one place, see ibid.
17 Davidson, *Inquiries*, p.xiv.
18 Platts, *Ways of Meaning*, pp.56–57.
19 Davidson, *Inquiries*, p.134.
20 Ibid., p.141.
21 Ibid., p.135.
22 Ibid., p.142.
23 Ibid., p.137.
24 Strangely, Platts writes (*Ways of Meaning*, pp.58–67) as if Davidson is unaware of this problem, and suggests that Davidson will have to modify his theory along the lines suggested in McDowell, "Truth-Conditions, Bivalence, and Verificationism" in order to deal with it. But as I hope I have shown, Davidson is actually acutely aware of the problem, and his solution to it takes us to the heart of his views on interpretation.
25 Davidson, *Inquiries*, p.xvii.
26 Ibid., p.137.
27 Hookway, *Quine*, p.173.
28 Davidson, *Inquiries*, pp.153–154.
29 Ibid., p.153.
30 Ibid., p.125.
31 This is connected with Davidson's view that reference, as applied to subsentential expressions, has a purely instrumental role to play within the theory of meaning. See "Reality Without Reference", in ibid. See also Field, "Tarski's Theory of Truth", and McDowell, "Physicalism and Primitive Denotation" for an important exchange on this and related issues.
32 Davidson, *Inquiries*, p.125.
33 For a critique of Davidson's use of the measurement-scales analogy, see Hossein-Khani, Kripke's Wittgenstein's Sceptical Solution and Donald Davidson's Philosophy of Language, Chapter 3.
34 Hookway, *Quine*, p.174.
35 Davidson, *Inquiries*, p.136.
36 For examples of applications of the Davidsonian approach, see Lepore and Ludwig, *Donald Davidson's Truth-Theoretic Semantics*.

Chapter 9

Sense, world, and metaphysics

In this chapter, we return to an issue which loomed large in Chapter 3: the relationship between the philosophy of language and metaphysics. Philosophers no longer believe the positivist idea that philosophy of language can enable us to dispense with metaphysical debate: instead, the philosophy of language has come to be viewed as a *tool* we might use in attempting to get clearer on what metaphysical questions are and on how they might be answered. In particular, philosophy of language has come to be regarded by some leading thinkers as central in the metaphysical debates between *realists* and their opponents. The literature on this topic is vast and complex: in this chapter our aim is simply to provide a rough map of some of the main territory. Inter alia, we will also introduce some issues about the relationship between language, mind, and world: in particular, about the relationship between a speaker's grasp of sense and facts about the nature of his worldly environment.

9.1 Realism

What is it to be a realist about the subject matter of a particular discourse or area of inquiry? Take discourse about ordinary, macroscopic objects such as tables, chairs, sticks, and stones; or discourse about what is morally right and wrong, about ethical value. *To be a realist about the subject matters of these discourses is to think that our talk and thought about them aspires to reflect an objective reality, and sometimes succeeds in this aim*. Thus, when I judge that there are two chairs in the room, I express a *belief* which aims to reflect an objective fact; and beliefs such as this often do *succeed* in reflecting the objective facts of the matter, so that the judgement in question is *true*. Likewise in the case of moral discourse: when I judge that torture is wrong, I am expressing a belief which aims to reflect an objective fact and which is successful in this aim.

But can we say something more specific about what it is for our thought "to aspire to reflect an objective reality"? One way to start on this would be as follows: to be a realist about the subject matter of, for example, discourse about ordinary, macroscopic objects is to think that the sentences of this discourse *possess objective truth-conditions*. Likewise, to be a realist about morals is to think that the sentences of moral discourse possess objective truth-conditions. And to think that we sometimes succeed in this aim of getting our thought to reflect an objective reality is to think that these objective truth-conditions are sometimes satisfied.

Now we have the task of spelling out what is meant by the term "objective" in this characterisation of realism: what does it mean to say that the sentences of a discourse have objective – as opposed to, say, subjective – truth-conditions? This task will occupy us, in one way or another, in much of this chapter. We'll work towards answering it by first considering some ways of *rejecting* realism about particular areas of inquiry.[1]

9.2 Non-cognitivism and the Frege–Geach problem

One way of denying that the sentences of a given discourse have objective truth-conditions would be to deny that they have

truth-conditions *at all*. Say that if a sentence has truth-conditions then it is *truth-apt*: apt to be true or false, or apt to be assessed in terms of truth and falsity. One way of denying that a given set of sentences have objective truth-conditions would be to deny that they are truth-apt. In fact, we have already seen an example of this sort of view in §3.6, when we looked at the logical positivist account of ethical language. According to Ayer, a sentence such as "Stealing is wrong" is not the sort of sentence that can be assessed in terms of truth and falsity: in uttering it we are not even purporting to represent some fact about the world, some condition in the world whose obtaining would render the utterance true. The function of "Stealing is wrong" is not to express beliefs which aim to represent the world and which can be true or false, but rather to express feelings, emotions, or, generally, *non-cognitive sentiments* which are incapable of being true or false.

One way of opposing realism about morals, therefore, would be to deny that moral sentences are truth-apt: they do not express beliefs, which can be true or false, but rather non-cognitive sentiments which can neither be true nor false. This way of opposing realism is known as *non-cognitivism*.

If non-cognitivism were plausible, it would solve many worries philosophers have had about the metaphysics of morality. Since the *purpose* of moral discourse is not to state facts, we do not have to include moral facts in our inventory of the sorts of fact the world contains in order to legitimise our moral practices.[2] Given that many philosophers find moral facts hard to countenance – "queer metaphysical entities" – they would say that this is just as well.

Unfortunately, however, non-cognitivism faces very serious difficulties. The main problem faced by non-cognitivism is known as the "Frege–Geach Problem". According to non-cognitivism, when I judge that murder is wrong, I am not expressing a belief, but rather expressing some non-cognitive attitude, incapable of being true or false. This is an account of the semantic function of "Murder is wrong" in a context in which it is apparently being asserted that murder is wrong (call this an "asserted context"). But the sentence can occur in many, many "unasserted contexts" in which it is clearly *not being asserted*: for example, if I say that *if murder is wrong, then getting your little brother to murder people*

is wrong, I am not necessarily expressing disapproval of murder. (If I assert a conditional generally, I need not commit myself to asserting its antecedent: for example, if I say that if it is raining, the streets will be wet, I am not thereby asserting that it is actually raining.) The task for the non-cognitivist is therefore this: give an account of the semantic function of the occurrence of "Murder is wrong" in the antecedent of the above conditional in terms of the sentiment it allegedly expresses in asserted contexts. Since I am not expressing disapproval of murder in uttering the conditional, the semantic function of "murder is wrong" as it appears in the antecedent must be different from its semantic function in the asserted context. But now we are going to have a problem in accounting for the following apparently valid inference ("moral modus ponens"):

(1) Murder is wrong.
(2) If murder is wrong, then getting your little brother to murder people is wrong.

Therefore:

(3) Getting your little brother to murder people is wrong.

If the semantic function of "Murder is wrong" as it appears in (1) is different from its semantic function as it appears in (2), isn't someone arguing in this way simply guilty of *equivocation*? In order for the argument to be valid, the occurrence of "murder is wrong" in (1) has to *mean the same thing as* the occurrence of "murder is wrong" in (2). But if "murder is wrong" has a different semantic function in (1) and (2), then it certainly doesn't mean the same thing in (1) and (2). So the above argument is apparently no more valid than:

(1*) My beer has a head on it.
(2*) If something has a head on it, then it must have eyes and ears.

Therefore:

(3*) My beer must have eyes and ears.

This argument is obviously invalid, because it relies on an equivocation on two senses of "head", in (1*) and (2*) respectively.[3]

Another way of putting essentially the same point: in elementary logic, how would we try to figure out how a given argument is valid? One way would be to construct a truth-table (see Chapter 1), and check whether there are any cases in which all of the premises are true and yet the conclusion is false. If there are, the argument is invalid; if there are not, the argument is valid. But how can this procedure even make sense when some of the premises (e.g. (1)) in the argument are not even assessable in terms of truth and falsity?

The Frege–Geach challenge to the non-cognitivist is thus to answer the following question: How can you give a non-cognitivist account of the occurrence of moral sentences in "unasserted contexts" – such as the antecedents of conditionals – without jeopardising the intuitively valid patterns of inference in which those judgements figure?

Simple as this problem seems, non-cognitivists have found it extremely difficult to answer.[4] This suggests that opposition to realism about a particular declarative discourse cannot take the form of a simple denial that the sentences of that discourse are truth-apt or possess truth-conditions.

9.3 Realism and verification-transcendent truth

Denying that sentences have truth-conditions does not therefore give us a plausible way of denying realism, the claim that they have objective truth-conditions. This takes us back to the question about objectivity that we raised at the end of §9.1. Maybe opposition to realism can be framed as the claim that sentences have truth-conditions (so that the Frege–Geach problem is avoided), but that these truth-conditions are not objective. But what does this mean? What does it mean to say that the truth-conditions of a range of sentences *are* "objective"? One way of fleshing this out has been suggested in a series of influential works by Michael Dummett: to say that a range of sentences have objective truth-conditions is to say that those truth-conditions are potentially *verification-transcendent*. To say that a truth-condition is potentially verification-transcendent is to say that it determinately obtains

(or not) even though we may be incapable, even in principle, of determining whether or not it obtains. Thus, consider discourse about the past, and in particular the sentence "James II[5] suffered a migraine in 1665, on the afternoon of his 32nd birthday". Intuitively, we are inclined to say that the truth-condition of this sentence determinately obtains or fails to obtain even though we have no means at our disposal guaranteed to turn up evidence one way or the other. Since all of the evidence may have vanished and time-travel is impossible, we are prepared to say that the truth-condition determinately obtains (or not) even though we may be incapable, even in principle, of verifying that it does (or doesn't). In other words, its truth-condition is potentially verification-transcendent.[6] Or consider arithmetic and take Goldbach's conjecture, "Every even number greater than two is the sum of two primes". A realist would say that this is determinately true or false, even though no-one has ever found a proof or a counterexample, and have we have no means at our disposal of producing one or the other: its truth-condition determinately obtains (or not) even though we may be incapable, even in principle, of proving that it does or finding a counterexample to show that it doesn't. Its truth-condition is, like that of the sentence about James II considered above, potentially verification-transcendent.[7]

Thus, sentences about the past and about arithmetic have potentially verification-transcendent truth-conditions: in this sense their truth-conditions are objective. Now, why is the claim that the sentences of a discourse are potentially verification-transcendent a way of cashing out realism about the subject matter of that discourse? In other words, why is it a way of cashing out the idea that our thought about that subject matter aspires to reflect an objective reality? In order to see this we have to recall that the sense of a sentence is given by its truth-conditions, and that understanding a sentence consists in grasping its sense. Thus, understanding a sentence consists in grasping its truth-conditions. Any thesis about the truth-conditions of a set of sentences is inter alia a thesis about what our understanding of those sentences consists in. In a slogan, a *theory of meaning is also a theory of understanding*. Thus, someone who accepts that the truth-conditions of a region of discourse are potentially verification-transcendent also accepts that our understanding of the sentences of that discourse consists in our

grasp of potentially verification-transcendent truth-conditions. And now the connection with realism about that discourse is relatively easy to see. As Crispin Wright puts it:

> To conceive that our understanding of statements in a certain discourse is fixed . . . by assigning them conditions of potentially [verification]-transcendent truth is to grant that, if the world co-operates, the truth or falsity of any such statement may be settled beyond our ken. So we are forced to recognise a distinction between the kind of state of affairs which makes such a statement acceptable, in the light of whatever standards inform our practice of the discourse to which it belongs, and what makes it actually true. The truth of such a statement is bestowed on it independently of any standard we do or can apply; acceptability by our standards is, for such statements, at best merely congruent with truth. Realism in Dummett's sense is thus one way of laying the essential semantic groundwork for the idea that our thought aspires to reflect a reality whose character is entirely independent of us and our cognitive operations.[8]

We can thus conceive of the metaphysical debate between realists and their opponents – anti-realists – in a particular region of discourse D as concerning whether the sentences of D can plausibly be viewed as possessing potentially verification-transcendent truth-conditions.

Realism: The sentences of D have truth-conditions (are truth-apt) and these truth-conditions are potentially verification-transcendent.

Anti-realism: The sentences of D have truth-conditions (are truth-apt) but those truth-conditions are not potentially verification-transcendent.[9]

Realism, as thus conceived, appears highly intuitive, and is apparently hinted at by some of the most influential philosophers of language that we have considered in this book. Davidson, for example, writes that

> We can be realists in all departments. We can accept objective truth-conditions as the key to meaning, and we can insist that knowledge is of an objective world independent of our thought and language.[10]

And Frege writes:

> A thinker does not create [thoughts] but must take them as they are. They can be true without being grasped by a thinker.[11]

If a thought – the sense of a sentence – can be determinately true or false even though that thought is not even grasped by a thinker, then the sentence in question can be true or false even though thinkers may be incapable, even in principle, of determining its truth-value.

How, then, can we resolve the metaphysical debate between realism and anti-realism about a particular region of discourse? In the next section we look at some of the main arguments against realism.

9.4 Acquisition, manifestation, and rule-following: The arguments against verification-transcendent truth

The debate between realism and anti-realism about a region of discourse is a debate about the nature of the truth-conditions possessed by the sentences of that discourse. Any account of the truth-conditions of a range of sentences will be unacceptable if it cannot cohere with a plausible account of what our *understanding* of those sentences consists in. Dummett's strategy is thus to argue that the account of linguistic understanding which realism leads to is completely implausible, so that realism must be rejected. The metaphysical debate concerning the plausibility of realism thus boils down to a debate within the philosophy of language.[12]

Why, then, does the anti-realist think that we cannot plausibly construe linguistic understanding as grasp of potentially verification-transcendent truth-conditions? There are three main arguments.

9.4.1 The acquisition challenge

Suppose that we are considering some region of discourse D, the sentences of which we intuitively understand. Suppose, for reductio, that the sentences of D have potentially verification-transcendent truth-conditions. Thus,

(1) We understand the sentences of D.
(2) The sentences of D have verification-transcendent truth-conditions.

Now, from (1) together with thesis 10 (Chapter 2) we have

(3) We grasp the senses of the sentences of D: i.e. we know their truth-conditions.

We now add the apparently reasonable constraint on ascriptions of knowledge:

(4) If a piece of knowledge is ascribed to a speaker, then it must be at least in principle possible for that speaker to have *acquired* that knowledge.

So

(5) It must be at least in principle possible for us to have acquired knowledge of the verification-transcendent truth-conditions of D.

But

(6) There is no plausible story to be told about how we could have acquired knowledge of verification-transcendent truth-conditions.

So, by reductio, we reject (2) to get:

(7) The sentences of D do not have verification-transcendent truth-conditions, so realism about the subject matter of D must be rejected.

The crucial premise here is obviously (6). Wright puts the point as follows:

> How are we supposed to be able to *form* any understanding of what it is for a particular statement to be true if the kind of state of affairs which it would take to make it true is conceived, *ex hypothesi*, as something beyond our experience, something which we cannot confirm and which is insulated from any distinctive impact on our consciousness?[13]

However, as Wright notes, this argument is at best inconclusive. It really only presents the realist with a challenge:

> In order to be more than a challenge, [it] would need the backing of a proven theory of concept-formation of a broadly empiricist sort. [And] the traditional theories of that sort have long been recognized to be inadequate.[14]

The challenge to the realist is thus: give some plausible account of how the knowledge of verification-transcendent truth-conditions which you impute to speakers could have been acquired on the basis of the sorts of facts to which they are exposed in the course of their linguistic training. Whether or not this challenge can be met by the realist is very much an open question, in the absence of a proven theory of concept acquisition.[15]

9.4.2 The manifestation argument

Suppose that we are considering region of discourse D as before. Then this argument can be set out as follows:

(1) We understand the sentences of D.

Suppose, for reductio, that

(2) The sentences of D have verification-transcendent truth-conditions.

From (1) and thesis 10 we have:

(3) We grasp the senses of the sentences of D: i.e. we know their truth-conditions.

We then add the following premise, which stems from the Wittgensteinian insight that understanding does not consist in the possession of an inner state, but rather in the possession of some practical ability (see §5.1):

(4) If speakers possess a piece of knowledge which is constitutive of linguistic understanding, then that knowledge should be *manifested* in speakers' use of the language i.e. in their exercise of the practical abilities which constitute linguistic understanding.

It now follows from (1), (2), and (3) that:

(5) Our knowledge of the verification-transcendent truth-conditions of the sentences of D should be manifested in our use of those sentences, i.e. in our exercise of the practical abilities which constitute our understanding of D.

Since

(6) Such knowledge is never manifested in the exercise of the practical abilities which constitute our understanding of D,

it follows that

(7) We do not possess knowledge of the truth-conditions of D.

(7) and (3) together give us a contradiction, whence, by reductio, we reject (2) to obtain:

(8) The sentences of D do not have verification-transcendent truth-conditions, so realism about the subject matter of D must be rejected.

The basic point is that, so far as an account of speakers' understanding goes, the ascription of knowledge of verification-transcendent truth-conditions is simply *redundant*: there is no good reason for ascribing it. Consider one of the sentences we considered earlier as candidates for possessing verification-transcendent truth-conditions, "James II had a migraine on the afternoon of his

32nd birthday" or "Every even number greater than two is the sum of two primes". The realist account views our understanding of these sentences as consisting in our knowledge of a potentially verification-transcendent truth-condition. But

> How can that account be viewed as a description of any *practical* ability of use? No doubt someone who understands such a statement can be expected to have many relevant practical abilities. He will be able to appraise evidence for or against it, should any be available, or to recognize that no information in his possession bears on it. He will be able to recognize at least some of its logical consequences, and to identify beliefs from which commitment to it would follow. And he will, presumably, show himself sensitive to conditions under which it is appropriate to ascribe propositional attitudes embedding the statement to himself and to others, and sensitive to the explanatory significance of such ascriptions. In short: in these and perhaps other important respects, he will show himself competent to use the sentence. But the headings under which his practical abilities fall so far involve no mention of evidence-transcendent truth-conditions.[16]

This establishes (6), and the conclusion follows swiftly. A detailed assessment of the plausibility of this argument is impossible here: but we should note that premise (4) depends upon an interpretation of Wittgenstein's rule-following considerations, and that this (see Chapters 5 and 6) is an extremely controversial matter. In particular, one issue that needs to be addressed is whether the interpretation of premise (4) required for the anti-realist argument is left intact by McDowell's interpretation (§6.7) according to which understanding can harmlessly be construed as a state of mind.[17]

9.4.3 The argument from rule-following

The third anti-realist argument involves a more direct use of Wittgenstein's rule-following considerations. In expounding this argument Wright introduces a number of different species of objectivity. First of all, there is the sort of objectivity which

is possessed by regions of discourse whose sentences have potentially verification-transcendent truth-conditions. Wright calls this the *objectivity of truth*:

> To hold that a class of statements may be fully intelligible to us although resolving their truth-values may defeat our cognitive powers (even when idealized) may naturally be described as believing in the *objectivity of truth*.[18]

In addition, there is a distinct notion of objectivity, which Wright calls the *objectivity of meaning*: this is the notion that

> The meaning of a statement is a real constraint, to which we are bound, as it were, by contract, and to which verdicts about its truth-value may objectively conform, quite independently of our considered opinion on the matter.[19]

Now, Wright argues that if a discourse exhibits objectivity of truth then it follows that it must also exhibit the objectivity of meaning:

> The objectivity of meaning is a manifest implication of the objectivity of truth. If statements of certain sorts can be undetectably true, then we have no alternative but to think of their meanings as, so to speak, reaching into regions where we cannot follow: there is *already* a verdict about the truth-value of such a statement which – if it is intelligible to suppose that our cognitive powers could be appropriately extended – our present understanding of its constituents and syntax would oblige us to give once we had investigated matters properly.[20]

Since objectivity of truth entails objectivity of meaning, we can attack the idea that a discourse exhibits objectivity of truth – has sentences with verification-transcendent truth-conditions – by attacking the idea that it exhibits objectivity of meaning. For a given discourse D, the shape of the argument is therefore:

(1) If D exhibits objectivity of truth then it exhibits objectivity of meaning.
(2) D does not exhibit objectivity of meaning.

Therefore:

(3) D does not exhibit objectivity of truth i.e. the sentences of D do not have potentially verification-transcendent truth-conditions, so that realism about D must be rejected.

This argument is certainly valid, and premise (1) seems plausible enough. But what is the argument for premise (2)? Recall that in §6.6 we looked at Wright's "judgement-dependent" conception of meaning. Wright argues that we must view meaning as judgement-dependent if we are to avoid the sceptical paradox about meaning, which Kripke's sceptic reads into Wittgenstein's rule-following considerations. The proper upshot of the rule-following considerations, Wright thinks, is thus that meaning is judgement-dependent. *But if meaning is judgement-dependent, we have to give up the objectivity of meaning*. If the meaning we attach to a statement S is determined by our best judgements, then it is certainly not the case that the meaning of S "is a constraint . . . to which verdicts about its truth-value may objectively conform, or fail to conform, *quite independently of our considered opinion on the matter*"(emphasis added).

The plausibility of the argument from rule-following against verification-transcendent truth thus depends on the plausibility of Wright's judgement-dependent conception of meaning. This is a highly controversial matter. Recall from §6.8 that McDowell opposes Wright's interpretation of Wittgenstein, and he argues that if we give up the judgement-independent conception of meaning we will not be able to find room for the idea that meaning is genuinely normative. In effect, McDowell thinks we have to accept the objectivity of meaning, on pain of losing the notion of meaning altogether. And given that we have to accept the objectivity of meaning, there is no threat to the idea that our sentences can have potentially verification-transcendent truth-conditions:

> There is no standpoint from which we can give a sense-making characterization of linguistic practice other than that of immersion in the practice: and from that standpoint our possibly verification-transcendent world is certainly in the picture.[21]

Evaluating the anti-realist argument from rule-following, then, depends upon attaining a plausible perspective on the correct interpretation and implications of Wittgenstein's rule-following considerations. This provides us with additional impetus for tackling the issues raised in Chapter 6.

Those, then, are the three main anti-realist arguments developed by Dummett and his followers. Evaluation of these arguments is far outwith the scope of this book. In the remainder of this section, we simply note some features of the anti-realist position which they attempt to establish.

Note 1: A sentence is said to be *effectively decidable* if there is some procedure which a speaker can in principle apply in order to determine whether or not the sentence is true. Thus, "2 + 2 = 4" and "Donald Trump had cornflakes for breakfast yesterday" are both effectively decidable: we can carry out an elementary arithmetical calculation in the first case, and we can gather the obvious sorts of evidence in the second case, in order to determine the truth-values of the respective sentences. But "James II had a migraine on the afternoon of his 32nd birthday" and "Every even number greater than two is the sum of two primes" are *not known to be decidable*: in neither case do we know a procedure which we can apply to determine whether or not they are true. Now intuitively, we think that even though these sentences are not known to be decidable, we can nevertheless still assert that they are either true or false: "Every even number greater than two is the sum of two primes" has a determinate truth-value, it's just that we have no guaranteed way of working out what this truth-value is. In other words, even though the sentence is not known to be decidable, we still think that the *principle of bivalence*, that every (non-vague) sentence is determinately either true or false, applies to it. Now this is an idea which is put under pressure by the conclusion of the anti-realist arguments we have been considering. If truth is not verification-transcendent, it is *epistemically constrained*. One way to spell out what it means to say that truth is epistemically constrained is to say that it must be construed in terms of some notion like *correct* or *warranted assertability*: to say that a sentence is true is to say that there is a warrant to assert it, or that it possesses some other property which is constructed out of warranted assertability.[22] Now given that truth is thus epistemically constrained, what can we say about "Every even number greater than two is the sum of

two primes"? We do not have a warrant to assert this – since no one has yet been able to construct a mathematical proof of it – not do we have a warrant to assert its negation – since no-one has yet produced a counterexample to it, or established that such a counterexample must exist. Given this, and given that truth is to be to construed in terms of warranted assertability, we cannot assert that the sentence "Every even number is the sum of two primes" is either true or false. That is to say, *we cannot assert a priori the principle of bivalence for sentences that are not known to be decidable*: we cannot assert, *a priori*, that they are either true or false.[23]

Note 2: The anti-realist thus claims that we cannot assert *a priori* the principle of bivalence, at least as applied to sentences that are not known to be decidable. Now the principle of bivalence – that every (non-vague) sentence is determinately either true or false is closely associated with the theorem of classical logic known as *the law of excluded middle*:

$$\vdash P \vee -P$$

Refusing to assert *a priori* the principle of bivalence, as the anti-realist proposes, thus appears to threaten the law of excluded middle, and the classical system of logic which is founded upon it. There is much debate among anti-realists as to whether anti-realism implies *revisionism* about classical logic: Dummett has argued that anti-realism implies that classical logic must be given up in favour of some form of *intuitionistic* logic which does not have the law of excluded middle as a theorem.[24]

Note 3: It is important to be clear that although the anti-realist claims that we cannot assert that sentences not known to be decidable are either true or false, he is *not* claiming that we *can* assert that they are *neither* true nor false. Dummett is explicit that although the anti-realist does not wish to assert *a priori* the principle of bivalence, he does not reject the principle of *tertium non datur*, that there is no third truth-value ("neither true nor false") standing between truth and falsity.[25]

Note 4: Note that the anti-realist attitude to sentences which are not known to be decidable is completely different from the logical

positivist attitude to sentences which are not in principle verifiable. Whereas the logical positivist claims that such sentences – because they are in principle unverifiable – are literally meaningless (see Chapter 3), the anti-realist claims that sentences which are not known to be decidable *are* meaningful but that their meanings have to be construed in terms of an epistemically constrained notion of truth.

Note 5: The approach to metaphysical questions which is exhibited here is indicative of Dummett's view that the philosophy of language – in particular, the theory of meaning – has a *foundational* role to play within philosophy. Dummett believes that one of the reasons why philosophical speculation about metaphysical issues has made little progress over the centuries is that the opposing positions in various metaphysical disputes have only been explained in pictorial, or metaphorical terms:

> Even to attempt to evaluate the direct metaphysical arguments, we have to treat the opposing theses as though their content were quite clear and it were solely a matter of deciding which is true; whereas . . . the principal difficulty is that, while one or another of the competing pictures may appear compelling, we have no way to explain in non-pictorial terms what accepting it amounts to.[26]

Dummett's approach is intended to remedy this: the metaphysical disputes are recast as disputes about "the correct model of meaning for statements of the disputed class", thus giving the debates some non-metaphorical content, and enabling the disputes to be resolved within the theory of meaning.[27]

Note 6: Dummett's view that the philosophy of language has a foundational role to play within philosophy is mirrored by his stance on the linguistic meaning/mental content priority issue (see §6.1):

> It has until recently been a basic tenet of analytical philosophy . . . that the philosophy of thought can be approached only through the philosophy of language. That is to say, there can be no account of what thought is, independently of its means of

> expression; but the purpose of the philosophy of thought can be achieved by an explanation of what it is for words and sentences to have the meanings that they bear, an explanation making no appeal to an antecedent conception of the thoughts those sentences express.[28]

We shall return to the debate between realism and anti-realism in §9.6. First, an interlude in which we consider another important issue about the relationship between meaning and world.

9.5 Twin-Earth, meaning, mind, and world

Recall from §2.5 that, according to Kripke, proper names are rigid designators: they refer to the same individual in every possible world in which that individual exists. Kripke, and another famous American philosopher, Hilary Putnam, argue that just as the proper name "Hesperus" rigidly designates Hesperus, a kind term such as "water" rigidly designates a kind of stuff. And just as the sentence "Hesperus is Phosphorus"' expresses a necessary truth which is nevertheless *a posteriori*, the sentence "Water is H_2O" expresses a necessary truth which is discoverable only via empirical investigation. In discovering that water is H_2O, we are discovering the *essence* of water. Simplifying greatly, just as Hesperus rigidly denotes an individual – the planet Venus – which stands in a certain sort of privileged causal relationship to our uses of the name "Hesperus" (in that there is a causal chain leading back from our current uses to an event in which the planet was christened with the name "Hesperus"), the expression "water" rigidly denotes a scientifically identified kind which stands in a certain sort of privileged causal relationship to our perceptions of the features used in the initial baptism to fix the reference of "water". We could sum this fact about "water" up as follows. The superficial features of water include the fact that it is clear, odourless, thirst-quenching, falls from the sky, fills rivers and lakes, and so on.

> (W) "water" refers to that kind of stuff which is dominantly causally responsible for our perceptions of co-instantiations of the superficial features.

It follows from this that, for English speakers, "water" denotes H_2O, since that is in fact the kind of stuff which is relevantly causally responsible. This shows that "water", as we use it, is a *natural kind* term. If we discovered a substance with the same superficial features as water, but with a different underlying molecular structure, that substance would not be water, the superficial similarities notwithstanding. A test we can apply to a term X to see whether it denotes a natural kind is to ask whether it is possible for there to be "fools-X", stuff with all the superficial features of X but which is nevertheless not actually X. "Gold" is thus a natural kind term since not only is "fool's gold" possible, there actually is such a substance (iron pyrites).

Hilary Putnam devised some ingenious arguments designed, as he put it in his classic paper "The Meaning of 'Meaning'", to show that "meanings just ain't in the head", or to suggest a form of what is nowadays known as *externalism*, the view that the physical and social environments in which speakers actually live at least partly constitute the fact that they mean what they mean by at least some of their words. Putnam's most famous argument for externalism is that involving the thought-experiment of "Twin-Earth".

In Putnam's thought-experiment we imagine that Earth and Twin-Earth are very distant parts of the same possible world. There is only one difference between Earth and Twin-Earth: whereas on Earth, water – the clear, odourless, thirst-quenching liquid that fills rivers and lakes, etc. – has underlying molecular structure H_2O, on Twin-Earth the superficially indistinguishable clear, odourless, thirst-quenching liquid that fills rivers and lakes, etc. has underlying structure XYZ. Suppose that Oscar is an Earthian speaker, while Toscar is a Twin-Earthian speaker, where Oscar and Toscar are atom-for-atom doppelgängers. If Oscar were transported to Twin-Earth, his application of "water" to the clear, odourless liquid in the glass would be false, since only things with the molecular structure H_2O count as water in his language; likewise if Toscar were transported to Earth, his application of "water" to the clear, odourless liquid in the glass would be false, since only things with the molecular structure XYZ count as water in his language. Thus, "water" as used by Oscar has a different extension from "water" as used by Toscar. So, the thought expressed by Oscar's utterance of "The River Severn is full of water" is different

from the thought expressed by Toscar's utterance of "The River Severn is full of water". Since Oscar and Toscar are atom-for-atom doppelgängers, it follows that "meanings ain't in the head".[29]

We can spell out the consequences of this argument using the Fregean notions of sense, extension, and semantic value, as developed in Chapters 1 and 2. The extension of "water" as used by Oscar is different from the extension of "water" as used by Toscar. Since the semantic value of a predicate is determined by its extension, "water" as used by Oscar has a different semantic value from "water" as used by Toscar. So the sentence "The River Severn is full of water" as uttered by Oscar expresses a different thought (sense) from that expressed by the same sentence as uttered by Toscar. Thus, the nature of Oscar's environment constrains the extension of his term, the extension of the term constrains its semantic value, and its semantic value constrains its sense. So the nature of Oscar's environment constrains the sense of his term "water". Likewise for Toscar.

Putnam develops another argument in terms of what he calls "The Division of Linguistic Labour". Suppose that Osbert uses the terms "Elm" and "Beech", but is himself incapable of distinguishing between them or even of saying in what the difference between them consists. All he knows is that Elms and Beeches are different kinds of deciduous trees. Now some members of Osbert's speech-community – the tree experts – do know what the difference between beeches and elms consists in, and are capable of distinguishing between them. Putnam contends that because Osbert defers to the tree experts, when he uses "Beech" it refers to beeches and when he uses "Elm" he refers to elms. Now imagine Tosbert – an atom-for-atom doppelgänger of Osbert – on Twin-Earth, where the only difference between there and Earth is that "Beech" is used by Twin-Earthers to refer to elms and "Elm" is used by them to beeches. Tosbert, like Osbert, defers to the experts on his planet, and does not himself grasp the difference between beeches and elms. When Tosbert uses "Beech" it refers to elms and when he uses "Elm" it refers to beeches. The sentence "There is a beech tree in the quadrangle", as uttered by Osbert, expresses a different thought from that expressed by the same sentence, as uttered by Tosbert. Since Osbert and Tosbert are atom-for-atom doppelgängers it follows again that "meanings ain't in the head".[30]

The externalist conclusion applies to states of mind as well as to language. Recall from §6.1 that just as linguistic expressions have meaning, mental states such as beliefs have content. The Twin-Earth argument can simply be rerun to establish that the belief Oscar expresses by "The River Severn is full of water" has a different content from the belief expressed by Toscar's utterance of the same sentence: the former belief is true if and only if the Severn is full of H_2O, the latter belief is true if and only if the Twin-Severn is full of XYZ.[31]

What are the implications of Putnam's Twin-Earth argument? In order to get clear on the possible options, we can consider the following claims, focusing on the case of Oscar and Toscar:

A. The sense of "water" determines its extension.

B. Grasping the sense of "water" is a matter of being in a certain psychological state.

(A) is a simple consequence of Frege's thesis 8, as outlined in Chapter 2 above. (B) should be uncontroversial: regardless of one's views on the relationship between language and thought, the idea that linguistic understanding is a psychological state should seem anodyne.

We can now add:

C. Psychological states are "psychological states in the narrow sense" ("states whose attribution to a subject entails nothing about her environment"[32]).

D. As far as psychological states in the narrow sense are concerned, Oscar and Toscar are identical.

E. Given (A), (B), (C), and (D), the extension of "water", as used by Oscar, is the same as the extension of "water", as used by Toscar.

(C) is an expression of the "methodological solipsist" view of psychological states as *internal* states of agents: an example of this kind of internalist view of the mind might be physicalist views according to which mental states consist in complex facts about

physical states of the brain.[33] (D) is a straightforward aspect of the thought-experiment: since Oscar and Toscar are atom-for-atom doppelgängers they are equivalent with regard to narrow psychological states. (E) is a simple consequence of (A), (B), (C), and (D).

Now, from the Twin-Earth argument rehearsed above:

> F. The extensions of "water", as used by Oscar, is different from the extension of "water", as used by Toscar.
>
> G. Contradiction, from (E) and (F).
>
> We must give up either (A), (B), (C), or (D).

Since (G) is a contradiction, something in the argument up to that point has to be rejected to resolve the contradiction. The main options include:

Putnam's "meanings ain't in the mind": This reaction to the argument involves denying that grasping a sense is a matter of being in a psychological state, and hence resolving the contradiction by jettisoning (B). This is arguably the reaction advocated by Putnam himself in "The Meaning of 'Meaning'": since psychological states are "in the head" and states of grasping senses of natural kind terms like "water" aren't, it follows that "Meanings ain't in the head", and so ain't in the mind either.[34]

Fodor's "bifurcated content": According to (A) the sense of "water" determines its extension. One way of avoiding the contradiction would be to argue that the extension of "water" is determined, not by the sense of "water" alone, but by the sense of "water" together with the *context of utterance*. Just as the reference of an indexical like "I" varies with context of utterance (if I utter it it refers to me, if you utter it it refers to you) so too does the extension of "water" vary with context of utterance: as used on Earth its extension includes only H_2O, as used on Twin-Earth its extension includes only XYZ. Fodor now defines *narrow content* as a *function* from *context of utterance to reference*: Oscar and Toscar assign the same narrow content to "water" (just as you and I attach the same meaning to "I"), but since Oscar and Toscar

inhabit different contexts, the narrow content together with their respective contexts, determine H_2O and XYZ as the referents of their respective uses of "water".[35] We thus resolve the contradiction by modifying (A). Oscar and Toscar thus differ in regard to the broad content (truth-conditions) of "The Severn is full of water", but coincide in point of narrow content (function from context of utterance to truth-condition of "The Severn is full of water").

McGinn's "structured content": Building on considerations similar to those broached by Fodor in order to resolve the contradiction, McGinn considers a "duplex" or "two component" conception of contentful states as *composite*; according to the duplex conception "the concept of command of a meaning . . . is the concept of something that is, in itself, in the head, but conceived in terms of its relations to what is outside the head".[36] Thus the state which constitutes the "in-the-head" or "narrow" component of the thought expressed by Oscar's "The Severn is full of water" could be possessed by Toscar, where, because of its sustaining some different relation to what is outside the head, it constitutes an "in-the-head" component of an entirely different "broad" thought (i.e. the thought expressed by Toscar's "The Severn is full of water"). The states expressed by Oscar's and Toscar's respective utterances of "The Severn is full of water" thus share a narrow (but not broad) component.

McDowell's externalism: Possibly the most radical of the main responses to the Twin-Earth scenario is the externalist attempt to resolve the contradiction by rejecting (C) and fully embracing the idea that linguistic understanding and mental content are "world-involving". According to McDowell, we ought to reject (C), thereby opening up the possibility of a conception of psychological states which will allow us to embrace both (A) and (B):

> What is to be learned from [the Twin-Earth] reflections is not, as Putnam himself argues, that it cannot be true both that "knowing the meaning of a term is just a matter of being in a certain psychological state" and that "the meaning of a term determines its extension"; so that if we retain the second of

> these assumptions, we must renounce the first. This presupposes that anyone who embraces the first assumption must be restricting psychological states to "narrow" states. Rather, the moral of Putnam's considerations is that the idea of a psychological state, as it figures in the first assumption, cannot be the idea of a "narrow" state. That is: we should not leave in place an idea of the mind that is shaped by the tenets of "methodological solipsism", and conclude that meanings are not in the mind, since they are not in the head. Rather, we should read the two assumptions in such a way that they can be true together and exploit such a reading to force us into explicit consideration of a different conception of the mind.[37]

It would be a mistake to view McDowell's externalist as arguing that at least some mental states are "broad" or that Putnam's reflections suggest a "duplex" conception of content: rather, in rejecting (C) we are rejecting a conception of the mind that appears to require a distinction between two kinds of content, narrow and broad, or between two aspects of content, the "in-the-head" aspect and the aspect from which the "in-the-head" component is viewed as related to external reality:

> [Putnam] assumes that anyone who wants to conceive knowledge of a meaning as wholly a matter of how it is with someone's mind must be already committed to a theoretical conception of the mind . . . which, in conjunction with Putnam's reflections about meaning, guarantees that the wish cannot be fulfilled . . . What [he] never seems to consider is the possibility of a position that holds that command of a meaning is wholly a matter of how it is with someone's mind, and combines that with the determination of extension by meaning so as to force a radically non-solipsistic conception of the mind to come to explicit expression.[38]

Dissolving Twin-Earth: An option distinct from all of the above would defuse the apparent contradiction by denying (F) and arguing that in the Twin-Earth scenario the extension of "water", as used by Oscar, is actually identical to the extension of "water",

as used by Toscar. Tim Crane argues along these lines, suggesting that in the initial Twin-Earth thought-experiment we can take the extension of "water", as used by *both* Oscar *and* Toscar, to include only substances that are either H_2O *or* XYZ.[39] Although experts on Earth, who have a complete understanding of "water" as it is used on Earth, refer to H_2O in their uses of "water", and experts on Twin-Earth, who have a complete understanding of "water" as it is used on Twin-Earth, refer to XYZ in their uses of "water", we can view Oscar and Toscar, both non-experts, as having an *incomplete* understanding of "water" as it is used respectively on Earth and Twin-Earth. The senses attached by them via their incomplete understanding of "water" determine the property H_2O *or XYZ* as the referent of their respective uses of the term. Crane thus blocks the Twin-Earth argument at the crucial stage, and since there is no need to introduce the notion of broad content, there is likewise no need to introduce narrow content, in the technical sense defined by (C).[40]

The literature spawned by Putnam's thought-experiment is large and complex and in this section we have attempted only the barest outline of some of the philosophical views that have been developed in reaction to it. In the remainder of the chapter we return to the issue about the role of philosophy of language in metaphysical disputes between realists and anti-realists.

9.6 Grades of objectivity: Wright on anti-realism

To be a realist about a certain area of discourse is to view the sentences of that discourse as possessing objective truth-conditions. Earlier in this chapter, we looked at Dummett's way of cashing out the notion of an objective truth-condition: a truth-condition is objective if it is potentially verification-transcendent. Thus, one way to oppose realism, to espouse anti-realism, is to deny that truth is potentially verification-transcendent and to argue that truth must be viewed as epistemically constrained. But is this the best way of cashing out the metaphysical debates between realists and their opponents? Although Dummett's way of characterising the metaphysical debate seems to be appropriate in some cases – e.g.

mathematics, statements about the past – there are other cases where it simply seems besides the point. Consider discourse about, for example, morals or comedy. It seems that in these cases a moral realist may not wish to claim that the truth-conditions of the relevant sentences are potentially verification-transcendent and that both the moral realist and the moral anti-realist can *agree* that statements about comedy or moral value do not have verification-transcendent truth-conditions:

> There are, no doubt, kinds of moral realism [or realism about comedy] which do have the consequence that moral [or comic] reality may transcend all possibility of detection. But it is surely not essential to any view worth regarding as realist about morals [or comedy] that it incorporate a commitment to that idea.[41]

Intuitively, a sensible version of realism about "That remark was funny" or "That deed was wrong" does not have to view facts about funniness or wrongness as potentially verification-transcendent. So although construing objective truth-conditions as verification-transcendent truth-conditions may be useful for characterising realism about some areas of discourse, there are other areas for which this is not a useful characterisation. The upshot of this is that we need other ways of fleshing out the notion of an objective truth-condition, other ways of construing objectivity. The most important recent work in the area of overlap between the philosophy of language and metaphysics has consisted of attempts to do just this. In this section, we'll briefly sketch one such attempt in very broad outline, that developed by Wright.

Wright argues that non-cognitivism – the denial that the sentences of a discourse are truth-apt or even possess truth-conditions – does not provide a useful way of formulating opposition to realism. The debate between realism and anti-realism about a discourse takes place only *after* it has been granted that the sentences of that discourse are truth-apt. There are two main parts to Wright's sketch of the shape of the debates. Firstly, he develops a version of *minimalism about truth-aptness*, according to which all of the discourses – including morals, comedy, the external world, mathematics, the past, and so on – do turn out to be truth-apt.

Secondly, he develops a *number* of ways of characterising realism and anti-realism about discourses whose truth-aptness has already been granted – that is, a number of different ways in which truth-conditions can be objective. It turns out that viewing the sentences of a discourse as having potentially verification-transcendent truth-conditions is only *one* of a *number* of ways of characterising realism and the notion of objective truth-conditions. We'll look at these two parts of Wright's sketch in turn.

9.6.1 Minimalism about truth-aptness

According to minimalism about truth-aptness, a class of sentences is truth-apt if the following two conditions are satisfied:

Discipline: the sentences figure in an area of discourse which is disciplined, where this means that there are standards operative with respect to which uses of those sentence are deemed to be warranted or not. In other words, there are acknowledged standards for the proper and improper use of those sentences.

Syntax: the sentences possess the right sort of syntactic features: in particular, they must be capable of conditionalisation, negation, embedding in propositional attitudes, and so on.

If these conditions are met, then on the minimalist theory of truth-aptness, there is simply no option but to concede that the sentences are truth-apt, are apt to be evaluated in terms of truth and falsity. And these conditions are met in all of the discourses which have been the focus of metaphysical debate. For example, moral discourse is certainly disciplined: there are acknowledged standards governing the use of moral sentences. Take the sentence "Jones did the right thing in lying to Smith". The utterance of this sentence would (arguably) be warranted in circumstances in which Smith, intending to murder Jones's family, had requested information

about their whereabouts, and it is easy to think of different circumstances in which the utterance of the sentence would not be warranted. And there seems to be no problem in satisfying the condition concerning syntax: we can perfectly well say things like "If Jones did the right thing in lying to Smith then that shows that it is sometimes morally permissible to lie", "it is not the case that Jones did the right thing in lying to Smith", "I believe that Jones did the right thing in lying to Smith". All of these are syntactically kosher. So, both of the conditions which minimalism imposes on truth-aptitude are satisfied by "Jones did the right thing in lying to Smith" (and by other declarative sentences of moral discourse). Those sentences are therefore truth-apt.[42] The same goes for "That joke was hilarious", "Every even number greater than two is the sum of two primes", and "James II had a migraine on the afternoon of his 32nd birthday".

9.6.2 The marks of realism

Given that the sentences of a region of discourse are, courtesy of minimalism, truth-apt, how can we characterise realism about that discourse? One way is given by Dummett's characterisation of realism:

> ***Mark 1*** Verification-transcendence: The sentences of the region have truth-conditions which are potentially verification-transcendent.

But as we saw, this mark doesn't give us a good way of getting a handle on the realist/anti-realist debates about morals and comedy, because it is possible that a realist and an anti-realist can agree in those cases that discourse about morals and discourse about humour do not satisfy this mark. In order to deal with this, and to widen the range of realist/anti-realist disputes that we can capture, Wright develops a number of other "marks of realism":

> ***Mark 2*** Cognitive Command: It is *a priori* that, if two practitioners of the discourse differ in their judgements, then given that the divergence of opinion is not due to vagueness, at least one of them is making some *cognitive* error.

Discourse about comedy, for example, might appear to fail cognitive command: if Jones thinks Benny Hill is funnier than Billy Connolly, and Smith thinks Billy Connolly is funnier than Benny Hill, then even if this divergence in opinion is not attributable to vagueness, it is not *a priori* that either Smith or Jones is prey to some sort of cognitive malfunction. It may just be that they have different non-cognitive senses of humour. This contrasts with beliefs about the shape of middle-sized physical objects: if Smith says the object is square and Jones says it is round, then it is *a priori* that unless the disagreement is explicable on the grounds of vagueness, at least one of Smith and Jones will be prey to some sort of cognitive shortcoming.

> ***Mark 3*** Width of Cosmological Role: The truth-conditions of the sentences of the discourse feature in the explanation of facts other than the holding of speakers' beliefs, and do so other than via their role in the explanation of the holding of those beliefs.

Moral discourse appears to fail this condition: moral facts only seem to have an impact on reality via the contribution they make to explanations of our moral beliefs (e.g. the injustice of a given social system, such as slavery, certainly can have effects on the world (e.g. it can lead to the eventual downfall of that system), but only via human *attitudes* to slavery). This contrasts with discourse about middle-sized physical objects, which appears to satisfy the condition: facts about middle-sized physical objects can figure in explanations of other sorts of facts and other than via the role they play in explaining why we have beliefs about them (e.g. the fact that the snowman melted explains why the ground is wet).

> ***Mark 4*** Judgement-Independence: The central predicates of the region of discourse are *judgement-independent* (in the sense of §6.6).

As we saw in §6.6, Wright argues that discourse about meaning and discourse about intention both fail this condition (and he argues that it is satisfied by discourse about the *shape* of medium-sized physical objects).

These four marks require substantial elaboration and development which we cannot attempt here. What is important is the shape they allow us to impose on realist/anti-realist debates. One way in which the sentences of a discourse can have objective truth-conditions is for that discourse to satisfy Mark 1: but other ways are provided by the satisfaction of some of the other marks. And there is no reason why a discourse cannot satisfy some of the marks, but not all of them: for example, discourse about colour may satisfy Mark 3, but not Marks 1, 2, and 4. What this means is that the shape of the debate between realism and anti-realism is far messier than Dummett conceived. At the realist end of the spectrum, there will be discourses whose sentences have objective truth-conditions, in the senses of "objective" given by each of Marks 1–4. And the anti-realist end of the spectrum there will be discourses which fail to satisfy any of the various marks of realism, which have truth-conditions which are not objective in any of the senses defined by these marks. But in between there will be many discourses which satisfy some of the marks, but not all, which may possess objective truth-conditions in some of the senses defined by the marks, but not all.[43]

On Wright's conception of the realist/anti-realist debate, then, the debate has to proceed in a much more *piecemeal* fashion than that envisaged by Dummett.[44]

9.7 Two threats of quietism

That, then, is an outline of Wright's conception of the general shape of the realist/anti-realist debate. This conception of the

shape of metaphysical debate is threatened by *quietism*, which is the view that "significant metaphysical debate is impossible". Quietism threatens to close down metaphysical debate in two ways, one direct, one indirect.

The *direct* route to quietism proceeds via Wittgenstein's rule-following considerations. As we saw above, Wright believes that the rule-following considerations endanger the objectivity of meaning, the view that "the meaning of a statement is a real constraint, to which we are bound . . . by contract, and to which verdicts about its truth-value may objectively conform, quite independently of our considered opinion on the matter". The meaning of a statement imposes requirements on what counts as correct use of the statement, which, once they are in place, determine which uses are correct and incorrect independently of any opinions we may subsequently form. The problem is that if the rule-following considerations undermine the idea that meanings are objective in this sense, they thereby threaten the various ways in which anti-realists attempt to draw comparisons concerning objectivity. For example, we saw in §9.4 that objectivity of truth implies objectivity of meaning. If a global rejection of objectivity of meaning is forced by the rule-following considerations, then so will a global rejection of objectivity of truth. And if no discourses possess objectivity of truth, appealing to failure of objectivity of truth will be useless for drawing metaphysical comparisons between discourses. Likewise, since the truth of any statement is a function of its meaning together with facts about the world (see §5.3), rejection of objectivity of meaning may entail that *all* predicates are judgement-dependent. The possibility of appealing to the judgement-independent/judgement-dependent distinction in order to draw a metaphysical contrast between discourses will be endangered. In short, the rule-following considerations seem to threaten us with quietism about objectivity: the view that no principled, metaphysically interesting contrasts between grades of objectivity can be drawn.

Anti-realists try to find ways of avoiding quietism, while retaining their interpretation of the rule-following considerations.[45] But we should note that some philosophers, such as McDowell, think that there is a different, *indirect*, way in which quietism can close down metaphysical debate. As we saw above, McDowell does not view the

rule-following considerations as threatening the objectivity of meaning. So the metaphysical relevance of the various marks of truth, such as verification-transcendence and judgement-independence, is not threatened in the direct manner envisaged in the previous paragraph. The distinctions the anti-realist wishes to draw can still be drawn: instead, the rule-following considerations are taken to threaten the idea that there is any interesting metaphysical *point* to be made by appealing to the distinctions in the first place. For example, it might be argued that the metaphysical relevance of Wright's judgement-independent/judgement-dependent distinction depends upon a conception of detecting or tracking facts that the rule-following considerations actually display to be untenable. The idea would be that the distinction depends upon a notion of detection or tracking reminiscent of the "master-thesis" discussed in §6.7: we can only think of ourselves as detecting or tracking facts or states of affairs which just "stand there", which are normatively inert. Wright would be interpreted as arguing that since facts about meanings are not normatively inert (they sustain substantial, *a priori* relations to facts about correct use), we cannot think of ourselves as detecting or tracking them, and must instead view them as judgement-dependent. McDowell would argue that once this notion of detection or tracking has been rejected, as it must be if we are to avoid losing our grip on the notion of normativity of meaning altogether, the judgement-independent/judgement-dependent distinction loses its metaphysical bite, since it was a distinction between discourses in which our best judgements play a tracking or detecting role and discourses in which such judgements play a constitutive role.

The issue of whether anti-realists can block these two threats of quietism about metaphysics and objectivity is complex and difficult. For one thing, it depends on getting clear about the proper implications of Wittgenstein's reaction to KW's sceptical problem about rule-following (§6.7). It also depends on getting clearer about the precise formulation of the "marks of realism" discussed in §9.6. But it is very much a *live* issue, and represents perhaps the most interesting question concerning the relationship between meaning and metaphysics facing philosophers of language in the early decades of the twenty-first century.

Chapter summary

In this chapter we characterised realism about a particular subject matter as the view that our talk and thought about that subject matter aspires to reflect an objective reality, and sometimes succeeds in this aim, and we cashed this out as the idea that sentences about that subject matter have objective truth-conditions. We looked at Dummett's conception of objective truth-conditions as potentially verification-transcendent truth-conditions, and outlined three main anti-realist arguments against objective truth-conditions, so conceived. We then looked at how Wright modifies Dummett's conception of what it is for a truth-condition to be objective and introduced quietism, the view that it is not possible to have genuine debate about the objectivity of truth-conditions in the manner proposed by Dummett and Wright. Inter alia, we looked at a distinct set of issues about language and how it relates to the world, the debate between semantic externalism and internalism.

Study questions

1. Give two examples – not involving sentences about the past or mathematics – that intuitively have potentially verification-transcendent truth-conditions.
2. What sorts of facts, in general, are speakers exposed to in the course of their linguistic training? How does your answer to this question impact on Dummett's "acquisition challenge"?
3. Can you think of a practical ability whose exercise would manifest grasp of a potentially verification-transcendent truth-condition?
4. Are there any good reasons to reject the "objectivity of meaning", as characterised by Wright?
5. Outline the differences between the responses to Putnam's Twin-Earth thought-experiment developed by Putnam himself, Fodor, McGinn, McDowell, and Crane. Which is the most plausible?
6. Is Wright's conception of the nature of realism an improvement on Dummett's?
7. What grounds might there be for adopting quietism about the debate between realism and anti-realism?

Further reading

Good, book-length introductions to the issues discussed in this chapter are Kirk, *Relativism and Reality* and Brock and Mares, *Realism and Antirealism*. For a useful discussion of the relationship between metaphysics and the philosophy of language, see Morris, "Metaphysics, Philosophy, and the Philosophy of Language".

For recent attempts to defend non-cognitivism against the Frege–Geach problem, see Blackburn, *Spreading the Word*, Chapters 5 and 6; Gibbard, *Wise Choices, Apt Feelings*, Chapter 5, and *Thinking How to Live*, Chapter 3. For an overview of the difficulties faced by non-cognitivism, see Hale, "Can There Be a Logic of Attitudes?", and Miller, *Contemporary Metaethics: An Introduction*, Chapters 3–5.

For a good introductory chapter on the debate between internalism and externalism, see Haukioja, "Internalism and Externalism". For good, book-length studies, see McCulloch, *The Mind and Its World*, and Kallestrup, *Semantic Externalism*. Externalist positions are defended in Evans, *The Varieties of Reference*; McDowell, *Mind and World*; and McCulloch, *The Life of the Mind*. For internalism, see Segal, *A Slim Book about Narrow Content*. Two useful (but demanding) volumes on the debate are Woodfield (ed.), *Thought and Object*, and McDowell and Pettit (eds), *Subject, Thought and Context*. A good collection of central sources is Pessin and Goldberg (eds), *The Twin-Earth Chronicles*.

Dummett's writings are extensive and difficult. An excellent overview and exposition can be found in Weiss, *Michael Dummett*. A clear account of the general project can be found in the introduction to Dummett, *The Logical Basis of Metaphysics* (though the rest of this book is extremely tough). Relevant papers are "Truth", "Realism", "The Philosophical Basis of Intuitionistic Logic", and "The Reality of the Past", all reprinted in Dummett's *Truth and Other Enigmas*, and "What is a Theory of Meaning? (II)", reprinted in his *The Seas of Language*. Important articles critical of Dummett include McDowell's "On 'The Reality of the Past" and "Anti-Realism and the Epistemology of Understanding". See also Miller, "Realism and Antirealism", "What Is the Acquisition

Argument?", and "What Is the Manifestation Argument?", as well as Hale, "Realism and Its Oppositions". For an approach to the issues that contrasts sharply with that of Dummett and Wright, see Devitt, *Realism and Truth*. For yet another approach see Fine, "The Question of Realism". A good place to start on Dummett, and on anti-realism generally, is the introduction to Wright's collected papers, *Realism, Meaning, and Truth* (in which he also responds to some of McDowell's criticisms). Wright's own conception of the shape of the realist/anti-realist debate is developed in his *Truth and Objectivity*: the first five chapters of this are devoted to developing the "marks of realism" discussed in §9.5, and the sixth chapter contains an extensive discussion of the threats of quietism discussed in §9.6. For more on quietism, see McDowell, "Wittgensteinian Quietism", and Wright, "Comrades against Quietism". See also the papers collected in Wright's *Saving the Differences*.

Notes

1 Note that in this chapter we will be concerned exclusively with issues about realism construed as issues about the objectivity of truth-conditions: we will not concern ourselves here with issues about realism construed as ontological issues about the existence of objects. Here we are concerned only with realism about e.g. mathematics construed as an issue about the objectivity of the truth-conditions of mathematical statements, not as an ontological issue about the existence of mathematical objects. For discussion of how these approaches to questions about realism are related, see the introduction to Wright, *Realism, Meaning, and Truth* and *Frege's Conception of Numbers as Objects*.

2 Compare the strategy which the non-cognitivist about morals adopts with that which is adopted in KW's sceptical solution (Chapter 5) concerning meaning.

3 Compare with the following argument:

> (i) It is raining; (ii) If it is raining then the streets are wet; therefore, (iii) the streets are wet.

In this non-moral case of modus ponens, why is this argument not similarly invalid in virtue of the fact that "It is raining" is asserted in (i), but not in (ii)? The answer is of course that "It is raining" has the same *truth-condition* as it appears in both (i) and (ii); this

answer is not available to the non-cognitivist in the case of examples of modus ponens involving moral premises, since he claims that "Murder is wrong" doesn't *have* a truth-condition.

4 See Further reading for details of some recent attempts to solve the Frege–Geach problem.

5 Actually, James *VII* (of Scotland). But I refrain from correcting Wright's historical error in the text!

6 "Potentially" because we are not saying that evidence never will be found, just that we have no guarantee that it will be.

7 Or, alternatively, potentially proof-transcendent (since in mathematics, verification is a matter of proof).

8 Wright, *Truth and Objectivity*, p.4.

9 Note again that this form of opposition to realism does not involve the claim that the sentences of D are not truth-apt: so it can avoid the Frege–Geach problem which afflicts non-cognitivism. Note also that there are serious questions about what it means to say that the truth of a sentence is not verification-transcendent. Does this mean that the sentence can be verified by us as we actually are? By someone, somewhere, as they actually are? By someone, somewhere, given some suitable idealisation of the their present cognitive powers? And what is permissible as a "suitable idealisation"? And how can the notion of effective decidability (cf. §9.4) be extended from the mathematical case to the empirical domain? These questions must be answered if anti-realism is to have any determinate content. See Wright, *Realism, Meaning, and Truth*, p.32.

10 Davidson, "A Coherence Theory of Truth and Knowledge", p.307, in Lepore (ed.) *Truth and Interpretation: Essays on Davidson*.

11 Frege, "Thought", p.345.

12 The issues about metaphysical realism and truth are in fact considerably more complex than the previous section would suggest; see Miller, "Realism and Antirealism" for an account of some of the complexities.

13 Wright, *Realism, Meaning, and Truth*, p.13.

14 Ibid., p.15.

15 See Miller, "What is the Acquisition Argument?" for a much fuller treatment.

16 Wright, *Realism, Meaning, and Truth*, p.17.

17 See Miller, "What is the Manifestation Argument?" for a much fuller treatment.

18 Wright, *Realism, Meaning, and Truth*, p.5.

19 Ibid.

20 Ibid.

21 McDowell, "Anti-Realism and the Epistemology of Understanding", p.342.
22 This is greatly oversimplified; for more detail, see Wright, *Realism, Meaning, and Truth*, §5, and *Truth and Objectivity*, Chapter 2, where he suggests that for certain discourses, truth may be modelled on "superassertibility". For another attempt to construe truth as essentially epistemically constrained, see Putnam, *Realism, Truth, and History*.
23 Note that another way of describing matters here would involve characterising decidability and undecidability in explicitly epistemic terms: a statement is decidable if we *know* a procedure for determining its truth-value and undecidable if we don't *know* such a procedure. The class of statements the realist and anti-realist disagree over would then be characterised as the class of undecidables. Either way of describing the situation is fine so long as our terminology is consistent. See Shieh, "Undecidability in Anti-Realism" for an illuminating discussion.
24 The issues here are complex. See Wright, "Anti-Realism and Revisionism" and "Realism, Bivalence, and Classical Logic" in his *Realism, Meaning, and Truth*.
25 See Dummett, *Truth and Other Enigmas*, p.xxx. This might seem puzzling. Suppose that the principle of bivalence corresponds to the law of excluded middle: $\vdash$ P v –P, and that the principle of *tertium non datur* corresponds to $\vdash$ – – (P v –P) (is not the case that neither P nor not-P). Since it is a logically valid sequent that – – P $\vdash$ P, doesn't it follow that rejecting $\vdash$ P v –P entails the rejection of $\vdash$ – – (P v –P) i.e. that rejection of the principle of bivalence entails rejection of the principle of *tertium non datur*? The crucial point is that the sequent – – P $\vdash$ P which licenses this entailment is valid in classical logic *but not valid in intuitionistic logic*. Rejection of bivalence entails rejection of *tertium non datur* only given classical logic; but the Dummett-style anti-realist *rejects* classical logic, and so can reject bivalence whilst holding on to *tertium non datur*.
26 Dummett, *Logical Basis of Metaphysics*, p.12.
27 Perhaps the clearest statement of Dummett's general approach to metaphysics can be found in the introduction to *The Logical Basis of Metaphysics*. For criticism, see Blackburn, "Metaphysics", §2, and also Devitt, "Dummett's Anti-Realism".
28 Dummett, *Logical Basis of Metaphysics*, p.3.
29 We'll here simply ignore the complication that Oscar and Toscar aren't quite atom-for-atom doppelgängers, since Oscar's body contains molecules of H_2O whereas Toscar's contains molecules of XYZ.

30 In his classic paper "Individualism and the Mental", Burge develops another thought-experiment that raises similar issue's to Putnam's. Burge's argument is more general than Putnam's, in that it is designed to apply to all terms, not just those that stand for natural kinds.
31 Note that in the literature "semantic externalism" refers to externalism about linguistic meaning, while "cognitive externalism" refers to externalism about mental content.
32 McDowell, "Putnam on Mind and Meaning", p.277. McDowell is here merely characterising the idea of a psychological state's being "narrow": as we'll see, he is not agreeing with the characterisation.
33 Although an internalist need not be a physicalist: see Crane, "All the Difference in the World", §3. As Crane points out, the attraction of (C) stems from considerations relating to *causation*. Psychological states cause behaviour, and they often cause behaviour in virtue of their content: for example, my belief that there is beer in the fridge contributes to the explanation of my reaching for the fridge door in virtue of its having the content *that there is beer in the fridge*. However, if causation involves only intrinsic properties of objects and events, then the idea that some psychological states are broad (rather than narrow) appears to threaten the idea that they can cause behaviour in virtue of their contents. For a useful discussion of some related issues, see Jackson and Pettit, "Functionalism and Broad Content".
34 Note that according to Putnam the difference in meaning between Earth and Twin-Earth does not depend on the fact that experts on Earth and Twin-Earth are able to tell that the stuff they call "water" is respectively H_2O and XYZ: Putnam holds that the difference in meaning obtained even before the development of the scientific expertise necessary to distinguish between the two substances at the level of microstructure. The development of chemistry doesn't change the meaning of "water", rather it tells us something new about the stuff we always referred to by our uses of that term.
35 See Fodor, *Psychosemantics*, Chapter 2.
36 This is McDowell's description of what is involved in the duplex conception; see McDowell, "Meaning and Intention in Wittgenstein", p.37.
37 McDowell, "Putnam on Mind and Meaning", p.282.
38 Ibid. McDowell argues that his reaction to the Twin-Earth scenario is better than that advocated by Putnam by relying on the response to Kripke's Wittgenstein outlined in §6.8. For a critique of this line of argument, see Miller, "Rule-Following and Externalism".

39 Crane actually uses a different example involving aluminium and molybdenum for reasons that we needn't go into here; Crane, "All the Difference in the World".

40 McCulloch, who follows McDowell in taking Putnam's argument to force a rejection of the solipsistic conception of mind enshrined in (C), argues in "The Spirit of Twin-Earth" that Crane's argument is effective against the externalist rejecter of (C) only if something like an internalist theory of meaning is assumed, an assumption that would of course be question-begging in the present context. This seems right but, likewise, McCulloch can be sure that Crane's argument is at least sometimes ineffective against the externalist if he assumes an externalist theory of meaning of the type eschewed by Crane. So all we get is a *stand-off* between externalism and its opponents. I think that this shows that viewing Putnam's thought-experiment as an *argument* for particular philosophical theories has its limitations: in general, thought-experiments only yield theoretical conclusions if they are interpreted against some theoretical background, the assuming of which is likely to beg the question against philosophers wishing to avoid the theoretical conclusion in question. So it is probably more fruitful to view the Twin-Earth thought-experiment, not as an *argument* for a particular theory of mind and meaning, but as *highlighting* aspects of mind and meaning that theories of mind and meaning have to accommodate or explain away. Which theory wins the day will then be settled by higher-level theoretical constraints – what do we want a theory of mind and meaning to explain? – and by which theory of mind and meaning best meets them. This is surely no surprise: we wouldn't expect the deep theoretical truth about the mind to be delivered by a simple thought-experiment anyway, so McCulloch's argument that it isn't won't surprise those who already know that if externalism is true, internalism is false.

41 Wright, *Truth and Objectivity*, p.9.

42 So minimalism about truth-aptness simply undercuts the possibility of non-cognitivism, which denies that the sentences are truth-apt. There has been some vigorous debate about the relationship between non-cognitivism and minimalism: see e.g. Smith, "Why Expressivists about Value Should Love Minimalism about Truth"; Miller and Divers, "Why the Expressivist about Value Should Not Love Minimalism about Truth"; and Horwich, "The Essence of Expressivism" in the symposium on "Minimalism and Expressivism", in *Analysis* (January 1994).

43 So e.g. the realist and anti-realist about comedy or morals can agree that discourse about comedy and discourse about morals fail to satisfy Mark 1, but they can still argue about whether these discourses are objective in any of the senses defined by Marks 2–4.

44 Another way of opposing realism, which we have not considered here, is to adopt an error-theory. This admits that the sentences of the discourse are truth-apt and do have truth-conditions, but denies that such truth-conditions are ever actually satisfied. That is, it argues that the (positive, atomic) sentences of the discourse are s*ystematically and uniformly false*. For an example of an error-theory about moral discourse, see Mackie, *Ethics: Inventing Right and Wrong*; for an example of an error-theory about mathematical discourse, see Field, *Science Without Numbers*. For argument that error-theories do not provide a good way of formulating opposition to realism, see Wright, *Truth and Objectivity*, Chapter 1. See also Miller, *Contemporary Metaethics*, Chapter 6, which in addition to discussing error-theories covers various forms of fictionalist opposition to moral realism.

45 See especially Wright, *Truth and Objectivity*, Chapter 6.

Bibliography

Ayer, A. J. *Language, Truth, and Logic* (New York: Dover Press 1946).

Ayer, A. J. (ed.) *Logical Positivism* (Glencoe, IL: Free Press 1959).

Ayer, A. J. *Part of My Life* (London: Collins 1972).

Ayer, A. J. *More of My Life* (London: Collins 1983).

Bach, K. "Speech-Acts and Pragmatics". In Devitt and Hanley (eds) *The Blackwell Guide to the Philosophy of Language*.

Baldwin, T. "Philosophy of Language in the Twentieth Century". In LePore and Smith (eds) *The Oxford Handbook of Philosophy of Language*.

Beaney, M. *Frege: Making Sense* (London: Duckworth 1996).

Beaney, M. (ed.) *The Frege Reader* (Oxford: Blackwell 1997).

Beaney, M. "Wittgenstein on Language: From Simples to Samples". In LePore and Smith (eds) *The Oxford Handbook of Philosophy of Language*.

Beaney, M. "What is Analytic Philosophy?". In M. Beaney (ed.) *The Oxford Handbook of the History of Analytic Philosophy* (Oxford: Oxford University Press 2013).

Beebee, H. *Hume on Causation* (London: Routledge 2006).

Bell, D. "How 'Russellian' Was Frege?", *Mind* 99 (1990).

Benacerraf, P. and Putnam, H. (eds) *The Philosophy of Mathematics* (Cambridge: Cambridge University Press 1983).

Blackburn, S. *Spreading the Word*. (Oxford: Oxford University Press 1984).

Blackburn, S. "Wittgenstein's Irrealism". In L. Brandl and R. Haller (eds) *Wittgenstein: Eine Neubewehrung* (Vienna: Hölder-Pickler Tempsky 1990).

Blackburn, S. "Circles, Finks, Smells, and Biconditionals", *Philosophical Perspectives 7* (1993).

Blackburn, S. *Ruling Passions* (Oxford: Oxford University Press 1998).

Blackburn, S. "The Individual Strikes Back". In Miller and Wright (eds) *Rule-Following and Meaning*.

Blackburn, S. "Metaphysics". In N. Bunnin and E. Tsui-James (eds) *The Blackwell Companion to Philosophy*, 2nd edn (Oxford: Blackwell, 2002).

Boghossian, P. "Review of McGinn's Wittgenstein on Meaning", *Philosophical Review* 98 (1989).

Boghossian, P. "Naturalizing Content". In G. Rey and B. Loewer (eds) *Meaning in Mind: Essays for Jerry Fodor* (Cambridge, MA: MIT Press 1990).

Boghossian, P. "The Rule-Following Considerations". In Miller and Wright (eds) *Rule-Following and Meaning*.

Boghossian, P. "The Status of Content", Philosophical Review 99 (1990).

Boghossian, P. "The Status of Content Revisited", Pacific Philosophical Quarterly (1990).

Boghossian, P. "The Normativity of Meaning", *Philosophical Issues* 13 (2003).

Boghossian, P. "Is Meaning Normative?". In C. Nimtz and A. Beckermann (eds) *Philosophy - Science - Scientific Philosophy* (Paderborn: Mentis 2005).

Boghossian, P. "Is (Determinate) Meaning a Naturalistic Phenomenon?". In S. Gross, N. Tebben, and M. Williams (eds) *Meaning Without Representation: Essays on Truth, Expression, Normativity, and Naturalism* (Oxford: Oxford University Press 2015).

Boghossian, P. "Postscript: Further Thoughts about Analyticity 20 Years On". In B. Hale, C. Wright, and A. Miller (eds) *A Companion to the Philosophy of Language*. 2nd edn.

Boghossian, P. "Analyticity". In Hale, Wright, and Miller (eds) *A Companion to Philosophy of Language*, 2nd edition.

Borg, E. "Intention-Based Semantics". In LePore and Smith (eds) *The Oxford Handbook of Philosophy of Language*.

Brock, S. and Mares, E. *Realism and Antirealism* (Chesham: Acumen 2007).

Burge, T. "Individualism and the Mental", *Midwest Studies in Philosophy* 4 (1979).

Byrne, A. "On Misinterpreting Kripke's Wittgenstein", *Philosophy and Phenomenological Research* 56 (1996).

Carnap, R. "The Elimination of Metaphysics through the Logical Analysis of Language". In Ayer (ed.) *Logical Positivism*.

Carnap, R. "The Old and the New Logic". In Ayer (ed.) *Logical Positivism*.

Carnap, R. "Empiricism, Semantics, and Ontology". In Benacerraf and Putnam (eds) *The Philosophy of Mathematics*.

Carruthers, P. *Introducing Persons* (London: Croom Helm 1986).

Carruthers, P. *Tractarian Semantics* (Oxford: Blackwell 1989).

Chomsky, N. "Quine's Empirical Assumptions". In D. Davidson and J. Hintikka (eds) *Words and Objections* (Dordrecht: Reidel 1969).

Church, A. "Review of Ayer *Language, Truth and Logic*", *Journal of Symbolic Logic* 14 (1949).

Collins, J. *Chomsky: A Guide for the Perplexed* (London: Continuum 2008).

Crane, T. "All the Difference in the World", *Philosophical Quarterly* 41 (1991).

Crane, T. *The Mechanical Mind*, 2nd edition (London: Routledge 2003).

Crary, A. and Read, R. (eds) *The New Wittgenstein* (London: Routledge 2000).

Creath, R. "Logical Empiricism". In E. N. Zalta (ed.) *Stanford Encyclopedia of Philosophy* (2017). <https://plato.stanford.edu/archives/fall2017/entries/logical-empiricism/>.

Currie, G. *Frege: An Introduction to his Philosophy* (Brighton: Harvester Press 1982).

Davidson, D. *Essays on Actions and Events* (Oxford: Oxford University Press 1980).

Davidson, D. *Inquiries into Truth and Interpretation* (Oxford: Oxford University Press 1984).

Davidson, D. *Subjective, Objective, Intersubjective* (Oxford: Oxford University Press 1997).

Davidson, D. *Truth, Language and History* (Oxford: Oxford University Press 2005).

Davies, D. "How Sceptical is Kripke's 'Sceptical Solution'?", *Philosophia* 26 (1998).

Davies, M. *Meaning, Quantification, and Necessity* (London: Routledge Kegan-Paul 1981).

Davies, M. "Philosophy of Language". In E. James and N. Bunnin (eds) *The Blackwell Companion to Philosophy* (Oxford: Blackwell 1996).
De Gaynesford, M. *McDowell* (Cambridge: Polity Press 2004).
Devitt, M. *Designation* (New York: Columbia University Press 1981).
Devitt, M. "Dummett's Anti-Realism", *Journal of Philosophy* (1981).
Devitt, M. "Transcendentalism about Content", *Pacific Philosophical Quarterly* 71 (1990).
Devitt, M. *Realism and Truth*, 2nd edition (Princeton: Princeton University Press 1991).
Devitt, M. and Hanley, R. (eds) *The Blackwell Guide to the Philosophy of Language* (Oxford: Blackwell 2006).
Divers, J. *Possible Worlds* (London: Routledge 2002).
Donnellan, K. "Reference and Definite Descriptions", *Philosophical Review* 75 (1966).
Dummett, M. *Frege: Philosophy of Language* (London: Duckworth 1973).
Dummett, M. *Truth and Other Enigmas* (London: Duckworth 1978).
Dummett, M. "Wittgenstein's Philosophy of Mathematics". In *Truth and Other Enigmas* (London: Duckworth 1978).
Dummett, M. *The Interpretation of Frege's Philosophy* (London: Duckworth 1981).
Dummett, M. *The Logical Basis of Metaphysics* (Cambridge, MA: Harvard University Press 1991).
Dummett, M. *The Nature and Future of Philosophy* (New York: Columbia University Press 2010).
Dummett, M. *The Seas of Language* (Oxford: Oxford University Press 1993).
Ebbs, G. *Rule-Following and Realism* (Cambridge, MA: Harvard University Press 1997).
Edwards, J. "Best Opinion and Intentional States", *Philosophical Quarterly* 42 (1992).
Edwards, J. "Secondary Qualities and the A Priori", *Mind* 101 (1992).
Edwards, J. "Debates about Realism Transposed to a New Key", *Mind* 103 (1994).
Evans, G. "Semantic Theory and Tacit Knowledge". In S. Holtzmann and S. Leich (eds) *Wittgenstein: To Follow a Rule* (London: Routledge 1981).
Evans, G. *The Varieties of Reference* (Oxford: Oxford University Press 1982).
Evans, G. "The Causal Theory of Names". In *Collected Papers* (Oxford: Oxford University Press 1985).
Evans, G. "Identity and Predication". In *Collected Papers* (Oxford: Oxford University Press 1985).

Evnine, S. *Donald Davidson* (Stanford: Stanford University Press 1991).
Field, H. *Science Without Numbers* (Oxford: Blackwell 1980).
Field, H. "Tarski's Theory of Truth". In M. Platts (ed.) *Reference, Truth, and Reality* (London: Routledge Kegan-Paul 1980).
Fine, K. "The Question of Realism", *Philosophers' Imprint* 1 (2001).
Fodor, J. *The Language of Thought* (Hassocks: Harvester Press 1976).
Fodor, J. *Psychosemantics* (Cambridge, MA: MIT Press 1987).
Fodor, J. *A Theory of Content and Other Essays* (Cambridge, MA: MIT Press 1990).
Fodor, J. and LePore, E. "Analyticity Again". In Devitt and Hanley (eds) *The Blackwell Guide to the Philosophy of Language*.
Forbes, G. "Scepticism and Semantic Knowledge". In Miller and Wright (eds) *Rule-Following and Meaning*.
Foster, J. *Ayer*. (London: Routledge 1985).
Frege, G. *The Foundations of Arithmetic*. (Evanston: Northwestern Illinois Press 1953).
Frege, G. *Begriffsschrift* (selections). In Beaney (ed.) *The Frege Reader*.
Frege, G. "On Concept and Object". In Beaney (ed.) *The Frege Reader*.
Frege, G. "Function and Concept". In Beaney (ed.) *The Frege Reader*.
Frege, G. "Letter to Jourdain". In Beaney (ed.) *The Frege Reader*.
Frege, G. "On *Sinn* and *Bedeutung*". In Beaney (ed.) *The Frege Reader*.
Frege, G. "Thought". In Beaney (ed.) *The Frege Reader*.
Fricker, E. "Semantic Structure and Speaker's Understanding", *Proceedings of the Aristotelian Society* 83 (1982).
Geach, P. and Black, M. (eds) *Translations from the Philosophical Writings of Gottlob Frege* (Oxford: Oxford University Press 1960).
Gibbard, A. *Wise Choices, Apt Feelings* (Cambridge, MA: Harvard University Press 1990).
Gibbard, A. *Meaning and Normativity* (Oxford: Oxford University Press 2012).
Gibbard, A. *Thinking How to Live* (Cambridge, MA: Harvard University Press 2003).
Gibson, R. (ed.) *The Cambridge Companion to Quine* (Cambridge: Cambridge University Press 2004).
Ginet, C. "The Dispositionalist Solution to Wittgenstein's Problem about Understanding a Rule". *Midwest Studies in Philosophy* 17 (1992).
Ginsborg, H. "Inside and Outside Language: Stroud's Non-Reductionism about Meaning". In J. Bridges, N. Kolodny, and W. Wong (eds) *The Possibility of Philosophical Understanding: Reflections on the Thought of Barry Stroud* (Oxford: Oxford University Press 2011).

Ginsborg, H. "Primitive Normativity and Skepticism About Rules", *Journal of Philosophy* 108 (2011).

Ginsborg, H. "Meaning, Understanding and Normativity", *Proceedings of the Aristotelian Society* Supplementary Volume 86 (2012).

Gluer, K. *Donald Davidson: A Short Introduction* (Oxford: Oxford University Press 2012).

Goldfarb, W. "Kripke on Wittgenstein on Rules". In Miller and Wright (eds) *Rule-Following and Meaning*.

Gower, B. (ed.) *Logical Positivism in Perspective* (London: Croom Helm 1987).

Grayling, A. *Russell: A Very Short Introduction* (Oxford: Oxford University Press 2002).

Grice, P. "Meaning". In *Studies in the Ways of Words*.

Grice, P. *Studies in the Ways of Words* (Cambridge, MA: Harvard University Press 1989).

Grice, P. and Strawson, P. "In Defense of a Dogma", *Philosophical Review* 65 (1956).

Haack, S. *Philosophy of Logics* (Cambridge: Cambridge University Press 1978).

Hahn, H. "Logic, Mathematics, and Knowledge of Nature". In Ayer (ed.) *Logical Positivism*.

Hahn, L. (ed.) *The Philosophy of A. J. Ayer* (Chicago: Open Court 1992).

Hahn, L. (ed.) *The Philosophy of Donald Davidson* (Chicago: Open Court 1999).

Hahn, L. and Schilpp, P. (eds) *The Philosophy of W. V. Quine* (Chicago: Open Court 1986).

Hale, B. "Can There Be a Logic of Attitudes?". In J. Haldane and C. Wright (eds) *Reality, Representation, and Projection* (Oxford: Oxford University Press 1993).

Hale, B. "Realism and Its Oppositions". In Hale, Wright, and Miller (eds) *A Companion to the Philosophy of Language*.

Hale, B., Wright, C. and Miller, A. (eds) *A Companion to the Philosophy of Language*, 2nd edition (Oxford: Blackwell 2017).

Hanfling, O. *Wittgenstein's Later Philosophy* (London: Macmillan 1989).

Harman, G. "Quine on Meaning and Existence (I)", *Review of Metaphysics* 21 (1967).

Hattiangadi, A. "Is Meaning Normative?", *Mind and Language* 21 (2006).

Hattiangadi, A. *Oughts and Thoughts: Scepticism and the Normativity of Content* (Oxford: Oxford University Press 2007).

Hattiangadi, A. "The Normativity of Meaning". In Hale, Wright, and Miller (eds) *A Companion to the Philosophy of Language*.

Haukioja, J. Rule-Following, Response-Dependence and Realism. PhD thesis, University of Turku (2000).

Haukioja, J. "Internalism and Externalism". In Hale, Wright, and Miller (eds) *A Companion to the Philosophy of Language*.

Heal, J. *Fact and Meaning: Quine and Wittgenstein on the Philosophy of Language* (Oxford: Blackwell 1990).

Heil, J. and Martin, C. "Rules and Powers". In J. Tomberlin (ed.) *Philosophical Perspectives: Language, Mind and Ontology* (Oxford: Blackwell 1998).

Higginbotham, J. "Truth and Reference as the Basis of Meaning". In Devitt and Hanley (eds) *The Blackwell Guide to the Philosophy of Language*.

Hintikka, M. and Vermazen, B. (eds) *Essays on Davidson: Actions and Events* (Oxford: Clarendon Press 1985).

Hodges, W. *Logic* (London: Penguin Books 1991).

Holton, R. "Intention Detecting", *Philosophical Quarterly* 43 (1993).

Hookway, C. *Quine* (London: Polity Press 1987).

Horwich, P. "The Essence of Expressivism", *Analysis* (1994).

Horwich, P. *Meaning* (Oxford: Oxford University Press 1998).

Horwich, P. "Meaning, Use and Truth". In Miller and Wright (eds) *Rule-Following and Meaning*.

Horwich, P. *Wittgenstein's Metaphilosophy* (Oxford: Oxford University Press 2012).

Hossein-Khani, A. Kripke's Wittgenstein's Sceptical Solution and Donald Davidson's Philosophy of Language. PhD thesis , University of Otago (2017).

Hume, D. *An Enquiry Concerning Human Understanding* (many editions, originally published 1748).

Hylton, P. *Quine* (Abingdon: Routledge 2007).

Jackson, F. and Pettit, P. "Functionalism and Broad Content", *Mind* 97 (1988).

Jeffrey, R. *Formal Logic: Its Scope and Limits* (New York: McGraw Hill 1991).

Johnston, M. "Objectivity Disfigured". In J. Haldane and C. Wright (eds) *Reality, Representation, and Projection* (Oxford: Oxford University Press 1993).

Joseph, M. *Donald Davidson* (Chesham: Acumen 2003).

Kallestrup, J. *Semantic Externalism* (Abingdon: Routledge 2012).

Kant, I. *Critique of Pure Reason* (many editions, originally published 1781).

Kemp, G. *Quine: A Guide for the Perplexed* (London: Continuum 2006).

Kenny, A. *Wittgenstein* (London: Penguin Books 1973).

Kenny, A. *Frege*. (London: Penguin Books 1995).

Kirk, R. *Translation Determined* (Oxford: Oxford University Press 1986).
Kirk, R. *Relativism and Reality* (London: Routledge 1999).
Kripke, S. *Naming and Necessity*. (Oxford: Oxford University Press 1980).
Kripke, S. *Wittgenstein on Rules and Private Language* (Oxford: Oxford University Press 1982).
Kusch, M. "Rule-Following", *Oxford Bibliographies Online* (2010), www.oxfordbibliographies.com/view/document/obo-9780195396577/obo-9780195396577-0106.xml.
Kusch, M. *A Sceptical Guide to Meaning and Rules: A Defence of Kripke's Wittgenstein* (Chesham: Acumen 2006).
Lemmon, E. J. *Beginning Logic*. (London: Nelsons University Paperbacks 1964).
Lepore, E. (ed.) *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson* (Oxford: Blackwell 1982).
Lepore, E. and Ludwig, K. *Donald Davidson: Meaning, Truth, Language and Reality* (Oxford: Oxford University Press 2005).
Lepore, E. and Ludwig, K. *Donald Davidson's Truth-Theoretic Semantics* (Oxford: Clarendon Press 2007).
Lepore, E. and Ludwig, K. (eds) *A Companion to Donald Davidson* (Oxford: Wiley-Blackwell 2013).
Lepore, E. and McLaughlin, B. (eds) *Actions and Events: Perspectives on the Philosophy of Donald Davidson* (Oxford: Blackwell 1985).
LePore, E. and Smith, B. (eds) *The Oxford Handbook of the Philosophy of Language* (Oxford: Oxford University Press 2006).
Lewis, D. *Convention*. (Cambridge, MA: Harvard University Press 1969).
Lewis, D. "How to Define Theoretical Terms", *Journal of Philosophy* 67 (1970).
Lewis, D. "Psychophysical and Theoretical Identifications", *Australasian Journal of Philosophy* 50 (1972).
Lewis, D. "Radical Interpretation". In *Collected Papers*, Volume 1 (Oxford: Oxford University Press 1983).
Lewis, D. "Statements Partly about Observation", *Philosophical Papers* 17 (1988).
Locke, J. *An Essay Concerning Human Understanding*, edited by P. Nidditch (Oxford: Oxford University Press 1975).
Longworth, G. "Semantics and Pragmatics". In Hale, Wright, and Miller (eds) *A Companion to the Philosophy of Language*.
Lowe, E. J. *Locke on Human Understanding* (London: Routledge 1995).
Lowe, E. J. *Locke* (London: Routledge 2005).
Ludlow, P. and Neale, S. "Descriptions". In Devitt and Hanley (eds) *The Blackwell Guide to Philosophy of Language*.

Ludwig, K (ed.) *Donald Davidson* (New York: Cambridge University Press 2003).
Lycan, W. "Names". In Devitt and Hanley (eds) *The Blackwell Guide to Philosophy of Language*.
Lyons, J. *Chomsky* (London: Penguin Books 1970).
Mackie, J. *Ethics: Inventing Right and Wrong* (London: Penguin 1978)
Makin, G. *The Metaphysicians of Meaning: Russell and Frege on Sense and Denotation* (London: Routledge 2000).
Malcolm, N. *Nothing is Hidden* (Oxford: Blackwell 1986).
McCulloch, G. *The Game of the Name* (Oxford: Oxford University Press 1989).
McCulloch, G. *The Life of the Mind* (London: Routledge 2003).
McCulloch, G. *The Mind and Its World*. (London: Routledge 1995).
McCulloch, G. "The Spirit of Twin-Earth", *Analysis* 52 (1992).
McDowell, J. *Mind and World* (Cambridge, MA: Harvard University Press 1994).
McDowell, J. "Anti-Realism and the Epistemology of Understanding". In *Meaning, Knowledge and Reality*.
McDowell, J. "Intentionality and Interiority in Wittgenstein". In *Meaning, Knowledge and Reality*.
McDowell, J. "Meaning and Intention in Wittgenstein's Later Philosophy". In his *Meaning, Knowledge and Reality*.
McDowell, J. *Meaning, Knowledge and Reality* (Cambridge, MA: Harvard University Press 1998).
McDowell, J. "Physicalism and Primitive Denotation". In *Meaning, Knowledge and Reality*.
McDowell, J. "Putnam on Mind and Meaning". In *Meaning, Knowledge and Reality*.
McDowell, J. "On 'The Reality of the Past'". In *Meaning, Knowledge and Reality*.
McDowell, J. "On the Sense and Reference of a Proper Name". In *Meaning, Knowledge and Reality*.
McDowell, J. "Truth-Conditions, Bivalence, and Verificationism". In *Meaning, Knowledge and Reality*.
McDowell, J. *Mind, Value and Reality* (Cambridge, MA: Harvard University Press 1998).
McDowell, J. "Projectivism and Truth in Ethics". In *Mind, Value and Reality*.
McDowell, J. "Reply to Crispin Wright". In C. McDonald, B. Smith, and C. Wright (eds) *Self-Knowledge and Externalism* (Oxford: Oxford University Press 1998).

McDowell, J. "Wittgenstein on Following a Rule". Reprinted in Miller and Wright (eds) *Rule-Following and Meaning*.

McDowell, J. "Are Meaning, Understanding, etc. Definite States?". In *The Engaged Intellect*.

McDowell, J. *The Engaged Intellect* (Cambridge, MA: Harvard University Press 2009).

McDowell, J. "How Not to Read *Philosophical Investigations*: Brandom's Wittgenstein". In *The Engaged Intellect*.

McDowell, J. "Wittgensteinian Quietism", *Common Knowledge* 15 (2009).

McDowell, J. and Pettit, P. (eds) *Subject, Thought and Context* (Oxford: Oxford University Press 1986).

McGinn, C. "Semantics for Non-Indicative Sentences", *Philosophical Studies* 32 (1977).

McGinn, C. "The Structure of Content". In A. Woodfield (ed.) *Thought and Object* (Oxford: Oxford University Press 1982).

McGinn, C. *Wittgenstein on Meaning* (Oxford: Basil Blackwell 1984).

McManus, D. "Boghossian, Miller and Lewis on Dispositional Theories of Meaning", *Mind and Language* 15 (2000).

Melia, J. *Modality* (Chesham: Acumen 2003).

Millar, A. "The Normativity of Meaning". In A. O'Hear (ed.) *Logic, Thought and Language* (Cambridge: Cambridge University Press 2002).

Miller, A. "An Objection to Wright's Treatment of Intention", *Analysis* 49 (1989).

Miller, A. "Review Article on Lowe's *Locke on Human Understanding*", *Locke Newsletter* 26 (1995).

Miller, A. "Emotivism and the Verification Principle", *Proceedings of the Aristotelian Society* 98 (1998).

Miller, A. "Rule-Following, Response-Dependence, and McDowell's Debate with Anti-Realism", *European Review of Philosophy* 3 (1998).

Miller, A. "Horwich, Meaning and Kripke's Wittgenstein", *Philosophical Quarterly* 50 (2000).

Miller, A. "What is the Manifestation Argument?", *Pacific Philosophical Quarterly* 83 (2002).

Miller, A. "Does 'Belief Holism' Show that Reductive Dispositionalism about Content Could Not Be True?", *Proceedings of the Aristotelian Society* Supplementary Volume 77 (2003).

Miller, A. "What is the Acquisition Argument?". In A. Barber (ed.) *Epistemology of Language* (Oxford: Oxford University Press 2003).

Miller, A. "Rule-Following and Externalism", *Philosophy and Phenomenological Research* (2004).

Miller, A. "Meaning Scepticism". In Devitt and Hanley (eds) *The Blackwell Guide to the Philosophy of Language*.

Miller, A. "Realism and Antirealism". In LePore and Smith (eds) *The Oxford Handbook of Philosophy of Language*.

Miller, A. "Another Objection to Wright's Treatment of Intention", *Analysis* 67 (2007).

Miller, A. "Hume: Necessary Connection and Distinct Existence". In R. Cameron, A. McGonigal, R. Le Poidevin, and P. Simons (eds) *The Routledge Companion to Metaphysics* (London: Routledge 2009).

Miller, A. "The Argument From Queerness and the Normativity of Meaning". In M. Grajner and A. Rami (eds) *Truth, Existence and Realism* (Frankfurt: Ontos Verlag 2010).

Miller, A. "Kripke's Wittgenstein, Factualism and Meaning". In D. Whiting (ed.) *The Later Wittgenstein on Meaning* (Basingstoke: Palgrave 2010).

Miller, A. "Primary Qualities, Secondary Qualities, and the Truth about Intention", *Synthese* 171 (2010).

Miller, A. "Rule-Following Skepticism". In S. Bernecker and D. Prichard (eds) *The Routledge Companion to Epistemology* (London: Routledge 2011).

Miller, A. "Semantic Realism and Motivational Internalism". In R. Schantz (ed.) *Prospects for Meaning* (Berlin: De Gruyter 2012).

Miller, A. *Contemporary Metaethics: An Introduction*, 2nd edition (Oxford: Polity Press 2013).

Miller, A. "Postscript to Jane Heal's 'Radical Interpretation'". In Hale, Wright, and Miller (eds) *A Companion to the Philosophy of Language*, 2nd edition.

Miller, A. "Tacit Knowledge". In Hale, Wright, and Miller (eds) *A Companion to the Philosophy of Language*.

Miller, A. "Rule-Following, Meaning and Primitive Normativity" *Mind* (2017), https://doi.org/10.1093/mind/fzx033.

Miller, A. and Divers, J. "Best Opinion, Intention Detecting, and Analytic Functionalism", *Philosophical Quarterly* 44 (1994).

Miller, A. and Divers, J. "Why the Expressivist about Value Should Not Love Minimalism about Truth", *Analysis* 54 (1994).

Miller, A. and Wright, C. (eds) *Rule-Following and Meaning* (Chesham: Acumen 2002).

Milne, P. "Tarski on Truth and Its Definition". In P. Kolar et al. (eds) *Logica '96* (Prague: Filosofia 1997).

Monk, R. *Ludwig Wittgenstein: The Duty of Genius* (New York: Penguin Books 1990).

Monk, R. *Bertrand Russell: The Spirit of Solitude* (London: Vintage 1997).

Monk, R. *Bertrand Russell: The Ghost of Madness* (London: Vintage 2001).

Morris, M. *Routledge Philosophy Guidebook to Wittgenstein and the Tractatus* (London: Routledge 2008).

Morris, M. "Metaphysics, Philosophy, and the Philosophy of Language". In Hale, Wright, and Miller (eds) *A Companion to the Philosophy of Language*.

Mumford, S. *Dispositions* (Oxford: Oxford University Press 1998).

Myers, R. and Verheggen, C. *Donald Davidson's Triangulation Argument: A Philosophical Inquiry* (London: Routledge 2016).

Newton-Smith, W. H. *Logic* (London: Routledge 1985).

Noonan, H. *Frege* (Cambridge: Polity Press 2000).

Noonan, H. *Routledge Philosophy Guidebook to Kripke on Naming and Necessity* (London: Routledge 2013).

Orenstein, A. *W. V. Quine* (Chesham: Acumen 2002).

Pears, D. *The False Prison: A Study in the Development of Wittgenstein's Thought*, Volume 1 (Oxford: Oxford University Press 1987).

Pessin, A. and Goldberg, S. *The Twin-Earth Chronicles* (New York: Sharpe 1996).

Pettit, P. "Affirming the Reality of Rule-Following", *Mind* 99 (1990).

Pettit, P. "The Reality of Rule-Following", *Mind* 99 (1990).

Pettit, P. "Realism and Response-Dependence", *Mind* 100 (1991).

Pettit, P. *The Common Mind*, 2nd edition (Oxford: Oxford University Press 1996).

Pettit, P. "A Theory of Normal and Ideal Conditions", *Philosophical Studies* 96 (1999).

Phillips-Griffiths, A. *A. J. Ayer: Memorial Essays* (Cambridge: Cambridge University Press 1990).

Platts, M. *Ways of Meaning* (London: Routledge Kegan-Paul 1979).

Putnam, H. "The Analytic and the Synthetic". In *Mind, Language, and Reality* (Cambridge: Cambridge University Press 1975).

Putnam, H. "The Meaning of 'Meaning'". In *Mind, Language, and Reality* (Cambridge: Cambridge University Press 1975).

Putnam, H. *Realism, Truth, and History* (Cambridge: Cambridge University Press 1981).

Quine, W. V. "Two Dogmas of Empiricism". In *From a Logical Point of View* (Cambridge, MA: Harvard University Press 1953).

Quine, W. V. *Word and Object* (Cambridge, MA: MIT Press 1960).

Quine, W. V. "On the Reasons for the Indeterminacy of Translation", *Journal of Philosophy* 67 (1970).

Quine, W. V. *Philosophy of Logic* (Englewood Cliffs: Prentice Hall 1970).

Quine, W. V. *The Pursuit of Truth* (Cambridge, MA: Harvard University Press 1990).
Quine, W. V. "Carnap on Logical Truth". In Benacerraf and Putnam (eds) *The Philosophy of Mathematics*.
Quine, W. V. "Truth By Convention". In Benacerraf and Putnam (eds) *The Philosophy of Mathematics*.
Ramberg, B. *Donald Davidson's Philosophy of Language* (Oxford: Blackwell 1989).
Read, R. and Lavery, M. (eds) *Beyond the Tractatus Wars: The New Wittgenstein Debate* (Abingdon: Routledge 2011).
Read, S. *Thinking About Logic* (Oxford: Oxford University Press 1994).
Rogers, B. *A. J. Ayer* (London: Chatto and Windus 1999).
Russell, B. *Introduction to Mathematical Philosophy* (London: Allen and Unwin 1919).
Russell, B. "On Denoting". In *Logic and Knowledge*.
Russell, B. "Logical Positivism". In his *Logic and Knowledge*.
Russell, B. *Logic and Knowledge*, edited by R. Marsh (London: Routledge 1956).
Russell, B. "The Philosophy of Logical Atomism". In *Logic and Knowledge*.
Russell, B. *The Problems of Philosophy*. (Oxford: Oxford University Press 1978).
Russell, B. and Whitehead, A. N. *Principia Mathematica*, 3 vols (Cambridge: Cambridge University Press 1910–1913).
Russell, G. *Truth in Virtue of Meaning: A Defence of the Analytic/Synthetic Distinction* (Oxford: Oxford University Press 2008).
Sainsbury, M. "Frege and Russell". In E. James and N. Bunnin (eds) *The Blackwell Companion to Philosophy* (Oxford: Blackwell 1996).
Sainsbury, M. "Philosophical Logic". In A. Grayling (ed.) *Philosophy: A Guide through the Subject* (Oxford: Oxford University Press 1995).
Schiffer, S. "Convention and Intention in the Theory of Meaning". In Hale, Wright, and Miller (eds) *A Companion to the Philosophy of Language*.
Schiffer, S. *Meaning* (Oxford: Clarendon Press 1972).
Schiffer, S. *Remnants of Meaning* (Cambridge, MA: MIT Press 1987).
Schlick, M. "Positivism and Realism". In Ayer (ed.) *Logical Positivism*.
Schlick, M. "The Turning Point in Philosophy". In Ayer (ed.) *Logical Positivism*.
Schroeder, S. *Wittgenstein: The Way out of the Fly-bottle* (Oxford: Polity Press 2006).

Searle, J. "Proper Names", *Mind* 67 (1958).
Searle, J. "What Is a Speech-Act?" In M. Black (ed.) *Philosophy in America* (London: Allen and Unwin 1965).
Searle, J. *Speech-Acts* (Cambridge: Cambridge University Press 1969).
Searle, J. *Intentionality* (Cambridge: Cambridge University Press 1983).
Segal, G. *A Slim Book About Narrow Content* (Cambridge, MA: MIT Press 2000).
Segal, G. "Truth and Meaning". In LePore and Smith (eds) *The Oxford Handbook of Philosophy of Language*.
Shieh, S. "Undecidability in Antirealism", *Philosophia Mathematica* 6 (1998).
Smith, M. *The Moral Problem* (Oxford: Blackwell 1994).
Smith, M. "Why Expressivists about Value Should Love Minimalism about Truth", *Analysis* 54 (1994).
Soames, S. *Philosophical Analysis in the Twentieth Century*, Volume 1: *The Dawn of Analysis* (Princeton: Princeton University Press 2003).
Strawson, P. "On Referring", *Mind* 59 (1950).
Strawson, P. (ed.) *Philosophical Logic* (Oxford: Oxford University Press 1968).
Strawson, P. "Meaning and Truth". In his *Logico-Linguistic Papers* (London: Methuen 1971).
Sullivan, P. "Problems for a Construction of Meaning and Intention", *Mind* 103 (1994).
Summerfield, D. M. "On Taking the Rabbit of Rule-Following out of the Hat of Representation: A Response to Pettit's 'The Reality of Rule-Following'", *Mind* 99 (1990).
Tarski, A. "The Semantic Conception of Truth", *Philosophy and Phenomenological Research* 4 (1949).
Tarski, A. "Truth and Proof", *Scientific American* (June 1969).
Tarski, A. "The Concept of Truth in Formalized Languages". In his *Logic, Semantics, Metamathematics* (Indianapolis, IN: Hackett Publishers 1983).
Tennant, N. *The Taming of the True* (Oxford: Oxford University Press 1997).
Textor, M. *Routledge Philosophy Guidebook to Frege on Sense and Reference* (Abingdon: Routledge 2011).
Thornton, T. *John McDowell* (Chesham: Acumen 2004).
Thornton, T. *Wittgenstein on Language and Thought* (Edinburgh: Edinburgh University Press 1998).
Tomassi, P. *Logic* (London: Routledge 1999).

Travis, C. "Pragmatics". In Hale, Wright, and Miller (eds) *A Companion to the Philosophy of Language*.

Uebel, T. "Vienna Circle", *Stanford Encyclopedia of Philosophy* (2016). https://plato.stanford.edu/entries/vienna-circle/

Verheggen, C. "Davidson's Treatment of Wittgenstein's Rule-Following Paradox". In C. Verheggen (ed.) *Wittgenstein and Davidson on Thought and Action* (Cambridge: Cambridge University Press).

Wee, D. Wittgenstein, Objectivity and Rule-Following: Towards Resolving the Communitarianism Versus Individualism Debate. PhD thesis, University of Otago (2016).

Weiner, J. *Frege* (Oxford: Oxford University Press 1999).

Weiss, B. *Michael Dummett* (Chesham: Acumen 2002).

White, R. *Wittgenstein's Tractatus Logico-Philosophicus: A Reader's Guide* (London: Continuum 2007).

Whiting, D. "The Normativity of Meaning Defended", *Analysis* (2007).

Whiting, D. "Is Meaning Fraught with Ought?", *Pacific Philosophical Quarterly* 90 (2009).

Whiting, D. "What is the Normativity of Meaning?", *Inquiry* 59 (2013).

Wiggins, D. "Meaning and Truth Conditions: From Frege's Grand Design to Davidson's". In Hale, Wright, and Miller (eds) *A Companion to the Philosophy of Language*.

Wilson, G. "Semantic Realism and Kripke's Wittgenstein", *Philosophy and Phenomenological Research* 58 (1998).

Wilson, G. "Kripke on Wittgenstein on Normativity". Reprinted in Miller and Wright (eds) *Rule-Following and Meaning*.

Wilson, G. "Rule-Following, Meaning and Normativity". In LePore and Smith (eds) *The Oxford Handbook of Philosophy of Language*.

Wilson, G. "On the Skepticism about Rule-Following in Kripke's Version of Wittgenstein". In A. Berger (ed.) *Saul Kripke* (Cambridge: Cambridge University Press 2011).

Wittgenstein, L. *Remarks on the Foundations of Mathematics* (Oxford: Basil Blackwell 1956).

Wittgenstein, L. *Tractatus Logico-Philosophicus* (London: RKP 1961).

Wittgenstein, L. *Philosophical Investigations* (Oxford: Basil Blackwell 1974).

Woodfield, A. (ed.) *Thought and Object* (Oxford: Oxford University Press 1982).

Wright, C. *Frege's Conception of Numbers as Objects* (Aberdeen: Aberdeen University Press 1983).

Wright, C. "Inventing Logical Necessity". In J. Butterfield (ed.) *Language, Mind, and Logic* (Cambridge: Cambridge University Press 1986).

Wright, C. "Scientific Realism, Observation, and the Verification Principle". In C. Wright and G. McDonald (eds) *Fact, Science, and Morality* (Oxford: Blackwell 1986).

Wright, C. *Realism, Meaning, and Truth*, 2nd edition (Oxford: Blackwell 1993).

Wright, C. "Theories of Meaning and Speakers' Knowledge". In *Realism, Meaning, and Truth*.

Wright, C. "Wittgenstein's Rule-Following Considerations and the Central Project of Theoretical Linguistics". In *Rails to Infinity*.

Wright, C. "Critical Notice of McDowell's *Mind and World*", *European Journal of Philosophy* 4 (1996).

Wright, C. "Human Nature?", *European Journal of Philosophy* 4 (1996).

Wright, C. "Self-Knowledge: The Wittgensteinian Legacy". In C. McDonald, B. Smith, and C. Wright (eds) *Knowing Our Own Minds* (Oxford: Oxford University Press 1998).

Wright, C. "On Making Up One's Mind: Wittgenstein on Intention". In *Rails to Infinity*.

Wright, C. "The Problem of Self-Knowledge (II)". In *Rails to Infinity*.

Wright, C. "Does PI 258–260 Suggest a Cogent Argument against Private Language?" In *Rails to Infinity*.

Wright, C. "Kripke's Account of the Argument Against Private Language". In *Rails to Infinity*.

Wright, C. *Rails to Infinity* (Cambridge, MA: Harvard University Press 2001).

Wright, C. "Study Note on Wittgenstein on the Nature of Philosophy and its Proper Method". In *Rails to Infinity*.

Wright, C. "Critical Notice of McGinn's *Wittgenstein on Meaning*". In Miller and Wright (eds) *Rule-Following and Meaning*.

Wright, C. "Comrades against Quietism". In *Saving the Differences*.

Wright, C. "Moral Values, Projection, and Secondary Qualities". In *Saving the Differences*.

Wright, C. *Saving the Differences* (Cambridge, MA: Harvard University Press 2003).

Wright, C. *Truth and Objectivity* (Cambridge, MA: Harvard University Press 1992).

Wright, C. "The Verification Principle: Another Puncture, Another Patch". In *Realism, Meaning, and Truth*.

Wright, C. "Indeterminacy of Translation". In Hale, Wright, and Miller (eds) *A Companion to the Philosophy of Language*.

Yolton, J. *Perceptual Acquaintance From Descartes to Reid* (Oxford: Blackwell 1984).

Zalabardo, J. "A Problem for Information Theoretic Semantics", *Synthese* 105 (1995).
Zalabardo, J. "Rules, Communities, Judgements", *Critica* 63 (1989).
Zeglen, U. (ed.) *Donald Davidson: Truth, Meaning and Knowledge* (London: Routledge 1999).
Ziff, P. "On H. P. Grice's Account of Meaning", *Analysis* 28 (1967).

Index

a posteriori 68–69, 114, 116, 117, 130, 139, 146, 399
a priori 116, 117, 128–137
acquisition challenge 390-391
adequacy conditions *see* Davidson
analysis 57–59
analytic priority 329–330
analytic/synthetic distinction 154, 155–161
analytic truth 66, 117, 156, 157
analyticity 3–4, 128–137, 146, 156–157, 165–167
anti-realism 6, 388–399, 406–11
assertability 211–212, 221–222, 231–235; *see also* warranted assertability
associationism 48–55, 240 n. 10
assymetric dependency *see* Fodor
Austin, J.L. 318, 341
axiom 26–27
Ayer, A.J. 3, 112–153, 155, 165; emotivism 142–145; holism 145–146, 165; necessity 128–137; verifiability 117, 119–122; verification principle 117–123, 143, 144, 155

Bach, K. 341
Baldwin, T. 32, n. 6
Beaney, M. 106 n. 26, 148
bearerless names 35–36, 41–42, 63–65, 77–85
Bedeutung 3, 18, 32 n. 7, 33 n. 9
Beebee, H. 241 n. 19
Begriffsschrift 32 n. 2
belief contexts 36–37, 43–44, 59–63, 85–88; and indirect reference 43–44, 59–63; and indirect sense 60–63

Bell, D. 107, n. 39
Berkley, G. 112, 310 n. 45
Blackburn, S. 238, 285–288, 306, 315 n. 111, 334–337, 341, 379, 415, 418 n. 27
Boghossian, P. 194, 196 n. 16, 212–217, 224–226, 237, 238, 239, 241 n. 18, 250–256, 258–260, 266–268, 277–278, 299, 305, 306, 309–310 n. 23; sophisticated dispositionalism 251–256, 258–260, 299
Boole, G. 32, n. 2
Borg, E. 341
Brock, S. 415
Burge, T. 419 n. 30
Byrne, A. 226, 239

Carnap, R. 112, 116, 133–134, 137–142, 148, 152 n. 51, 155, 164; external questions 116, 137–142; internal questions 116, 137–142; linguistic framework 138–142, 152 n. 51, 155
Carruthers, P. 106 n. 30, 198 n. 48
Chomsky, N. 188–189, 194, 380 n. 7
Church, A. 122
Cognitive Command 410
Cognitivism (about meaning) 223
Collins, J. 380 n. 7
compositional semantic value *see* principle of compositionality
compositionality 18, 26, 34–35, 36–37, 44–45, 59, 99, 100, 331–332, 334–336, 339–340, 347, 356, 372, 376
concept script *see Begriffsschrift*
concepts 23–26, 80, 89, 99, 161–164
conditional 9–10, 385–386
conjunction 24
constitutive scepticism 127, 173–174, 199–200, 239 n. 3, 268
contingent *a priori* truth 68–71
contingent truth 66
Convention (T) 354–356, 364–365, 372, 374–376
Crane, T. 264–265, 306, 308 n. 1, 309 n. 19, 406, 419 n. 33
Creath, R. 148, 195
Currie, G. 32, n. 2, 103

Davidson, D. 5, 163, 344–380, 388–389 adequacy conditions 346–348, 351–356, 366–372, 376: extensional adequacy 347–348, 351–356, 366–367, 370–372, 376: holism 371, 374–376; indeterminacy of interpretation 372–373; principle of charity 368–372; radical interpretation 367–372; translation 363–364, 366
Davies, D. 226, 234, 239, 245 n. 54, 249
Davies, M. 329, 341
De Gaynesford, M. 307
deflationary conception of truth 213–214
demonstratives 81–82
descriptions 55–57, 67–71, 77–98, 101, 104, 134
descriptive law 54–56; *see also* normative law; prescriptive law
Devitt, M. 103–104, 238, 416, 418 n. 27
disjunction problem 262–263
dispositionalism 206–208, 209–210, 247, 251–260, 265–266, 267, 272–273, 280–281, 282–285, 297–298, 299, 301; sophisticated 251–256, 258, 265–266, 272–273, 280–281, 297–298,

299, 301; *see also* partial reductionism
Divers, J. 104, 306
Donnellan, K. 95–96, 342–343 n. 31
Duhem/Quine thesis *see* underdetermination
Dummett, M. 6, 46–48, 61–63, 71, 73, 74, 103, 108 n. 56, 210–211, 249, 389–396, 397–399, 406–407 409, 411, 415, 416; arguments against verification-transcendent truth conditions 389–396

Ebbs, G. 307
Edwards, J. 306, 312 n. 53
effective decidability 396
elementary propositions 114
emotivism 116, 142–145, 286
epistemological priority 329–330
epistemological scepticism 127, 174, 222
error-theory 421 n. 44
ethocentrism 248, 278–288
Evans, G. 64–65, 72, 103, 110 n. 71, 177–185, 194, 332–333, 341, 342 n. 23, 415
Evnine, S. 379
existential quantification 13–14, 25–26
extension 21–26, 33 n. 15
extensional adequacy *see* Davidson
extensional theories of meaning 348–351
external questions *see* Carnap
externalism 400–406, 415

factualism 226–237, 239
Feigl, H. 112
Field, H. 381 n. 31, 421 n. 44
Fine, K. 416
Fodor, J. 194, 247, 260–265, 306, 309 nn. 19, 22, 403–404; asymmetric dependency 263–265; mentalese 309 nn. 19, 22
Forbes, G. 305
force 75, 76
formal theory of meaning xi–xii, 33 n. 17, 344
Foster, J. 123–128, 149
Frege, G. xi, 2, 3, 8–30, 32 nn. 2, 7, 33 n. 9, 34, 36–44, 48–56, 57–63, 73–74, 75, 76, 99–100, 102–103, 104 n. 1, 107 n. 40, 157–160, 161–162, 178, 195 n. 15, 241 n. 18, 317, 318, 344–346, 349, 355–356, 376–377, 389; *Bedeutung* 3, 18, 32, n. 7, 33, n. 9; *Begriffsschrift* 32, n. 2; belief contexts 36–37, 43–44, 59–63; concept 23, 24; concept script 32 n. 2; the False 20, 346; force 75, 76; function 21–26, 346; indirect reference 43–44, 60–61, 100; indirect sense 60–61; informativeness 37–39, 40–41, 57–59; normative law 54–55; normativity of sense 54–55; object 19–20, 21–22; objectivity of sense 48–54, 55–57; predicate logic 11–16; prescriptive law 54–56; proper names 19–20, 22, 107 n. 40; reference 3, 33 n.9; semantic value 3, 18–21, 99–100, 346; sense 3, 34, 48–54, 99–100, 178, 195 n.15, 241 n. 18, 344, 355–356; *Sinn* 3, 104 n. 1; syntactical categories 16; tone, 2, 73–74, 76, 100; the True 20, 346
Frege–Geach problem 286, 288, 313 n. 73, 384–386, 415, 417 n. 9
Fricker, E. 336–337

function/functional expression 21–26, 33 nn. 14, 15, 37, 80, 86–87, 209, 346; first-level, definition of 24; second-level, definition of 24; *see also* truth-function/truth-functional connective

Gibbard, A. 243 n. 39, 415
Ginet, C. 305
Ginsborg, H. 248, 297–304, 307
Gluer, K. 379
Gödel, K. 112
Goldfarb, W. 238
grammar *see* syntax
grammatical form 79, 81, 113
Grayling, A. 104
Grice, P. 5, 161, 163–164, 193, 249, 318, 319–341; indicator meaning 319–320; natural meaning 319–320; non-natural meaning, definition of 319–320

Haack, S. 379
Hahn, H. 112, 116, 129, 131, 132–133, 148
Hale, B. 243 n. 39, 415, 416
Hanfling, O. 307
Harman, G. 193–194
Hattiangandi, A. 239
Haukioja, J. 226, 239, 415
Heal, J. 216–217, 219
Heil, J. 305
Higginbotham, J. 379
Hodges, W. 31
holism 145–146, 165–169, 186, 371, 374–376
Holton, R. 306
Hookway, C. 141–142, 188, 189, 194, 197 n. 29, 372, 375, 379
Horwich, P. 306, 420 n. 42
Hossein-Khani, A. 381 n. 33
Hume, D. 112, 241 n. 19
Hylton, P. 195, 198 n. 47

illocutionary effect 325–328, 340
imperatives 75, 118–119, 144, 250, 377
indeterminacy of interpretation *see* Davidson
indeterminacy of translation 154, 169–192, 372
indexical expressions 367
indicator meaning *see* Grice
indirect reference 43–44, 60–61, 83, 100; and belief contexts 83
indirect sense 60–63
informal theory of meaning xi, 33 n. 17
informativeness 37–39, 40–41, 57–59, 88–90; paradox of analysis 57–59
inscrutability of reference 197–198 n. 31
intensional contexts 105 n. 3, 350–351
intensional theories of meaning 348, 350–351
intention xii, 118, 150 n. 9, 271–272, 275–277, 320–323
intentionality 260–261, 308 n.1
internal questions *see* Carnap
internalism 402–403, 415
interrogatives 118, 377

Jackson, F. 419 n. 33
Jeffrey, R. 31
Johnston, M. 306
Joseph, M. 379
judgement-dependence 247, 273–278, 292–293, 296, 297, 303, 395, 411–413

Kallestrup, J. 415
Kant, I. 66, 130, 156–157 161
Kemp, G. 195
Kenny, A. 32, n. 3, 103, 148
Kirk, R. 187, 189–191, 194–195, 415
Kripke, S. xii, 4, 65–71, 93, 96–98, 103–104, 192, 238, 399; causal-historical theory of reference 96–98; contingent *a priori* 68–71; necessary *a priori* truth 70–71; rigid designation 67–71, 399;
Kripke's Wittgenstein xii, xiv 127, 199, 200–237, 267, 289; rule-following 203–204, 210–212, 241 n. 16; sceptical paradox 199, 200–37, 266–268, 270–271, 295; sceptical solution 210–222, 226–262, 416 n. 2; straight solution 247–315, 419 n. 38
Kusch, M. 226, 239, 307

law of excluded middle 83, 397
law of non-contradiction 131, 216–217
Leibniz's Law *see* substitutability
Lemmon, E. J. 31,
Lepore, E. 194, 379, 381 n. 36
Lewis, D. 137, 149, 255, 256–258; reductionism 256–258
linguistic framework *see* Carnap
linguistic meaning xii, 24–51, 283, 303, 308 nn. 3, 6, 318–319, 323, 329, 371–372, 398, 419 n. 31
literal meaning *see* sentence-meaning
Locke, J. 48–54; associationism 48–54; systematic inversion 51–53
logical form 79–81, 87, 113–114
logical positivism xii, 3, 112–153, 164, 382, 397–398; *a priori* 116, 117, 128–137; verification principle 116, 117–123, 143, 144, 155, 164–167; *see also* Quine, analytic/synthetic distinction
logicism 32, n. 3
Longworth, G. 108 n. 58, 341
Lowe, E. J. 103
Ludlow, P. 111 n. 84
Ludwig, K. 379, 381 n. 36
Lycan, W. 104

McCulloch 103, 104, 105 n. 16 107–108 n. 36, 109 n. 61, 310 n. 40, 415, 420 n. 40
McDowell 103, 289, 291–297, 303, 307, 381 nn. 24, 31, 393, 395, 404–405, 415, 416, 419 nn. 32, 36, 38
McGinn 108 n. 56, 238, 239, 240 n.10, 247, 265–270, 271–273, 306, 308 n. 4, 404
Mackie, J. L. 286, 421 n. 44
McManus, D. 305-6
Makin, G. 104, 109 n. 68
Malcolm, N. 307
manifestation argument 391–393
Mares, E. 415
Martin, C. B. 305
Meinong, A. 78, 134
Melia, J. 104
mental content 248–251, 322, 329–330, 371–371, 398–399, 402, 404–405, 419 n. 32
mentalese *see* Fodor
meta-language 352
Mill, J. S. 129, 130
Millar, A. 244–245 n. 46
Miller, A. 148, 153 n. 58, 195, 198 n. 45, 241 n. 19, 243 n. 39, 245 nn. 46, 47, 306, 307, 312 n. 52,

315 n. 113, 341, 342 n. 22, 380, 415, 417 nn. 12, 15, 17, 419 n. 38, 420 n. 42, 421 n. 44
Milne, P. 379
minimalism 408–409, 420 n. 42
mirror constraint 180–183
Monk, R. 104, 307
mood 75–76, 118
Moore, G. E. 152 n. 58, 245 n. 47, 286
Morris, M. 148, 415
motivational internalism 223–226
Mumford, S. 197 n. 28
Myers, R. 380

natural kind terms 399–406
natural meaning *see* Grice
Neale, S. 111 n. 84
necessary *a priori* truth 70–71
necessary truth 66–70, 159, 399
necessity 128–137
negation 24, 82–85
Neurath, O. 112
Newton-Smith, W. 31
non-cognitivism 153 n. 58, 243 n. 39, 384–386; *see also* Frege–Geach problem
non-factualism 196 n. 16, 213–222, 228, 242 n. 26
non-natural meaning *see* Grice
Noonan, H. 32, n. 3, 103, 104
normative law 54–55
normativity of meaning 207–208, 210, 223–226, 230–237, 252, 265–269, 293, 413
normativity of sense 54–55, 100

object *see* Frege
object-language 352
objectivity 48–54, 55–57, 383–421; of meaning 394; of sense 48–54, 55–57; of truth 393–394, 412; of truth-conditions 383–421
ordered pairs/triples 21
Orenstein, A. 195

paradox of analysis 57–59
partial reductionism 297–304
Pears, D. 148
perlocutionary effect 325–326
Pettit, P. 248, 278–288, 306–307, 419 n. 33; ethocentrism 278–288
Plato 162 n. 7
Platts, M. 328–335, 341, 349–351, 353, 363–364, 379, 381 nn. 15, 24
possible worlds 66
pragmatics 76, 317
predicate expressions 11–12, 19–26
predicate logic 11–16
prescriptive law 54–56
principle of bivalence 396, 397, 418 n. 25
principle of charity 367–372
principle of compositionality 18, 26–27, 36, 99, 331–332
principle of *tertium non datur* 397, 418 n. 25
proper names 11, 15, 19–20, 22, 35–36, 44, 67–71, 77, 80–82, 86, 88–89, 91, 96–98, 113–114, 399
propositional attitudes 248–251, 331–334
propositional logic 9–10, 14–15
psychologism 106 n. 22
Putnam, H. 6, 194, 399–406; externalism 400–406; Twin-Earth thought experiment 400–406
psychologism 106 n. 22

quasi-realism 285–88
quietism 6, 291, 411–414
Quine, W. V. 4, 134–137, 149, 151 n. 46, 154, 169–191, 193,

197–198 n.31, 239 n.3, 372–373, 379, 380 n. 7; analytic/synthetic distinction 154, 155–161; holism 165–167; indeterminacy of translation 154, 169–172, 372–373; inscrutability of reference 197–198 n. 31; radical translation 171–184, 187, 189; stimulus meaning 173–184; underdetermination 185–191

radical interpretation 5, 366–372; *see also* Davidson
radical translation 170–184, 187, 189
Ramberg, B. 379
Ramsey, F. 309 n. 15
Read, S. 31
realism 223, 383, 386–399; arguments against 389–399
reductionism *see* Lewis; partial reductionism
reference 3
referential/attributive distinction 95–96
rigid designation 67–71, 399
Rogers, B. 149
rule-following 4, 203–204, 221, 280–285, 287–288, 393–396, 412–413; argument from 393–396; *see also* Kripke's Wittgenstein; Wittgenstein
Russell, B. 3, 77–98, 104, 107 n. 40, 133–134, 308 n.1, 337–339; bearerless names 77–85; definite descriptions 78–82; informativeness 88–90; names and descriptions 77–82, 107 n. 40; scope 83–85; substitution into belief contexts 85–88; theory of descriptions 77–90, 104, 337–339
Russell, G. 194
Sainsbury, M. 92, 104
satisfaction 358–363
sceptical paradox 199, 200–37, 266–268, 270–271, 295
Schiffer, S. 341
Schlick, M. 112, 117, 148
Schroeder, S. 148
scope 83–85
Searle, J. 318, 324–328, 340
Segal, G. 379, 415
semantic properties 16
semantic value 3, 17–21, 34–35, 39–41, 46–48; *see also* principle of compositionality; substitutabililty
sense 3, 34, 39–41, 46–57, 59, 178; and semantic value 39–41, 46–48; normativity of 54–55, 100; objectivity of 48–54, 55–57; transparency of 41–42, 59, 100
sentence-meaning 74, 317, 320, 323, 328, 329–334
sentential connective 10–16
sentential constant 10–16
set, definition of 33 n. 12
Shieh, S. 418 n. 23
Sinn 3, 104 n. 1
Smart, J. J. C. 188
Smith, M. 244 n. 40, 309 n. 13, 420 n. 42
Soames, S. 148, 149 n. 3
sophisticated dispositionalism *see* Boghossian
speaker's-meaning 5, 76, 317, 320–341; *see also* Grice
speech-act 76; *see also* illocutionary effect; perlocutionary effect; Searle
Stalnaker, R. 103
Stanley, J. 103
stimulus meaning 173–184

Strawson, P. F. 93–96, 161, 163–164, 193, 316, 317–318, 337–339, 340
substitutability 19, 99
Sullivan, P. 306
Summerfield, D. M. 307
syllogistic logic 14
synonymy 4, 58, 158, 160–161, 165
syntactic rule 15–16
syntax 14–16
synthetic truth 66–67, 128, 155–169
systematic inversion 51–57

Tarski, A. 346, 354–6, 364–366, 376–77, 379; Convention (T) 354–56, 364–66; truth-theory 354–356, 366
Tarskian truth-theories 354–363, 370–371, 376
tautology 114, 117, 149 n. 3
Tennant, N. 305
Textor, M. 103
theory of descriptions 77–90, 104, 337–339
Thornton, T. 307
thoughts 44–45, 51, 56, 64, 72, 100, 150 n. 6, 389
tone, 2, 73–74, 76, 100
translation 169–192, 363–64
transparency of sense 41–42, 59, 100
Travis, C. 108 n. 58
truth-aptness 142, 223, 230, 234–235, 384, 408–409
truth-conditions 29–30, 45–46, 71–73, 77, 100, 113, 149–156, 165, 210–211, 212–219, 242 nn. 26, 28, 243 n. 39, 307, 344–346, 355–363, 375–376, 383
truth-function/truth-functional connective 24
truth-tables 9–10, 16–18, 386
type–token distinction 7
Twin-Earth thought experiment 6, 400–406

Uebel, T. 148
universal quantification 12–13, 24–25
underdetermination 185–191
use–mention distinction 6

validity 9–10, 16–18, 34, 386; and meaning 17–18
variable 12, 21–22
Verheggen, C. 380
verifiability 117, 119–123
verification principle 116, 117–128, 143, 144, 155, 164–167
verification-transcendent truth 386–389; arguments against *see* acquisition challenge; manifestation argument; rule-following, argument from
Vienna Circle 112, 115, 137

Waismann, F. 112
warranted assertability 396–397
well-formedness *see* syntax
Weiner, J. 103
Weiss, B. 415
White, R. 148
Whiting, D. 239
Width of Cosmological Role 410
Wiggins, D. 379
Wilson, G. 226, 227–237, 239, 307
Wittgenstein, L. 4, 112–115, 210–211, 221, 241 n. 16, 248, 265, 289–297, 307, 393, 412–413; rule-following 4, 204, 211, 221, 241 n. 16, 307, 393, 412–413; *Tractatus* 112–115

Wright, C. 6, 148–149, 167–169, 194, 195, 215, 217–219, 221–222, 237, 238, 242 n. 34, 268–278, 289, 291–292, 306, 307, 314 n. 97, 318, 332–334, 341, 380, 388, 390–391, 393–396, 406–411, 412, 416 n. 1, 417 n. 9, 418 n. 22, 420 n. 42, 412 n. 45; Cognitive Command 410; minimalism about truth-aptness 408–409, 420 n. 42; on the realist/anti-realist debate 406–411; on quietism 6, 421 n. 45; Width of Cosmological Role 410

Yolton, J. 105 n.13

Zalabardo, J. 219–223, 237, 238, 306
Ziff, P. 324, 340, 341

Taylor & Francis eBooks

Helping you to choose the right eBooks for your Library

Add Routledge titles to your library's digital collection today. Taylor and Francis ebooks contains over 50,000 titles in the Humanities, Social Sciences, Behavioural Sciences, Built Environment and Law.

Choose from a range of subject packages or create your own!

Benefits for you

- Free MARC records
- COUNTER-compliant usage statistics
- Flexible purchase and pricing options
- All titles DRM-free.

REQUEST YOUR **FREE** INSTITUTIONAL TRIAL TODAY

Free Trials Available
We offer free trials to qualifying academic, corporate and government customers.

Benefits for your user

- Off-site, anytime access via Athens or referring URL
- Print or copy pages or chapters
- Full content search
- Bookmark, highlight and annotate text
- Access to thousands of pages of quality research at the click of a button.

eCollections – Choose from over 30 subject eCollections, including:

- Archaeology
- Architecture
- Asian Studies
- Business & Management
- Classical Studies
- Construction
- Creative & Media Arts
- Criminology & Criminal Justice
- Economics
- Education
- Energy
- Engineering
- English Language & Linguistics
- Environment & Sustainability
- Geography
- Health Studies
- History
- Language Learning
- Law
- Literature
- Media & Communication
- Middle East Studies
- Music
- Philosophy
- Planning
- Politics
- Psychology & Mental Health
- Religion
- Security
- Social Work
- Sociology
- Sport
- Theatre & Performance
- Tourism, Hospitality & Events

For more information, pricing enquiries or to order a free trial, please contact your local sales team:
www.tandfebooks.com/page/sales

Routledge
Taylor & Francis Group
The home of Routledge books

www.tandfebooks.com