The Presence of Mind

Daniel D. Hutto

Advances in Consciousness Research

3/w

THE PRESENCE OF MIND

ADVANCES IN CONSCIOUSNESS RESEARCH

ADVANCES IN CONSCIOUSNESS RESEARCH provides a forum for scholars from different scientific disciplines and fields of knowledge who study consciousness in its multifaceted aspects. Thus the Series will include (but not be limited to) the various areas of cognitive science, including cognitive psychology, linguistics, brain science and philosophy. The orientation of the Series is toward developing new interdisciplinary and integrative approaches for the investigation, description and theory of consciousness, as well as the practical consequences of this research for the individual and society.

EDITOR

Maxim I. Stamenov
(*Bulgarian Academy of Sciences*)

EDITORIAL BOARD

David Chalmers (*University of Arizona*)
Gordon G. Globus (*University of California at Irvine*)
Ray Jackendoff (*Brandeis University*)
Christof Koch (*California Institute of Technology*)
Stephen Kosslyn (*Harvard University*)
Earl Mac Cormac (*Duke University*)
George Mandler (*University of California at San Diego*)
John R. Searle (*University of California at Berkeley*);
Petra Stoerig (*Universität Düsseldorf*)
Francisco Varela (*C.R.E.A., Ecole Polytechnique, Paris*)

Volume 17

Daniel D. Hutto

The Presence of Mind

THE PRESENCE OF MIND

DANIEL D. HUTTO
University of Hertfordshire

JOHN BENJAMINS PUBLISHING COMPANY
AMSTERDAM/PHILADELPHIA

The paper used in this publication meets the minimum requirements of American National Standard for Information Sciences — Permanence of Paper for Printed Library Materials, ANSI Z39.48–1984.

Library of Congress Cataloging-in-Publication Data

Hutto, Daniel D.
The presence of mind / Daniel D. Hutto.
p. cm. -- (Advances in consciousness research, ISSN 1381-589X ; v. 17)
Includes bibliographical references and index.
1. Intentionality (Philosophy). 2. Intentionalism. 3. Connectionalism. 4. Act (Philosophy)
I. Title. II. Series.
B105.I56H87 1999
128'.2--dc21 99-22516
ISBN 90 272 5137 1 (Eur.) / 1 55619 433 1 (US) (Pb; alk. paper) CIP

© Copyright 1999 - John Benjamins B.V.
No part of this book may be reproduced in any form, by print, photoprint, microfilm, or any other means, without written permission from the publisher.

John Benjamins Publishing Co. • P.O.Box 75577 • 1070 AN Amsterdam • The Netherlands
John Benjamins North America • P.O.Box 27519 • Philadelphia PA 19118-0519 • USA

For my parents
ab imo pectore

Table of Contents

List of Figures xi
Acknowledgements xiii

Introduction 1
1. The Background History of the Debate 1
2. The Structure and Argument of the Book 5

Chapter 1. Nothing in Mind
1. Introduction 9
2. The Meta-Argument of Eliminativism 10
2.1. Theory-Theory 11
2.1.1. Causal Explanatory Commitments 11
2.1.2. Methodological Considerations 12
2.2. Theories and Conceptual Schemes 13
2.2.1. Minimal Coherentism 15
2.3. Major Conceptual Revolutions 17
3. Changing Our Minds 19
3.1. Connectionism: The Champion of Eliminativism 19
3.2. The Information Processing Debate 20
3.3. Representationalism and Expansive Classical Cognitivism 21
3.3.1. Why Mentalese Representations? 23
3.4. Propositional Attitude Psychology and the Language of Thought 25
4. The Connectionist Threat 27
4.1. Assessing the Connectionist Threat 28
5. Conclusion 29

Chapter 2. A Lack of Content

1. Introduction 31
2. The Problem of Causal Impotence 31
2.1. The Tension between Externalism and Strong Supervenience 34
2.2. Why Commonsense Psychology is Externalist 36
2.2.1. Two Factor Theories 37
2.2.2. Narrow Content 39
2.3. All Roads Lead to Externalism 41
3. The Problem of Content 42
3.1. Causal Theories of Content 42
3.2. Asymmetric Dependency Theory 46
3.3. The Rejection of Causal Theories 49
4. The Information-Theoretic Approach 49
4.1. Informational Content 50
4.2. Teleofunctional Representation 54
4.3. The Property-Specific Character of Digital Information 56
4.4. The Indeterminacy Problem 58
5. Conclusion 59

Chapter 3. Black Dots and Red Herrings

1. Introduction 61
2. Biosemantics: The Natural Choice 62
2.1. Etiological versus Propensity Accounts of Selection 63
2.2. The Reproducible Character of Biological Devices 65
2.3. Biosemantics and Intentionality 67
3. Defending High Church Biosemantics 70
3.1. The Intensional Indeterminacy Argument 70
3.2. The Capacity Argument 72
3.3. The Reply from on High 76
3.4. A Modest Proposal 78
4. Conclusion 86

Chapter 4. Seeing Without Believing

1. Introduction 87
2. Nonconceptual Content Introduced 87
2.1. The Virtues of Nonconceptual Content 89
2.2. The Role of Nonconceptual Content in Conceptual Development 93
2.2.1. Non-Objective Thought, Generality and Connectionism 94
2.2.2. The Need for Correctness Conditions 101
3. The Failure of Eliminativist Plan A 103
3.1. Structuring Causes and Biosemantic Explanations 105
3.2. Domestic Anthropology and the Limits of Propositional Ascription 108
3.3. Connectionist Processing and Situated Robotics 110
4. Conclusion 112

Chapter 5. Interpreting Minds

1. Introduction 113
2. Intensionality Introduced 113
2.1. The Intensionality Test 114
2.2. The Essentially Social Nature of Interpretation 115
2.2.1. The Very Idea of Asocial Triangulation 119
2.3. The Origin of Conceptual Perspectives 122
3. The Internal Structure of Propositions 124
3.1. Context Invariant Symbols and External Props 124
3.2. Syntactic Structure and Radical Interpretation 126
3.3. Theories of Truth and Theories of Meaning 128
4. Normativity and Propositional Ascription 129
4.1. Two Principles of Charity 129
4.2. The Further Need for the Principle of Correspondence 133
4.3. Irreducible Norms 136
5 Conclusion 138

Chapter 6. Davidson's Identity Crisis

1. Introduction 139
2. What Motivates Causalism? 140
2.1. The Neo-Wittgensteinian Orthodoxy 141
2.2. The Primary Reason Problem 141
2.2.1. Why Causalism doesn't Solve the Primary Reason Problem 143
3 The Mysterious Connection Problem and Anomalous Monism 146
3.1. The Token Identity Theory 147
3.1.1. Clarifying Davidson's Understanding of Supervenience 149
3.2. The Epiphenomenal Objection 151
3.3. Davidson's Replies 153
3.3.1. The Weak Extension Reply 154
3.3.2. The Strong Extension Reply 158
4. The Mysterious Connection Reconsidered 160
4.1. Agency and Non-Humean Causation 161
5. Conclusion 163

Chapter 7. The Poisoned Chalice

1. Introduction 165
2. Eliminativism: Plan B 165
2.1. Churchland's Coherentist Argument 166
3. Assessing the 'Lack of Coherence' Charge 168
3.1. Why Minds Matter 169
4. The Eliminativist Burden 172
4.1. What Brains Can't Do 172
4.2. The Training of Connectionist Networks 175
4.3. The Nonconceptualist Reply 176
4.4. Minding Our Language 179
5. Conclusion 181

Notes 182
References 195
Index 223

List of Figures

Figure 1.	Defusing strategy A - stage 1	6
Figure 2.	Defusing strategy A - stage 2	7
Figure 3.	Defusing strategy B	8
Figure 4.	Fodor's implementation levels	35
Figure 5.	The disality problem	43
Figure 6.	The misrepresentation problem	44
Figure 7.	The anomic nature of the asymmetric dependency relation	48
Figure 8.	The branches of binary decision-making	51
Figure 9.	Dretske's distinction between causal and informational links	52
Figure 10.	Actual and possible causal links	53
Figure 11.	The transitivity problem	59
Figure 12.	The bee dance: triangulation at the intentional level	68
Figure 13.	Neander's low church biosemantics	73
Figure 14.	Intentional direction and informational sensitivity	82
Figure 15.	A version of the Müller-Lyer illusion	88
Figure 16.	Davidson's example of triangulation in primitive learning	118
Figure 17.	The development of the concept of belief	118
Figure 18.	An Individual triangulating with itself diachronically	120
Figure 19.	The intensionality of explanation	155

Acknowledgements

The initial question that inspired this book was posed by my good friend Michael (the Dane) Busk-Jepsen when we were MPhil students together at St. Andrews. After I had just finished a dissertation on the topic of self-deception he asked me what I thought beliefs were. I, rather worryingly, had no ready answer. This, in conjunction with George Graham's warning that some philosophers denied the very existence of beliefs, fostered my interest in investigating and defending commonsense psychology. I am particularly grateful to George for his friendship, generous support and advice over the years. But there are also many others to whom I wish to thank.

The first and most important debts are owed to my wife, Farah, and my long-suffering parents, Margaret and William Hutto. Their moral, and financial support have been necessary conditions for the completion of this work. In particular, despite reservations, my parents have given me the encouragement and fostered my self-confidence in a way that has encouraged me to pursue my ambitions wherever they taken me – even into the depths of philosophical and metaphysical speculation. They have been the pillars upon which this project has rested, in one way or another, for the past nine years. On a different note, I would like to apologise to my two sons, Alexander and Rais, for not being able to play with them as often as they (and I) would have liked, and for locking my study door. Also I am very grateful to my father, my wife, Sue Anthony and Slyvester Kociniak for helping to proof-read the manuscript at various stages.

Debts of an intellectual kind are owed to: Martin Bell, José Bermudez, Filip Buekens, Ignar Brinck, Paul Coates, Anthony Chemero, Ron Chrisley, Charles Conti, Adrian Cussins, Robert Gordon, Brendan Larvor, Kathleen Lennon, Harry Lewis, John Lippitt, Mark Jones, Richard Menary, Sonya Sikka, Peter Steffenson, Karsten Stueber, Stephen Thomas, Julia Tanney and Christine Vergeheggen; all of whom helped clarify my thinking on various matters. I am also grateful to the audiences and organisers of the following conferences for giving me the opportunity to articulate and defend my views: 'New Directions in Cognitive Science' Finland, 1995; 'Fifth Annual European Society for Philosophy of Psychology', Barcelona, 1996; 'European Congress for Analytical Philosophy II' Leeds, 1996; 'New Trends in Cognitive Science', Vienna, 1997; "Fourth

International Conference on Persons', Prague, 1997; 'Contents and Concepts Conference', Copenhagen, 1998; 'Seventh Annual European Society for Philosophy and Psychology', Lisbon, 1998; 'Normativity and *A Priori* Knowledge Workshop', Merton College, Oxford 1998; and the 'Mind and Language Seminar' and 'Nonconceptual Content Workshop', C.R.E.A., Paris, 1998. Similarly, I am grateful to the following Philosophy Departments for inviting me to speak at their colloquia: City College of New York, College of the Holy Cross, Mass., University of Aarhüs, University of Connecticut, University of Leeds, and the University of Sussex. I would especially like to thank Donald Davidson and Ruth Millikan for taking the time to read and comment on parts of my work and for setting me right in places.

I would also like to thank several editors for allowing me to re-use some material in this book. Earlier versions of parts of chapter three, four and five have appeared in my articles "Nonconceptual Content and Objectivity" (1998) In R. Grush (ed) *The Electronic Journal of Analytic Philosophy: Special Issue on Gareth Evans* http://www.phil.indiana.edu/ejap/ejap.html and "Cognition without Representation?" in *Understanding Representation* (1999) In A. Riegler, A. Von Stein, and M. Peschl, M. (eds) New York: Plenum Press. Material from chapter six has been reworked and integrated from my articles "Davidson's Identity Crisis" (1998) *Dialectica*, 52 (1). and "A Cause for Concern: Reasons, Causes and Explanation" (1999) *Philosophy and Phenomenological Research*. 59 (2).

I would especially like to express my gratitude to the Arts and Humanities Research Board of the British Academy for awarding me a Research Leave Scheme grant. This provided sabbatical relief for the academic year 1998-1999 during which some of the research for this book was carried out. Finally, I would like to express my gratitude to the University of Hertfordshire for its steadfast support and encouragement in all my intellectual endeavours.

Introduction

"A philosopher is a man who has to cure many intellectual diseases in himself before he can arrive at the notions of common sense"

– Wittgenstein, *Culture and Value,* 1944

Sitting in the dock charged with a capital offence is a sombre experience, I imagine, even for an account of action explanation. But such is the current situation for 'folk psychology', or as I, with less prejudice, prefer to call it 'commonsense psychology'. Commonsense psychology maintains that the best explanations of our actions are those given in terms of our beliefs and desires. This is, of course, an extremely familiar practice. Nevertheless commonsense psychology has been charged with being a misleading theory of mind and action because it is out of step with the metaphysics inspired by natural science. The charge is clear enough, but as an interested onlooker, I believe that its most prominent defence lawyers have handled the case rather badly. They have allowed the prosecution to dominate the proceedings by failing to question the preferred characterisation of the accused and by far too readily accepting the basic assumption about what constitutes a proper metaphysics. My aim is to offer a different, and hopefully stronger, defence.

1 The Background History of the Debate

Before I begin my defence it is useful to remind the reader of the history of the debate and some key background assumptions underpinning it. It is commonly supposed that propositional attitude psychology is a kind of low-grade theory which posits the existence of 'inner' mental states or events, such as beliefs and desires, as a means by which to causally explain our outward behaviour and action. Crane rightly remarks, "It has become customary among many philosophers (and some psychologists too) to describe the assumptions and hypotheses we adopt when understanding other minds as a sort of theory of other minds" (Crane 1995: 53). This thesis has become known as the 'theory-theory' in philosophical circles. It is called such because the idea that commonsense

psychology is theoretical is just one theory about its nature – albeit the dominant one. As Lewis, the father of theory-theory, saw it terms such as belief, desire and hope are theoretical terms (cf. also Stich 1983: ch. 2 MacDonald 1989: ch. 2). He suggested they gain their meaning by the relations that hold between them and other like and unlike terms.[1] Some of these other terms are part of the same theory but others will be independent of it.

Lewis used the example of Cluedo to make his point clear. He treated the place-holders in the game of Cluedo, for example; Professor Plum, Colonel Mustard, Miss Scarlet, as if they were all terms of a single theory – call it suspect theory. These entities stand both in relation to one another and to other entities which are not a part of suspect theory. For instance, we define Professor Plum in terms of the total sum of his relations. We know he is in the drawing room, we know he doesn't have the rope because Colonel Mustard has it, and we know he doesn't have the lead pipe since Miss Scarlet has it. We know a fair deal about Professor Plum because we understand his role in relation to the other named suspects and various other things in the game. But we do not know anything interesting about the entity that fills his role. For example, we know nothing about Professor Plum's looks, personality or character. All we know is his logical place in relation to other entities in the universe of Cluedo. Lewis reckoned it was much the same with our mental concepts. Terms like 'belief' and 'desire' are similar place-holding terms defined by their relation to other folk psychological terms and to non-theoretical entities and events – such as their typical causes and effects. In this way, such terms must be understood by means of implicit functional definitions.

Treating such concepts as if they were theoretical was a big advance over the treatment they received when psychological behaviourism was orthodoxy. Behaviourists refused to accept that non-observable mental phenomena had any place in serious psychological theorising. In their view, any and all genuine mental phenomena needed to be identified with dispositions to behave and had to have operational definitions in terms of observable causes and effects. Behaviourists were encouraged to treat our everyday psychological concepts as reducible to visible stimuli and responses because they thought it was the best way to maintain scientific credibility. The word of the day was that psychologists should be concerned with nothing over and above what is observable and testable; in this case, charting and describing the stimulus-response behaviour of subjects. This attitude was fostered, at least in part, because the dominant philosophy of science of the time suggested that proper science was ultimately only concerned with those theoretical terms that would reduce to a set of observable facts. Thus by sticking solely to what could be observed, i.e. stimulus-response behaviour, no entities or processes could be introduced into psychological theorising which would fall beyond the reach of controlled analysis and testing. It is easy to see how a

behaviouristic understanding of psychological phenomena squared with positivistic views about science. Positivism, which is a species of empiricism, encouraged the idea that all genuine statements of higher-order theory would reduce to rather complex statements about sense-data. The point is that statements about the latter could be tested directly against experience. However the idea that proper science is ultimately about the reduction of theories to a collection of observation statements was not a permanent fixture. Today the epistemological concerns of the positivists are out of vogue and much of their philosophy of science has been rejected. Nonetheless their vision of how theories reduce in a downward direction from one to another has remained in a modified form. Much current thinking about theories and naturalisation has a lineage which can be traced directly back to positivism and the very roots of the analytical tradition in philosophy. In conjunction with these changes in the philosophy of science, and despite its obvious advantages over behaviourism, the advent of theory-theory has generated some difficult questions about the scientific status of commonsense psychology.

Contemporary physicalism has inherited the general structure and methodological goal of reductionism from positivism, but it has traded the epistemic bedrock of sense-data for an ontological ground floor of basic physical entities, events and their relations. Most of today's philosophers are happy to accept that theories carry ontological implications such that if a theory is true then the entities and events about which it speaks must exist and they must have a causal purchase on the world. When characterising physicalism, John Dupré suggests that the received view pictures nature as a kind of 'cosmic machine' (see Dupré 1993: 8). It is, in effect, a "...general picture of the world of causes which works according to natural laws or regularities" (Crane 1995: 4). The point of the analogy being that if we compare nature to a machine we can understand its workings by looking for mechanisms working in harmony with one another according to some set of general laws. It becomes an *a priori* demand of those who accept such a picture that all genuine phenomena ought to be explicable in terms of some series of natural laws which accord with the most basic laws of the universe.

The shifts from behaviourism to theory-theory, and from positivism to physicalism, go hand in hand to generate concerns about the status and future of commonsense psychology. These changes have prepared the ground for the current debate about whether or not it will 'naturalise' and become an accepted part of the scientific world view. It is against this background that, at present, serious questions are being raised about both the character and value of our familiar way of understanding ourselves. This situation brings certain questions to light: Do the advances of cognitive science and neuroscience offer us new, and more fertile,

means of understanding ourselves? Are we about to discover that human nature is not quite as we ordinarily think it is? Will our everyday account of ourselves be vindicated by a new science? Or will our self-understanding remain untouched by such developments? Philosophers are divided on these issues. Some argue that commonsense psychology needs complete replacement because it is out of step with natural science. Others anticipate it will find a happy home within modern scientific theories. Still others hold that while it cannot be regarded as a scientific account of mind it has a 'special status', and simply cannot be eliminated. Finally, and more radically, some maintain that this entire debate is wrong-headed because it operates against the background of a misunderstanding of the very nature of commonsense psychology.

The perceived stakes are high. The consequences of choosing to support revision, elimination or retention with respect to commonsense psychology is often set out in dramatic terms. Consequently, we find Fodor remarking that, "If commonsense intentional psychology really were to collapse, that would be, beyond comparison, the greatest intellectual catastrophe in the history of our species" (Fodor 1987: xii). While, on the other hand, Churchland puts the opposite case by saying, "...should we ever succeed in making the shift [away from commonsense psychology], we shall be properly at home in the physical universe for the very first time" (P.M. Churchland 1979: 35). In her book *Saving Belief*, Lynne Rudder Baker regards the elimination of commonsense psychology as akin to a kind of cognitive suicide because of its role in underwriting a wide variety of key practices. She offers the following list: (i) Social practices that depend upon ordinary explanation and prediction of behaviour would become unintelligible, (ii) Moral and legal practices would become senseless, (iii) Linguistic practices would become mysterious, (iv) Psychological theorising would become problematic (Baker 1988: 130–133). While this catalogue of vital roles reveals commonsense psychology's indispensability, my own view is that debate over its future is of key philosophical importance because of its links to questions about epistemology and the nature and existence of content. Where we stand in this debate has implications for our views about the philosophy of science and our position in the world. In particular, if we find we cannot do without the kind of content which is an integral part of propositional attitude psychology, yet we also find we cannot accommodate it within a purely scientific account of the world, then we must revise our metaphysics. Indeed this is my conclusion, but I leave the project of finding a replacement metaphysics to a sister book *Beyond Physicalism* (Hutto forthcoming-b). Bearing these background issues and high stakes in mind, I now outline the general structure and argument of this book.

2. The Structure and Argument of the Book

There are two questions this monograph seeks to address. First: Do the new trends in cognitive science pose a genuine threat to commonsense psychology? To answer this question properly requires an investigation into the nature of both current cognitive science and commonsense psychology. The bulk of this book is taken up with that investigation. In this regard it is vital to distinguish issues concerning intentional content from those to do with propositional content. I argue that, when correctly understood, the former has good prospects for naturalisation and is in harmony with recent developments in cognitive science. Yet, in contrast, I argue that propositional attitude psychology *proper* (which is distinct from intentional psychology) is best understood along interpretationalist lines. In taking this stance, I claim that it is misrepresented when regarded as 'theoretical' in a narrow sense. Its explanatory concerns are not causal in character and do not conflict with those of cognitive scientists. The complicated answer to the first question is that, for different reasons, neither intentional nor propositional attitude psychology are threatened by the developments in today's cognitive science. But this response only sets the stage for a second question: If propositional attitude psychology has no home in the metaphysics inspired by natural science, shouldn't it be abandoned? I argue that the best naturalistic account of content we can expect from cognitive science is inadequate to explain proper scientific theorising. Due to this *lacuna* the elimination of propositional attitude psychology is not a coherent option.

Chapter one outlines the nature of the eliminativist threat to commonsense psychology and articulates the minimal assumptions which are needed to motivate the most stable, and threatening, form of eliminativism. For example, it is argued that although commonsense psychology need not be thought of as a theory it must be at least regarded as a kind of conceptual scheme. And given that conceptual schemes are generally open to revision it becomes a live question whether or not commonsense psychology is due for revision or elimination. There are two strategies that the eliminativists have employed to convince us that it ought to be abandoned. I call these strategies A and B. They are:

A. To establish that commonsense psychology should be eliminated because it is out of step with recent developments in cognitive science - in particular, connectionist accounts of information processing.

B. To establish that commonsense psychology should be eliminated because it does not cohere with the natural sciences.

After having outlined the eliminativist meta-arguments and their two main strategies, I assess the strategy A's prospects for establishing that revolutionary connectionist theories pose a genuine threat to commonsense psychology. It is argued that such theories really only pose a threat to classical cognitivist theories of information processing. But the debate about the future of commonsense psychology involves more than a debate over the right information processing theory. For example, Fodor's version of expansive classical cognitivism – which is explicitly committed to full-blown linguaform representational contents – is meant to offer genuine hope of a scientific vindication of propositional attitude psychology. However despite the fact that the connectionist threat to classical cognitivism is not decisive, it is argued that a Fodorian-style *lingua mentis* fails in its bid to provide a straightforward defence of commonsense psychology. The trouble is that it cannot satisfactorily account for mental causation and, relatedly, it fails to provide a naturalistic theory of representational content. With respect to this latter issue I consider the fundamental flaws of causal theories, Fodor's asymmetric dependency theory and Dretske's teleofunctional theory of content. The structure of this stage of the argument is represented in figure 1.

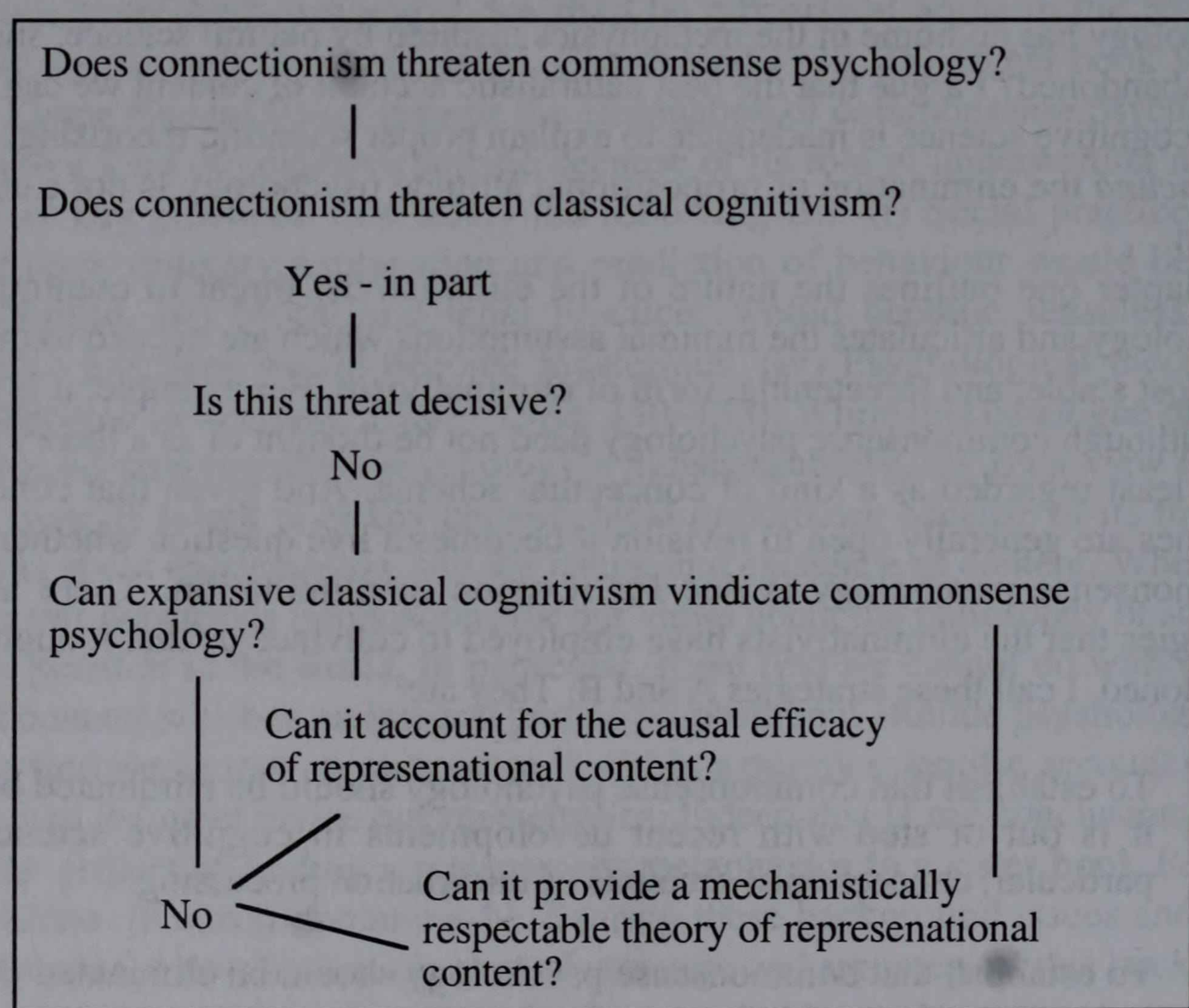

Figure 1: *Defusing strategy A – stage 1.*

In response to these difficulties, I go on to defend a modest biosemantic theory of intentionality, in chapters three and four, and argue that it complements and is complemented by an account of nonconceptual content. The good news is that this modest version of biosemantics is well partnered by connectionist theories of information processing. It also fits neatly alongside cutting edge work on embodied and embedded cognition. On these grounds, I conclude that a nonconceptual account of intentionality is not only compatible with today's cognitive science, it is in harmony with it. The less than good news is that the kind of nonconceptual content, which is the hallmark of intentionality, is not equivalent to propositional content. And the latter is required for an proper account of full-blown propositional attitude psychology. This provokes the interim verdict that, minimally, to fully defuse the connectionist threat it would have to be shown that a modest biosemantic theory of intentionality is not at odds with our best account of propositional content. In the wake of these conclusions, chapter five goes on to defend an interpretationalist approach to propositional attitude psychology and reveals that our capacity to interpret others non-reductively depends, in part, upon the kind of intentionality described in chapters three and four. Hence intentional and propositional attitude psychology are compatible. The structure of this stage of the argument is represented in figure 2.

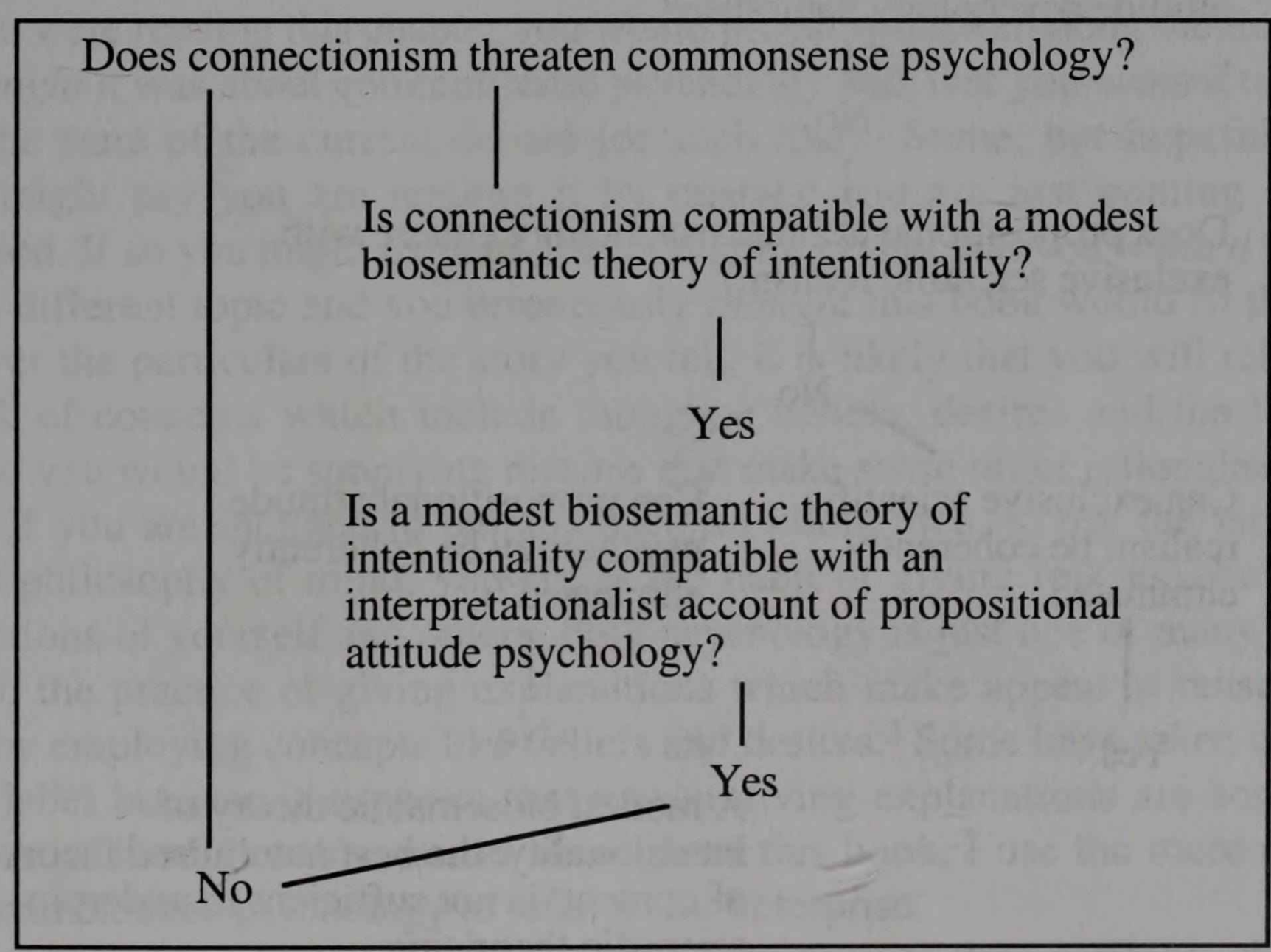

Figure 2. *Defusing strategy A – stage 2*

So far, so good. But there is still strategy B to deal with. For if propositional attitude psychology is best understood by interpretationalist lights then any attempt to absorb it into a scientific psychology, as is the project of its some self-styled friends, is misguided. A fully naturalised commonsense psychology is not in the cards. Connected with this, in chapter six, despite supporting his non-reductive interpretationalism, I criticise Davidson's compatibilist attempts to reconcile it with physicalism. In doing so, I argue against the standard causalist understanding ofthe propositional attitudes.

In the final chapter, I give attention to strategy B which proposes that commonsense psychology ought to be eliminated on the mere grounds of its irreduciblity. The problem with this idea is that the best naturalised theory of content we are likely to get from cognitive science, as outlined in chapters three and four, is inadequate to explain scientific theorising. But such an explanation is necessary to anyone committed to exclusive scientific realism. Thus, by appeal to the minimal coherentist epistemology that enabled us to motivate a secure eliminativism in the first place, it is revealed that the exclusive scientific naturalism sponsored by it is not just practically, but intellectually incoherent. For this reason we must question the fitness of an exclusive materialism.

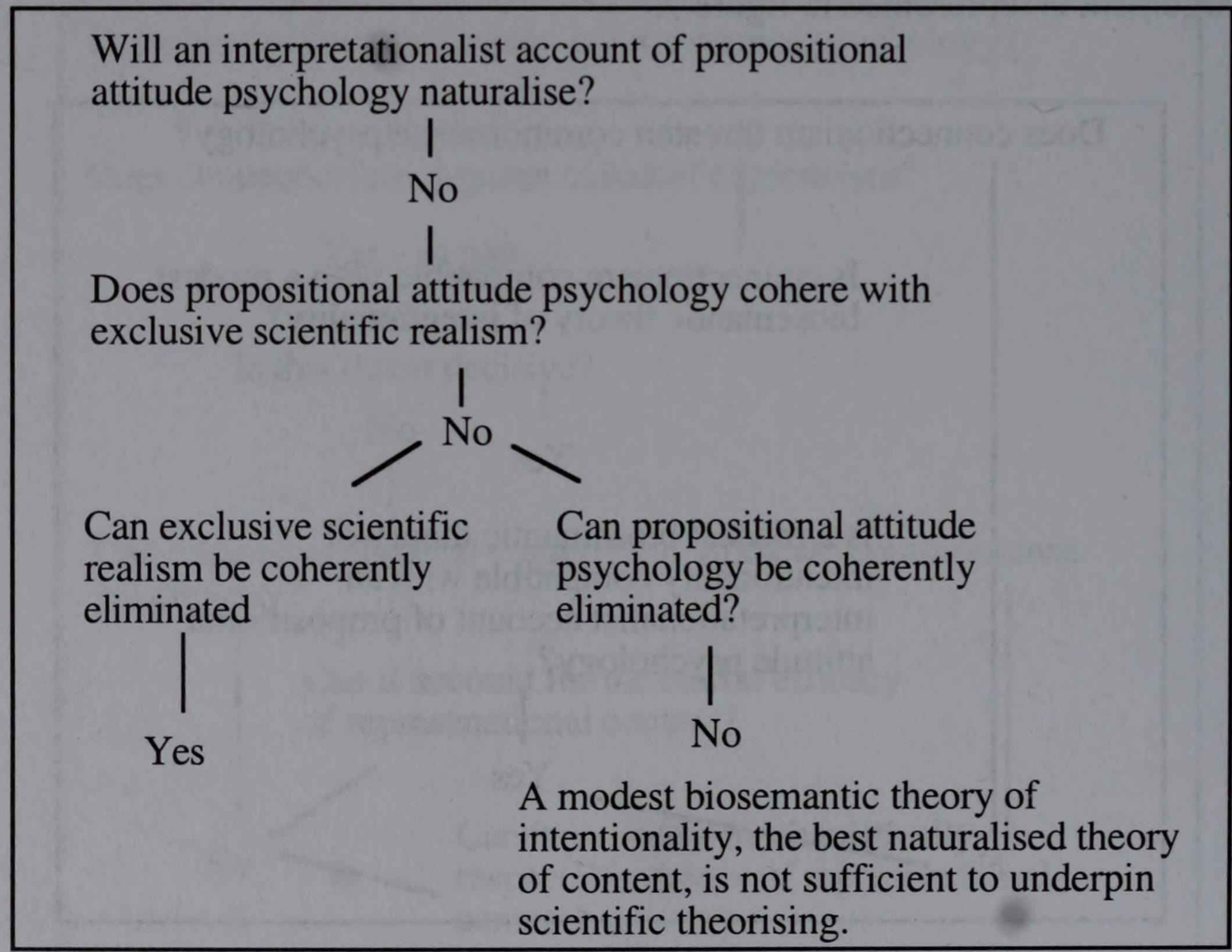

Figure 3. *Defusing strategy B*

CHAPTER ONE

Nothing in Mind

"I am strongly inclined to think that
the right thing to say is that there are
no such things as beliefs."

– Stich, *From Folk Psychology to Cognitive Science,* 1983

1. Introduction

Nothing could be more commonplace than reason explanations. If you were asked why you were reading this chapter you would probably answer along the lines that you *thought* it was about commonsense psychology and, that you *wanted* to know about the state of the current debate (or such like).[2] Some, but hopefully not many, might say you are reading it by mistake and are just coming to this realisation. If so you might explain your action by saying that you *wanted* to read about a different topic and you erroneously *thought* this book would fit the bill. Whatever the particulars of the story you tell, it is likely that you will rely on a network of concepts which include thoughts, beliefs, desires and the like. In doing so you would be supplying reasons that make sense of, or rationalise, your action. If you are inclined to employ such an idiom then, to use the jargon of today's philosophy of mind, you are in the habit of giving folk psychological explanations of yourself and others. Folk psychology is just one of many names given to the practice of giving explanations which make appeal to reasons for acting by employing concepts like beliefs and desires.[3] Some have taken offence to this label because it suggests that reason-giving explanations are somehow folksy or inferior. For this reason, throughout this book, I use the more neutral term commonsense psychology to refer to this enterprise.

According to Paul Churchland, we should be open to the possibility that this familiar way of understanding ourselves might be replaced by a more thorough-going, purely scientific account of cognition. Insisting that commonsense psychology is a species of theory, he argues that it is a pre-scientific folk theory

which is out of step with the rest of science. The view that commonsense psychology has dismal future prospects is typical of his programmatic and revisionary attitude toward all pre-scientific theories. His eliminativism is motivated by the desire to unify all theories with a coherent scientific world-view in what is envisioned to be a positive expansion of the human epistemic condition (P.M. Churchland 1979: 36, 1989: 242, 1993: 219). Commonsense psychology is denounced as false and misleading because it will not naturalise. Thus he believes the only acceptable attitude we can take towards it is to advocate its complete eradication. Quite literally, the eliminativist goal is to save us from our*selves*.

Before mounting a defence, it is important to be explicit about the assumptions which ground eliminativism. In order to make a viable case for abandoning commonsense psychology, the eliminativists must minimally convince us of the truth of claims (a) and (b) below.

(a) our practice of explaining actions by appeal to reasons depends, at least partially, on acceptance of a conceptual scheme involving terms such as belief, desire, etc.

(b) conceptual schemes are, *in general*, vulnerable to the tides of intellectual fashion and thus apt for change, revision or complete abandonment.

Only if we accept these two claims is it a live question whether or not we have good reason, in particular, to revise or eliminate commonsense psychology.[4] On the whole, I am prepared to subscribe to claims (a) and (b), but it is the eliminativist endorsement of a further claim (c) that is the real source of contention.

(c) there are good reasons to think commonsense psychology will be eliminated or revised.

I devote most of this chapter to a more detailed analysis of claims (a) and (b), before giving detailed attention to claim (c) in the rest of this book.

2. The Meta-Argument of Eliminativism

Minimally, the eliminativist must convince us of two points in order to run their argument. First, they must convince us that when we attempt to interpret, predict and understand actions by invoking reasons we are labouring under a potentially dispensable conceptual scheme (i.e. discourse, language game, etc.). This is vital.

For only if we accept that we are indeed using some form of conceptual scheme will we be prepared to see it as potentially revisable.

2.1. Theory-Theory

It is with this goal in mind that the Churchlands attempt to convince us that commonsense psychology is theoretical in nature (P.S. Churchland 1986: 299–312, P.M. Churchland 1979, 1981b, 1989, 1991). The view that reason-giving psychology is a kind of proto-scientific theory, or folk science, is known as theory-theory in philosophical and psychological circles. According to theory-theorists commonsense psychology is theoretical in two important, but distinct, ways.

2.1.1. Causal Explanatory Commitments

Firstly, commonsense psychology has ontological import. Putatively it theorises about what goes on *inside our heads* – in the black box that lies between perceptual stimuli and behavioural responses. In doing so, it allegedly posits inner entities which causally explain outward behaviour. It is theoretically committed to the entities which are responsible for our actions (cf. Loar 1981: 127). More precisely, the widespread assumption is that propositional attitude psychology, which explains our actions by appeal to the causal influences of contentful mental items such as beliefs and desires, is equivalent to our quotidian practice of offering reasons to explain behaviour.

'Propositional attitude' is a term of art. It refers to members of a class of complex mental states which have a dual nature; each such state is composed of a psychological attitude which is related to a given propositional content. A host of ordinary mental items, including beliefs, hopes and desires, fall under the umbrella of this term. A state's *attitudinal* aspect describes the kind of attitude which is connected with a given proposition. Examples of such psychological attitudes are legion; they include all the mental states that can be followed by a that-clause. For example, it is possible to believe-that-Φ, desire-that-Φ, recognise-that-Φ, hope-that-Φ, and so on. The *propositional* element is the 'propositional content' which fills the Φ place: it is the particular proposition to which a given psychological attitude relates. Hence just as one might adopt quite a range of psychological attitudes towards a single proposition one could also entertain a great number of propositions in relation to a single type of psychological attitude. For example, although you might believe that "This book is about the philosophy of mind", it is possible to replace the Φ, in "I believe-that-Φ", with any number of different propositions such as; "The meeting was at four o'clock", "London is twenty miles away", "Monday is the first day of exams", and so on. One can get a

feel for the scope of this kind of scheme by reflecting on the fact that any meaningful sentential assertion can be regarded as the expression of a proposition. Embedded in the right explanatory framework it is easy to see how such a generative arsenal of entities would be a powerful resource for those hoping to explain action.[5]

2.1.2. Methodological Considerations

This brings us to the next reason why we might wish to regard commonsense psychology as a theory. It concerns the way in which we ascribe propositional attitudes and make predictions on the basis of attitudes so ascribed. Theory-theorists claim that our ability to make explanatorily revealing propositional attitude ascriptions is based on our recourse to a shared theory about human psychology (cf. Heal 1995: 50–51, Nichols *et al* 1996: 49). Putting aside developmental concerns about nativism and the like, the essential thought is that such a theory provides knowledge of anthropocentric facts – i.e. how people generally react, tend to feel, or usually think in this or that situation. According to the orthodox version of this proposal, which is sometimes called narrow theory-theory, this law-like body of information stored in sentential form (cf. P.M. Churchland 1989: 234, Stich and Nichols 1992: 49). For example, using our workaday knowledge about human behaviour, we can anticipate that if a person bangs their leg on a desk, and they have no leg protection, then they will feel pain, *ceteris paribus*. Logically complex, general propositions of this kind are the rough basis upon which we make predictions and explanations. With reference to Lakatos, George Botterill has tried to explicitly state some of the principles of what he calls core theory-theory. He writes:

> If belief-desire psychology has a central principle, it must link belief, desire and behaviour. It could be formulated like this:
>
> [Action principle] An agent will act in such a way as to satisfy, or at least to increase the likelihood of satisfaction, of his/her current strongest desire in the light of his/her beliefs. (Botterill 1996: 115).

Naturally there are other such core principles and general background assumptions which, if theory-theorists are right, provide the basis for the successful ascription of propositional attitudes. In this regard, Loar is right to claim, " ...[that the] principles we normally rely on in our commonsense use of the theory" are *a part* of the standard scheme of commonsense psychology (cf. Loar 1981: 128-132). But it is important to note that to regard commonsense psychology as a kind of theory on this basis is driven by methodological concerns

rather than ontological ones. Here the theory-theorist is concerned with what goes on either side of the black box of the mind, not what occurs inside it. Thus Loar is also right to distinguish, "...two components: the belief-desire theory and the contingent principles" (Loar 1981: 133).

The idea that these principles form a theory gains further credibility from the fact that such generalisations have the same outward form as scientific generalisations in other domains. Here are two such examples borrowed from Paul Churchland:

(a) If a bar magnet is placed under a flat surface on which iron filings are evenly spread, then, if there are no other significant forces on the filings, they will shift into a pattern of whorls orientated on the line of the bar.

(b) If a sulphur-phosphorous match is struck, then, barring vitiating factors, it will burn. (P.M. Churchland 1970: 222).

The important thing to notice, in all three cases, is that the, "...generalization[s] purport to describe a regularity in nature" (P.S. Churchland 1986: 300). To further cement this association the Churchlands have noted that the propositional attitudes have structural affinities with theoretical predicates used elsewhere in the natural sciences. Thus Paul Churchland compares propositional attitudes such as '[X]...believes that p' with scientific expressions such as '[X]...has a length in metres of n' (P.M. Churchland 1979: 101, 1981b: 208, 1988a: 64, cf. also P.S. Churchland 1986: 304).

2.2. Theories and Conceptual Schemes

On their own, these comparisons hardly constitute conclusive grounds for thinking that commonsense psychology is a theory. Whatever outward affinities that exist between the ascriptions and generalisations of propositional attitude psychology and those of the natural sciences we must recognise the possibility that there are also significant differences. I have argued elsewhere that there are indeed such relevant differences which threaten the methodological claims of the orthodox version of theory-theory (cf. Hutto 1997: 70–72). Furthermore, in chapter six, I argue against the view that reason explanations are causal-explanatory (cf. also Hutto 1998a, Hutto 1999a). If they are not, then this undercuts the standard assumption which underpins theory-theorist claims about ontological commitments.

In this light, the most convincing argument of Paul Churchland's for thinking commonsense psychology is a theory emphasises the fact that our familiar mentalistic vocabulary (i.e. our talk of thoughts, feelings, and expectations) must be understood in the same manner as other theoretically embedded vocabularies

(i.e. talk of electrons, atoms and gravity). He argues that the way in which our mentalistic terms have meaning is no different from the way in which more explicitly theoretical terms do. In all cases, it is generated holistically. We are asked to accept that the meaning of ordinary psychological terms is constituted by their role in the network of sentences in which they figure. This is to accept the view that, "...the terms of commonsense psychology get their meaning from an implicit definition, or a network of sayings, platitudes even, connecting beliefs, desires, and the rest with typical causes, typical interrelations, and typical effects." (Blackburn 1991: 197). This account was outlined in the introduction where David Lewis' views are discussed. In adopting this view about the meaning of theoretical terms, Churchland is explicitly renouncing the empiricist view that meaning is somehow given in certain types of experience. Thus:

> ...if the meaning of our common observation terms is determined not by sensations, but by the network of common beliefs, assumptions, and principles in which they figure, then, barring some (surely insupportable) story about the incorrigibility of such beliefs, assumptions, and principles, our common observation terms are semantically embedded within a framework of sentences that has all the *essential properties* of a theory (P.M. Churchland 1979: 37, emphasis mine).

Yet while adoption of a broadly holistic view of meaning apparently allows the eliminativist to argue that in certain essential respects commonsense psychology is at least theory-*like*, it is not sufficient to show that it is a theory *per se*. It is only theory-like in virtue of its having a meaning-holistic character. However, since any and all concepts are arguably meaning-holistic, this line of argument is confused in its scope. If we accept that meaning holism applies to all conceptual terms, and that theoretical terms are a sub-set of conceptual terms, then it follows that all theoretical terms are meaning-holistic. Putting things in this light makes it clear why Churchland's remark about the essential properties of theories in the above quotation is off target. It does not follow that all conceptual schemes are theoretical, or even theory-like in any interesting sense, simply because they are meaning-holistic. Theoretical terms are a sub-set of the conceptual. Consequently even if we accept that the terms of commonsense psychology share this property in common with explicitly theoretical terms we must not call it a theory on these grounds alone.

Interestingly even some theory-theorists find Paul Churchland's use of the term theory objectionable in important respects. At times he moves, without argument, from acceptance of a modest form of theory-theory to the more dubious view that everyday psychological discourse is committed to a programmatic, continual refinement of its basic categories. In other words, he illegitimately claims that commonsense psychology is, "...a stagnant or degenerating research programme

and has been for millennia" (P.M. Churchland 1989: 8). This is surely false. It assumes commonsense psychology is an explicitly developing theory. But, as Blackburn says, adoption of such a view would be, "...overplaying anything gained by introducing the word 'theory' [into the debate]" (Blackburn 1991: 198, cf. also Preston 1989: 291, Horgan 1993: 294, Sterelny 1993: 308).

2.2.1. Minimal Coherentism

Neverthless even if we bar the insupportable research-programme interpretation, and even if we assume that the arguments of the theory-theorist are, at best, inconclusive, eliminativists can still threaten commonsense psychology. If we accept the idea that conceptual terms are meaning-holistic, and we adopt the weakest version of the attendant coherentist epistemology, it follows that nothing guarantees the truth of any particular conceptual scheme.[6] This is all that is needed in order to motivate questions about the potential revision and possible elimination of commonsense psychology. Eliminativist concerns can be generated even if we appeal to nothing stronger than a widely accepted picture of our epistemic condition; the view that human beings receive uninterpreted data from the external world *via* the senses. Such data is then categorised, or put under particular conceptions of *how the world is*. It is important to note that although eliminativists are committed to the idea that we conceptualise our experience in various ways they do not think that our concepts and categories are unalterable or fixed. They deny that the categories we use are either innate or that knowledge is simply given in experience. Quine helped secure this view for scientifically minded philosophers by overthrowing the positivist's analytic-synthetic distinction and concluding that, "...no statement is immune to revision" (Quine 1953: 43). For, whereas prior to this empirical foundationalists believed that knowledge could be given directly in experience, most modern-day coherentists and scientific naturalists accept that deciding what counts as knowledge or a valid judgement is, "...a rather more turbulent affair" (P.S. Churchland 1986: 281). In the main, today's philosophers of science accept this point and are critical of the positivistic view that raw perception can provide a neutral basis for the assessment of observation statements and can thereby act as the ultimate testing ground for the truth of theoretical statements. As Patricia Churchland notes even the, "...seemingly self-evident datum that the Earth is stationary is falsified by Newtonian theory" (P.S. Churchland 1986: 293).

Following Quine, the evaluation of the truth of any given statement must be carried out with reference to its fit with other statements which are not, at least currently, under question. To accept even this much coherentist epistemology commits one to the view that our implicit background assumptions play a crucial part in how we evaluate the truth of a given statement. Even in its weakest form,

coherentism forces us to conclude that, "...science meets experience, not sentence by sentence, but in large blocks of theories and laws" (Rosenberg 1985a: 6). The idea that our knowledge, including our scientific knowledge, forms a revisable framework seemingly destroys the distinction between absolute fact made true by experience, on the one hand, and mere interpretation or human projection, on the other. This is why Quine believed that any and all facts are open to alternative interpretation and why he famously proposed that, "The totality of our so-called knowledge or beliefs, from the most casual matters of geography and history to the profoundest laws of atomic physics or even of pure mathematics and logic is a *man-made* fabric which impinges on experience only at the edges" (Quine 1953: 42).

This fits with the stronger coherentist claim that our background beliefs can affect the way in which we interpret our experiences. If our conceptual judgements meet experience as a body, and we must allow for possible revision at any place within that body, then we must even allow for the possibility of changes in our verdicts about what is experienced itself. Paul Churchland demonstrates just how greatly theory can affect our observations. For example, he asks us to focus on the fact that although we believe the solar system is heliocentric we nevertheless see the heavens as fixed and stationary above us. He outlines a method by which we might actually alter the quality of our perceptual experience so that we can *see* the movement of the stars and *feel* the motion of the Earth below in a way which is consistent with received cosmological theory (P.M. Churchland 1979: 30–34).

It is a short step from accepting these coherentist claims to endorsing the further claim that our conceptual categories are essentially revisable. Thus in a vein even more radical than Quine's, the Churchlands want to overthrow every last residue of empiricism; including the assumption that knowledge is partly grounded experience, even at the edges (cf. Fodor & Lepore 1996: 146, Churchland and Churchland 1996, P.M. Churchland 1998). It is because conceptual schemes develop and change over time (and moreover, because others living now deploy different conceptual schemes – even if not radically different ones), that we must accept that our categories concerning 'what there is' are plastic, pliable and mutable. Indeed it is precisely because our conceptions of the world are not fixed and unchanging that we can make the sort of rare conceptual shifts that constitute the progressive march of science and the growth of knowledge. Hence even if commonsense categories could be shown to be importantly non-theoretical, the Churchlands still have it within their means to motivate eliminativist concerns by appeal to a weak coherentist epistemology. For even against this coherentist backdrop alone it is reasonable to suppose that commonsense psychology might be on the list of potentially dispensable conceptual schemes.

For example, if we forsake the idea that the truth of a conceptual judgement can be underwritten by a particular type of experience then there can be no special

defence of commonsense psychology on introspectivist or merely experiential grounds. If we admit that our introspective or experiential judgements (such that there are any) are informed by our training in the use of concepts, then introspection cannot be a neutral judge. We must accept that, "...an introspective judgement is just an instance of an acquired habit of conceptual response to one's internal states" (P.M. Churchland 1981b: 208). But if there is no justification of conceptual schemes by appeal to any unique kind of experience then we must also conclude that the integrity of our conceptual categories is dependent upon their more general virtues. This is why the Churchlands frequently speak of the global excellence of theory as the deciding factor in theory choice (or, in Cussins' words, the means by which to 'rule out' a discourse). In securing this point, the eliminativist hopes to have established that commonsense judgements and conceptual frameworks have no special or sacred status when it comes to understanding action. It is on these grounds that they parade the possibility that commonsense psychology's, "...principles are radically false" and its, "...ontology an illusion" (P.M. Churchland 1981b: 210). While I agree that the credentials of commonsense psychology, like any other conceptual scheme, must be fought for and not simply assumed, and I agree that they cannot be defended by appeal to introspection, I argue, nevertheless, that commonsense psychology does have a privileged place in our self-understanding. This case will be made in the final chapter. Yet that argument is a long way off, and in order to prepare the battlefield properly we must consider the second meta-condition of the eliminativist argument more fully.

2.3. Major Conceptual Revolutions

The next platform for serious consideration of the potential elimination of commonsense psychology is a general acceptance that theoretical revolutions of great magnitude are live possibilities. Getting us to accept this is vital to the eliminativist strategy for it softens us up to the possibility that:

> A detailed neurophysiological conception of ourselves might simply displace our mentalistic self-conception in much the same way that oxidation theory (and modern chemistry in general) simply displaced the older phlogiston theory of matter transformation (P.M. Churchland 1979: 5).

In this regard the eliminativist must convince us of two things. Firstly, they must convince us that the fact that we are profoundly accustomed to understanding each other in terms of commonsense psychology is 'historically accidental, culturally idiosyncratic, and epistemically irrelevant'. Secondly, they must convince us that our resistance to change and revision in matters psychological can, in no way, be a measure of commonsense psychology's internal virtues. To

achieve these ends the eliminativist turns to examples from the history of ideas. The rejection of the Ptolemaic cosmological system in favour of the Copernican theory is an oft-cited historical case in which a theory was displaced despite considerable public and intellectual resistance. In fact, this has become the classic case of a revolutionary conceptual shift. It has two features of interest. First, the revision required a re-thinking which, at the time, would have seemed as initially inconceivable to the informed public as the theoretic shifts which the eliminativist asks us to consider. For example, consider Milton's reaction, as described by this quotation from the introduction of *Paradise Lost*.

> In Milton's case we are presented with the interesting phenomenon of a mind apparently uncertain to the last which of the two systems, the Ptolemaic or the Copernican, was the true one, or perhaps being persuaded by the higher probability of the Copernican, but yet retained the Ptolemaic for poetical purposes. For Milton's life (1608–1674) coincides with the period of the struggle between the two systems (Masson 1897: 26).

Secondly, it is also a case in which the conceptual shift was of no small importance to human progress. It was not a mere theoretical tinkering with our views on cosmology but a change which had ramifications throughout many levels of society and its practices. This point is worth emphasising as several philosophers have attempted to defend commonsense psychology on the sole grounds that it is nothing more than a practical or heuristic device, and hence makes no serious ontological claims. I believe this defensive posture to be useless. For, even if we swap 'conceptual scheme' for 'theory' in the following quotation, Churchland is right to suggest that, "Theory is regularly an intimate part and constituting element of people's second-by-second practical lives" (P.M. Churchland 1993: 218). Given that theories (conceptual schemes) and practices are often tightly linked, he is also correct in thinking that, "...practices can be displaced as well as theories" (P.M. Churchland 1993: 218). Moreover the former are often displaced due to alterations in the latter. So it was with the change to Copernican cosmology. It was a change which not only affected our view of the heavens but also everyday agricultural and navigational (not to mention literary) practices.

The main point is that if we are open to the possibility of such a tremendous change in our conceptions and concomitant practices the eliminativists will have prepared us for the third and final stage of their project. That is: to convince us that we have good reasons, in particular, to maintain that commonsense psychology has been, or will be, superseded by a vastly superior theory within its own explanatory domain. It is at this point, as Blackburn says, that the eliminativists, "...retain responsibility for showing how the undermining [of commonsense psychology] works – in round terms" (Blackburn 1991: 209).

3. Changing our Minds

A major reason why eliminativists predict doom for commonsense psychology is because they believe it is in the process of being undermined by revolutionary advances in cognitive science. We are asked to believe that cognitive science is in the throws of an internal revolution akin to that which took place in cosmology during Milton's time. However there is one crucial difference between the two cases in that the revolution within cognitive science, unlike the Copernican revolution, purports to be an ontologically radical theory change. Eliminativists are right to say, "...it would be something of a joke to suggest that Copernicus and Galileo showed that the planets Ptolemy spoke of do not exist" (Ramsey *et al.* 1991: 95). That is to say, whereas the entities described by the Ptolemaic cosmology (i.e. planets) survived intact when cosmological theory went the way of Copernicus, the cognitive revolution is one which threatens to be eliminative in character (cf. Sterelny 1993: 314). Such a change is allegedly on par with other internal theory changes which have resulted in drastic changes in our view of 'what there is' – such as the loss of caloric fluid, phlogiston and their ilk.

3.1. Connectionism: The Champion of Eliminativism

The weakest version of this line of thinking, to which I shall not give much attention, is tied to the claim that we should anticipate that a completed neuroscience will someday displace commonsense psychology as a means of explaining and understanding human behaviour. Given that we know very little about the form of completed neuroscience, such talk can be nothing more than an assertion of faith in the march of science. Opponents of eliminativism have a right to require its proponents to 'put up or shut up' if this is all they have to offer (cf. Hannan 1993: 172). Paul Churchland, aware of this, no longer rests his arguments for eliminativism on such a claim. He once wrote:

> The anticipated super-framework [mature neuroscience] suitable for reducing the P-theory [intentional psychology] (or for explicating its instantiation in us, or for displacing it entirely) is not yet within our grasp, nor can we claim to discern its outlines with any notable clarity (P.M. Churchland 1979: 118).

Times change. That was Paul Churchland's remark of 1979. Today he, and others, believe they have found the makings of the champion theory, or super-framework, in the form of connectionism. More recently he writes, "Here in 1993, we do have some very specific and highly promising 'replacement' concepts under active exploration." (P.M. Churchland 1993: 220). This is important, for allegedly:

> what makes certain kinds of connectionist models genuinely revolutionary is the support they lend to a thorough-going eliminativism about some of the central posits of commonsense (or 'folk') psychology (Ramsey *et al.* 1991: 93).

The hope of many of today's eliminativists is to marshal evidence which will establish that commonsense psychology is already in the process of being displaced by a superior theory. The source of this hope is the revisionary theories which have been sweeping through the artificial intelligence and cognitive science communities; specifically, connectionist theories.[7] Connectionists employ models of information processing which consist of a layered network of nodes. Crudely, the layers consist of an input layer, an intermediate or hidden layer (which can be, in itself, many layered), and an output layer of nodes (see P.M. Churchland 1989, Bechtel and Abrahamsen 1991: ch. 2). It is worth noting that in connectionist networks, "...it is the *entire* network that manages to make the relevant discriminations" (P.M. Churchland 1989: 299, emphasis mine). More importantly, such networks do not process information by making transactions over discrete, atomic structures by means of logical rules. Instead, they deploy what can be best described as context-sensitive representations. The details of this form of cognitive processing will be discussed more fully in chapter four, section 2.2.1, but for the moment it is sufficient to note the difference between this account of the cognitive architecture and the more traditional theories.

3.2. The Information Processing Debate

Connectionism departs crucially from the traditional understanding of information processing which regards cognition as involving a series of computational transformations of structured symbols or symbol-strings. These are often thought to serve as an algorithmic, unconscious cognitive map. Such discursive maps are allegedly analogous to the those we use when we explicitly follow a set of serial directions in order to find a friend's home. The directions incorporate descriptions of features or objects of a particular environment and we respond to these by following a series of rules telling us what to do when we encounter these items. For example, such directions might be: 'Turn left at Bridge Street, go straight until you reach the red post box, then turn right on the High Street' and such like. The idea that symbol-manipulation of a similar kind is the essence of intelligence was self-consciously set out by Newell and Simon as a credible empirical hypothesis (Newell and Simon 1976). Clark articulates their claim as, "The strong-physical-symbol (SPSS) hypothesis. A virtual machine engaging in Von Neumann-style [serial] processing of standard symbolic atoms has the direct necessary and sufficient means for general intelligent action" (Clark 1989: 12). It is important to note that such theorists regard symbols after the fashion of the theoretical entities

found in any other foundational or special science. Hence they compare the ontological status of symbols to that of germs, tectonic plates, and the like. The basic entities of the science of the mind are symbols and the empirical claim is that 'if there is to be intelligence then there must be symbol-manipulation'.

3.3. Representationalism and Expansive Classical Cognitivism

As the above example suggests, classical cognitivists are standardly committed to the idea that the symbols manipulated in cognition represent worldly features and objects. Noting this is crucial, for without an explicit commitment to representationalism the information or data processed in a computer would have to be regarded as nothing more than meaningless symbol-strings. The versatility of computers derives from the very fact that various well-formed strings of 0s and 1s can be given alternative semantic interpretations. Bearing this in mind, mere symbol-strings, which constitute the information processed, do not intrinsically represent anything (cf. Searle 1990). It is for this reason that, in its unembellished form, the classical account of information processing is inadequate as a basis for a properly cognitive science. As Fodor has argued, cognitive science ought to be committed to the view that agents are organisms with internal, causally efficacious mental representations. This is because adoption of some form of representational theory is the only plausible means by which we can properly account for important phenomena such as; language and concept learning, perception, and decision making.

He argues that it is essential to any decision theory that the subjects are capable of choosing between a number of possible actions based on a spread of preferences in advance of taking action. But to do so a cogniser requires some means of representing states of affairs in order to choose, "...between possible (but) non actual outcomes" (Fodor 1975: 35). One might ask: Why can't this be done in the subject's natural language? Fodor's reply is that even those lacking natural language can take such decisions. Moreover he maintains that when we perceive, learn concepts and learn language we must do so by relying upon a representational medium. For example, he claims that when, in the process of learning its first concepts, a child sees what it takes to be an apple it advances the hypothesis that 'x is an apple' and its teacher informs it whether or not its conjecture is correct. His point is that to get to the stage of hypothesis-confirmation the child must have some means of couching its guesses prior to learning the concept or linguistic term. If this were not the case then the child would be unable to make any initial guesses at all. Thus he tells us that, "...concept learning [essentially involves] inductive extrapolation [which] commits one to postulating a representational system in which the relevant inductions are carried through" (Fodor 1975: 41, 1998: 124). Crucially, he also claims that,

"...no alternative view of concept learning has ever been proposed" (Fodor 1975: 41).

In precisely the same way, he tells us that in order to learn a first natural language the subject must have knowledge of a prior language of equally expressive power. Against objections that this embarks language learners on an infinite regress he says that the representational system that enables us to learn language is not itself a learned language; it is innate. His analogy, not surprisingly, is with the in-built machine 'language' of computers which provide them with a basis to develop other languages. Likewise he claims that in order to perceive some property of the environment the information received must be first detected and then re-described for use by the organism. This too requires a medium for representation. In all our best theory of mind must include an account of representational capacities because only by appeal to representations can we possibly explain several central cognitive activities. Essentially, Fodor has never changed his mind about this. He recently remarks "RTM [representational theory of mind] remains the only game in town, even after all these years" (Fodor 1998: 23).

Although some of Fodor's original arguments rely heavily on the controversial idea that all learning is a form of hypothesis-confirmation, his argument that cognitive science requires representations of *some kind*, for explanatory purposes, is well made and it is supported on other grounds. For example, without appeal to the representational contents it would be difficult to explain trivial events such as the regular appearance of a class of students in the same room at the same time, week after week, as anything more than a remarkable coincidence (cf. Hannan 1993, Sterelny 1993, Clark 1993a: 224). This has the interesting consequence that, whatever its final form, cognitive science must make an advance on a pure or minimalist version of theory-theory.

Defenders of theory-theory define mental states in terms of, "...causal relations among stimuli, hypothesised mental states, and behaviour" (Stich 1983: 18–23). But if cognitive science is to serve its particular explanatory ends it must commit to the idea that the hypothesised mental states of concern have, or are related to, representational contents. For example Stich draws attention to the fact that although such an account of mental states might achieve success in describing concepts such as pain, he insists that, "...there are a number of important differences between mental states like having a toothache, and mental states like believing *Socrates was wise*, differences which make it difficult to see how the theory-theory [functionalism] can provide an account of belief at all." (Stich 1983: 23). Fodor makes the same point in the following fashion, "Although functionalism tends, in general, to license Realistic readings of psychological ascriptions, you need *more* than functionalism to license Realism about propositional attitudes" (Fodor 1981c: 20). In other words, what comes between

the typical stimuli that initiates intelligent behaviour must either be inner functional states that serve as representations themselves or states which are related to representations. Most theorists are attracted to the second option according to which cognitive states are viewed as having two aspects: a certain functional role and a related representational content (Stich 1983: 28, Loar 1981: 57).

3.3.1. Why Mentalese Representations?

Exploring the nature of representational contents opens up several possibilities. We might say an organism is related either to: a sentence in a natural language, an abstract object, or a symbol of an internal code. Fodor chooses the third option and takes the objects of propositional attitudes to be tokens of a mental formulae. He dismisses the idea that we are related to tokens of natural language sentences on the grounds that non-verbal subjects, such as infants and animals, can entertain propositional attitudes. Like many others, he rejects the Fregean abstract object account on the grounds that it is, "...hard to see how entities that have no causal powers can play an indispensable role in a theory whose job is to explain actual physical motion." (Lycan 1988: 9).

In contrast, Fodor maintains, "A symbol (token) is a spatiotemporal particular which has syntactic and semantic properties and a causal role" and that, "...to have a propositional attitude is to have an occurrent mental symbol tokened *in the head*" (Fodor 1990b: 167, 1987: 17, 1975: 198). One virtue of viewing thought processes in this way is that the physical tokens of an internal code are allegedly mechanically realisable, thus they can enter into causal transactions with other physical objects and states. The symbols of this code are not only semantically evaluable, they also have logico-syntactic properties which link them inferentially to other semantically evaluable symbols and mechanically to other physical states of an agent's body. It is because these representations have logico-syntactical properties that they are ideally suited to take part in computational operations. These properties give them a kind of 'shape' analogous to the shapes of the syntactical forms of ordinary sentence tokens. The relation is no more than analogous because syntax does not reduce to physical shape, nor are the syntactical forms of ordinary language unambiguous in the way that the symbols of mentalese are supposed to be. Nonetheless in order for representational contents to make a difference their interactions must be mediated by their formal (or purely logical) properties. Tim Crane makes this point well.

> The idea is that since it is in virtue of some intrinsic property – its 'shape' – that a symbol has effects and since there is no difference in syntax without a difference in shape, then the syntax of the [language of thought] is actually the causal mechanism for the interactions of intentional states (Crane 1990: 195).

Although Fodor isn't explicit about the way in which the formal aspects of mind are to be physically implemented, as long as the formal properties are intrinsic, there is no logical barrier to explaining this feature. For example, Crane elsewhere remarks:

> ...as modern computing vividly reminds us, sentences can be stored and transformed in patterns of electrical impulses – the possible 'shapes' of sentences are not restricted to these familiar ones. The [language of thought] hypothesis holds that sentences can be stored in the electrochemical states of the brain. In fact, the phenomena of modern computing provide more than an analogy to this aspect of the [language of thought] hypothesis: they provide part of its inspiration (Crane 1990: 189).

This encourages the hope that the existence of sophisticated reasoning processes would not violate the requirements of a mechanistic world-view. Fodor insists that:

> ...the operations of such a machine consist entirely of transformations of symbols; in the course of performing these operations, the machine is sensitive *solely* to the syntactic properties of the symbols; and the operations the machine performs on the symbols are entirely confined to alterations in their shapes (Fodor 1991b: 42, emphasis mine).

If each representational content is related to a uniquely shaped symbol-type, and well-formed symbol-strings form whole propositions, then each proposition could, in virtue of its unique shape, play an effective role in the inferential and causal processes of a purely mechanical mind. We are asked to imagine that, "...the syntax of a symbol might determine the causes and effects of its tokenings in much the way the geometry of a key determines which locks it will open" (Fodor 1991b: 41).

This image of a key is revealing. If we accept that each proposition has its own unique geometry (mirroring its unique formal properties) this could also explain some other characteristics of propositional attitude psychology. It could account for the fact that it isn't possible to make semantic substitutions into belief contexts. Using the lingo of computationalism, if we swapped the symbol that stands for 'water' with a symbol that stands for 'H_2O', we would alter the representational content of a given symbol-string containing it. Such a substitution would make a difference to the overall content of the representation and inferential patterns of response even though 'water' and 'H_2O' refer to the same type of external stuff. According to the mentalese story, the 'shape' of the symbol which denotes 'water' would need to differ from the shape of the symbol which denotes 'H_2O' if they

were to be distinguished in thought. To put the point in more familiar Fregean terms, the fact that propositional contents are referentially opaque sits well with the idea that symbols must have unique shapes and the idea that when they are combined with other symbols according to well-formed rules, the symbol-strings they compose can form whole representations. Fodor pointedly suggests that the best means of explaining Fregean sense is by appeal to, "...mental representations – syntactic structures in the language of thought" (Fodor 1994a: 24, 1998: 37–39).

In light of these considerations, the language of thought version of computationalism is attractive for the very reason that it promises to undermine, "...Brentano's claim that the intentionality of mental states precludes their being physical" (Loar 1981: 3). It is against this background that Fodor confidently asserts that, "...the cost of not having a Language of Thought is not having a theory of thinking" and reminds us that "Remotely plausible theories are better than no theories at all." (Fodor 1987: 147, 1975: 27).

3.4. Propositional Attitude Psychology and the Language of Thought

Even more crucially, it is precisely in the above mentioned respects that Fodor sees the potential for defending propositional attitude psychology by appeal to his version of expansive classical cognitivism. He draws the parallel in the following way:

> To begin with, LOT [the language of thought hypothesis] wants to construe propositional-attitude tokens as relations to symbol tokens....Now all symbol tokens have intentional contents and their tokens are physical in all known cases. And – *qua* physical – symbol tokens are the right sorts of things to exhibit causal roles. [But remember] that these mental states that have content also have syntactic structure – constituent structure in particular – that's appropriate to the content they have (Fodor 1987: 135–137).

To reiterate, the language of thought hypothesis postulates the existence of inner mental symbols which have both representational capacities and formal structures. It would appear that propositional attitude psychology requires just such an account of the nature of representational contents and formal processes.[8] Its mode of explanation fits a very simple, but useful, model of means-end reasoning as idealised by the practical syllogism which Aristotle articulated.[9] On a metaphysical footing, to see how action depends on the existence of logical roles and representational contents let us consider some central features of unreduced rational decision making. Take, for example, the schema as it is outlined by Loar. He writes, "...a person X's reasons for an action A [can be represented] as follows" (Loar 1981: 6–7):

X desires that q;
X believes that if p then X's doing A will lead to q;
X believes that p.

What is important to note about this type of rationalistic explanation is that we are only able to fully explain why X acts in *this* way as opposed to *that* by filling in the propositional contents (p) and (q) in relation to X's various psychological attitudes (i.e. beliefs, hopes, fears, etc.). Inferential processes can be much more sophisticated than those described in the above example. A thinker may have all sorts of fears, doubts, hidden desires and other psychological attitudes. A subject might also have various degrees of confidence in propositions. Thus they might only partially believe or desire certain things and naturally this too will affect decision making. While all of this is of interest, it presents no obstacle to forging a link between the logical aspects of propositional attitude psychology and a classic computational theory of mind. It may be that beliefs and desires form a complicated network, and they might need to be variously weighted (as decision theory suggests), but, in any case, rational thought processes would still require the re-occurrence of logical structures (the tokens of various symbols). In the end, it appears that the rational aspects of our psychology crucially depend on the logical roles played by our thoughts. This gives us reason to think that the classical version of the computer metaphor could provide a naturalised account of our ordinary decision making processes. *Prima facie*, the logical aspects of propositional attitude psychology fit rather snugly with the classical account of information processing to which the language of thought hypothesis is committed.

While this is so, the main point with respect to the connectionist-classical cognitivist debate is that it is only by also committing to representationalism that classical cognitivism has any possibility of providing a scientific vindication of commonsense psychology. In Fodor's words, "...common sense would be vindicated if some good theory of the mind proved to be committed to entities which – like the attitudes – are *both* semantically evaluable and etiologically involved" (Fodor 1987: 26, emphasis mine).[10] His version of expansive classical cognitivism, combines a linguaform theory of structured representations with a traditional serial symbol manipulation account of information processing. In marrying these two views Fodor promotes the idea that the representational system which underwrites our cognitive activities is a language of thought composed of syntactically, well-formed representations.

We are now in a position to see the chain of reasoning that provides a basis for Fodor's potential defence of commonsense psychology. It rests on a number of crucial assumptions. These are:

(i) commonsense psychology is fully and properly characterised as propositional attitude psychology.

(ii) propositional attitude psychology is ontologically committed to entities similar to those posited by the language of thought hypothesis.

(iii) the language of thought hypothesis will be endorsed by modern cognitive science.

4. The Connectionist Threat

So what is the nature of the perceived connectionist threat to expansive classical cognitivism that has excited the hopes of eliminativists? Many connectionists object to the claim that the essence of intelligence boils down to symbol-manipulation since they do not accept that basic information processing involves symbols at all. This debate has apparent consequences for the very unity of cognitive science as a multi-disciplinary investigation into the nature of mind insofar as its unity has historically assumed the truth of the traditional account of symbolic processing (with input from linguistics and philosophical logic). However if the core assumption of cognitive science is only that the 'mind is a computer', then as Von Eckardt suggests, connectionism is no threat to it. With reference to Boyd's vision of scientific metaphors, which encourage investigation of the unknown as opposed to directing attention to the hitherto unnoticed, she regards cognitive scientists as making a literal claim about the nature of mind (Von Eckardt 1993: 116). Minimally what drives the analogy is the simple fact that computers are information processing devices. But note that this leaves it open how we are to understand the nature of information processing. Connectionists are therefore no less committed to the project of cognitive science than those Clark has dubbed classical cognitivists. This is because the target of the connectionist challenge is only the received view of the nature of information processing which has dominated the field since its inception. As Horgan puts it, the real debate here is between those advocating different theories of 'psychotechtonics' which he defines as, "...scientific theorizing about cognitive 'functional architecture'." (Horgan 1993: 285).

The heralded merits of connectionism in this domain are generally viewed as fourfold and I simply mention them here as they are well documented by other authors (cf. Clark 1989, Sterelny 1990: 175, Bechtel and Abrahamsen 1991: 56–65). Firstly, the symbolic structures posited by classical cognitivists are not robust. That is to say, since the alleged units of information are dependent upon their syntactic shapes any physical damage to those structures or any overloading of the system drastically and rapidly affects the cognitive capacity of the system (i.e. it crashes). Connectionist systems, in contrast, are more likely to suffer from

what is called 'graceful degradation' if damaged physically or overworked. They will begin to operate sub-optimally and deteriorate slowly. Compared to connectionist systems, classical cognitivist systems are the more fragile.

Secondly, classical architectures are deemed to be biologically implausible because in such systems even the simplest intelligent acts depend on a quite sophisticated array of rules and representations. But we know evolution does not programme systems in advance. In keeping with the biological facts, connectionist networks develop discriminatory power slowly over time and, broadly speaking, they are capable of 'learning' from experience.

Thirdly, classical computational systems are very poor at performing simple perceptual and recognitional tasks. As they only perform serially operations, they demand a large number of time consuming symbol manipulation steps in order to complete these tasks. In this they compare unfavourably to the relatively speedy and efficient processing times achieved by connectionist networks. Consequently connectionist systems tend to meet 'real time' psychological constraints on many sorts of recognitional tasks better than classical cognitivist systems.

Finally, because classical cognitivism is ontologically committed to the existence of internal syntactic structures, it is commonly argued that, "...classical architectures are insufficiently attuned to the fact that our brains are built of neurons" (P.M. Churchland 1989: 194, 1, 56, 159–163, 181–188). Once again, in contrast, a chief merit of the connectionist networks is that, to a greater degree than their classical rivals, they resemble the neural configurations that have actually been discerned within the microstructure of brains.

4.1. Assessing the Connectionist Threat

How should we judge this competition? The key question is: Should connectionism challenge our faith in our commonsense psychological categories and explanations? The putative threat arises because the connectionist revolution in cognitive science has led to widespread scepticism concerning the previously unchallenged truth of classical cognitivism. But, as we have seen, it is vital in assessing the substance of this threat that we draw a distinction between classical cognitivism understood as nothing more than a theory of information processing in terms of serial, symbol manipulation and a version with more expansive commitments involving a theory of mental representation.

Expansive classical cognitivism, as epitomised by the language of thought hypothesis, can be characterised as advancing the following claims (and, in some cases, a fourth hypothesis about innate categories):

(1) Propositions are the fundamental molecular units of cognition.

(2) The constituent parts of propositions have representational capacities and syntactic structure.

(3) The syntactic structural parts of propositions, in virtue of their unique shapes, underpin cognitive processing.

Only to the degree that connectionism actually threatens the expansive reading of classical cognitivism, does it threaten the prospects of a scientific vindication of propositional attitude psychology. But, in this respect, it is not clear that connectionism will emerge victorious. There are two possible manoeuvres open to the expansive classical cognitivist: one weaker, one stronger. Firstly, it can be argued that connectionists have to do more than provide a better story about information processing. They must establish that this story is incompatible with classical accounts of symbolic representation such that these cannot retained even if we abandon the classical account of information processing. Secondly, it can be argued, as Fodor has, that there are reasons to think that any decent theory of cognition needs to posit linguaform representations. This, in turn, could provide leverage for the claim that a classical account of information processing cannot be so readily abandoned since it is required to provide the vehicles of linguaform representation.

Ironically, given the pretext of this debate, there are serious problems with the second, stronger reply which stem, not from its attachment to a classical account of information processing, but from its inability to provide a naturalistic account of representational content. Thus, appearances to the contrary, the language of thought hypothesis offers no real projection for commonsense psychology. Exposing the real problems with Fodor's account will be the focus of the next chapter.

5. Conclusion

We have so far reviewed the meta-argument of eliminativism and examined the basis for their claim that connectionism threatens the credibility of commonsense psychology. However this conclusion depends on several controversial premises. The first is that the fate of commonsense psychology is linked to the fate of classical cognitivism. The second is that connectionism is decisively superior to classical cognitivism. With respect to the second claim, it has been the burden of section four to show that connectionists need to do more than parade the information processing virtues of their theory if they wish to win the war against the classical cognitivists. Nevertheless, in the next chapter, I review some familiar reasons for thinking that expansive classical cognitivism, as epitomised by the language of thought hypothesis, fails to provide a scientific vindication of

propositional attitude psychology. This is because it fails to provide naturalistic accounts of mental causation and representational content. This critique is an important prelude to the more positive investigation of the prospects of naturalising commonsense psychology in the central chapters. It is will not be until chapter four that I openly challenge the first premise in order to defuse the threat connectionism apparently poses.

CHAPTER TWO

A Lack of Content

"Psychologists have no right to assume that there *are* intentional states unless they can provide, or anyhow foresee providing, or anyhow foresee no principled reason why someone couldn't provide, naturalistic sufficient conditions for something to be *in* an intentional state."

– Fodor, *The Elm and the Expert,* 1994

1. Introduction

There are two internally related problems with Fodor's attempt to provide a scientifically respectable linguaform version of propositional attitude psychology. These are the problems of causal impotence and content. In this chapter I review the nature of these problems in order to lay the ground for a proper assessment of the prospect of naturalising commonsense psychology by other means. Specifically, in section two I aim to establish that no headway can be made with respect to the problem of causal impotence in the absence of a naturalistic theory of content. This demand is made of any theorist who accepts that content is, even in part, externally determined – including those who endorse accounts of narrow content. In section three I review some familiar reasons for thinking that crude causal theories of content do not provide a sound basis for naturalising content. I conclude the section with a critique of Fodor's asymmetric dependency theory. Finally, in section four, I examine Dretske's *prima facie* more promising teleofunctional approach but argue that because it relies on an information-theoretic base it too fails to determine what fixes content.

2. The Problem of Causal Impotence

Imagine that McX spies a glass of water and forms the true indexical belief that "This is water". Imagine also that McX is quite thirsty and that upon forming this belief he moves to satisfy his thirst. It is natural to suppose that by appealing to the

content of McX's belief we can explain his subsequent water-seeking action. To put the point counterfactually, we might say of McX: 'Had he not believed that the glass in front of him contained water he would not have tried to drink from it'. But herein lies the problem for the expansive classical cognitivist. Either the content of McX's belief about water is responsible for his drinking or his syntactically structured inner states are. On the one hand, if we say that both are responsible then his behaviour is over-determined. On the other hand, if McX's syntactic states are solely responsible for his action then their content is causally and explanatorily redundant. This is because, viewed merely as computational states, they are not intrinsically about water or anything else. The counterintuitive result is that there is no need, and indeed no room, to appeal to the content of McX's beliefs when designating the cause of his action. All that is required in order to causally explain (or predict, or control) behaviour is to determine which syntactic tokens are produced by the agent's information processing machinery (in relation to certain typical stimuli) and to figure out how they, in turn, causally interact with other such tokens so as to produce action (cf. Stich 1983).

By considering this example it becomes clear that the problem of causal impotence stems from two conspiring sources. The first is the view that a proper science only deploys causal explanations. The second is an acceptance of the formality condition which states that, "...in doing psychology....it's the fully opaque reading which tells us what the agent has in mind, and it's what the agent has in mind that causes his behaviour" (Fodor 1981c: 239). Famously this is why Fodor says that the, "...acceptance of the computational theory of mind leads to a sort of methodological solipsism" (Fodor 1981c: 232). Recall also that according to the classical account of information processing what the agent has 'in mind' is *nothing but* a series of syntactically structured symbolic states which interact in virtue of their shapes. Put all of this together and the result is that the syntactic structures, not contentful properties, do all the work in the production and explanation of action (cf. McGinn 1989: 135).

Given these assumptions, cognitive science should at best only trade in psychological laws of a narrow kind. That is to say, its laws should only make reference to the formal, or syntactic, features of cognition. Or, to put the point so as to highlight its threat to the explanatory value of the language of thought hypothesis, the causal laws of cognitive science ought not to mention content (even if we are prepared to recognise its existence). Information conceived as uninterpreted strings of code is both necessary and sufficient to generate and explain behaviour. The Churchlands also note this consequence by saying, "Computers are formal machines, in the sense that they operate on symbols in virtue of the form of the symbol, not in virtue of how the symbol may be interpreted" (Churchland and Churchland 1990: 303). They go on to note that, "...the machine goes from one state to another because it is caused to do

so....without reference to whatever semantic properties our cognitive states may or may not have" (Churchland and Churchland 1990: 303). Therefore, even if contents were mapped in a unique way onto particular syntactic structures, they could contribute nothing to the production and causal explanation of action. Even if propositional content exists, on such a picture, it is epiphenomenal.

Stich invokes a vivid picture to illustrate the tension. He suggests that the computational stories about the mind are initially compelling because we, "...suppose....that human mental sentences are written on little cerebral CRTs, and that the sentences in the language of thought look, for all the world, like tiny English sentence tokens." (Stich 1983: 53). Accordingly, sententialists become realists about beliefs. They regard them as subject to existential quantification:

(∃t) (t is a token of 'Ouagadougou is the capital of Upper Volta' and Stich believes that t). (Stich 1983: 39).

By imagining that the tokens of the relevant English sentences are inscribed on tiny CRT units in our heads we come to think that these sentences can play the relevant causal roles in virtue of their contents. Moreover they do so in just the way described by the defender of the language of thought hypothesis (Stich 1983: 37). All of this is fine, thinks Stich, until we stop reifying contents and abandon the pretence that English sentences somehow fill our heads. For, "On dropping the pretence....[we must ask] what is the relation between the English content sentence and the object of belief which, *ex hypothesis*, is a sentence in the language of thought?" (Stich 1983: 41–42). The relation of which he speaks is the relation between computational syntax and representational content. But for Stich that is no relation – at least, not one of interest to the cognitive scientist. Ultimately, we must pass the verdict that if we are to vindicate propositional attitude psychology we need apparatus other than that offered by a mechanistically respectable computationalism – expansive or otherwise. Some theorists when faced with this dilemma opt for, "...the important and rather surprising result [that]....a belief's semantical properties *per se* contribute nothing to the belief's causal role" (Lycan 1988: 22). They find that, "...the honest answer is the one Stich gives: Meaning doesn't do any work, so forget it" (Cummins 1991a: 124).

In contrast, Fodor wants to construe explanations of our actions by appeal to psychological laws that explicitly require mention of the contents of our beliefs and desires. He wants the laws and generalisations of psychology to be contentfully-construed. Crudely, an example of such a law might be: 'if a subject, X, believes that 'I am late', and desires not to be late, then X will act so as to avoid being late'. For Fodor this is a hedged predictive law since it assumes that X is not irrational, that X does not have any stronger counter-veiling beliefs and desires, and that nothing else prevents X from acting. Importantly, if the laws of

our best scientific psychology didn't quantify over contents they would offer no safe haven for propositional attitude psychology.

2.1. The Tension between Externalism and Strong Supervenience

The worry is: How to keep a science of the mind concerned with content-based psychological laws in step with one that is only interested in causal explanatory computational processes? The problem is especially gripping because Fodor, like Stich, accepts that any scientifically respectable explanation must be of a causal character. Essentially, his methodological solipsism plus his commitment to physicalism adds up to a commitment to a version of Stich's principle of autonomy.[11] The principle of autonomy states, "...psychological states and relations, or at least those that should be of concern in a serious scientific psychology, must supervene on the current, internal, physical states of the organism(s) in question" (Stich 1982: 197). Supervenience (in some varieties) is thought to provide a means of describing the nature of a non-reductive relationship of dependency between two types of properties, aspects or predicates.

The consequence of accepting such a principle is that it promotes the view that scientific psychology must only concern itself with the immediate causes of action located inside the skin of the agent. However Fodor maintains that the language of thought hypothesis only presupposes that the implementation mechanisms are classically computational. He claims that contents have an internal level of implementation upon which they strongly supervene. Nonetheless he also argues that contents are no more robbed of their causal powers by the fact that they are implemented by means of brain activity, than chemical compounds are robbed of their causal powers by the fact that they are implemented by interactions at the atomic level. In this way, he sees a scientific version of propositional attitude psychology as having the same kind of status as any other higher-order special science.[12] His hope is that if content strongly supervenes upon computational states, which in turn strongly supervene upon brain states, then it might be possible to see how the laws of commonsense psychology, which make appeal to contents, could be implemented computationally and efficaciously. He tells us that:

> This is not, by the way, any sort of epiphenomenalism; or if it is, it's patently a harmless sort. There are plenty of cases in the respectable sciences where a law connects a pair of properties, but where the properties the law connects *don't figure in the story about how the law is implemented.* (Fodor 1990a: 286).

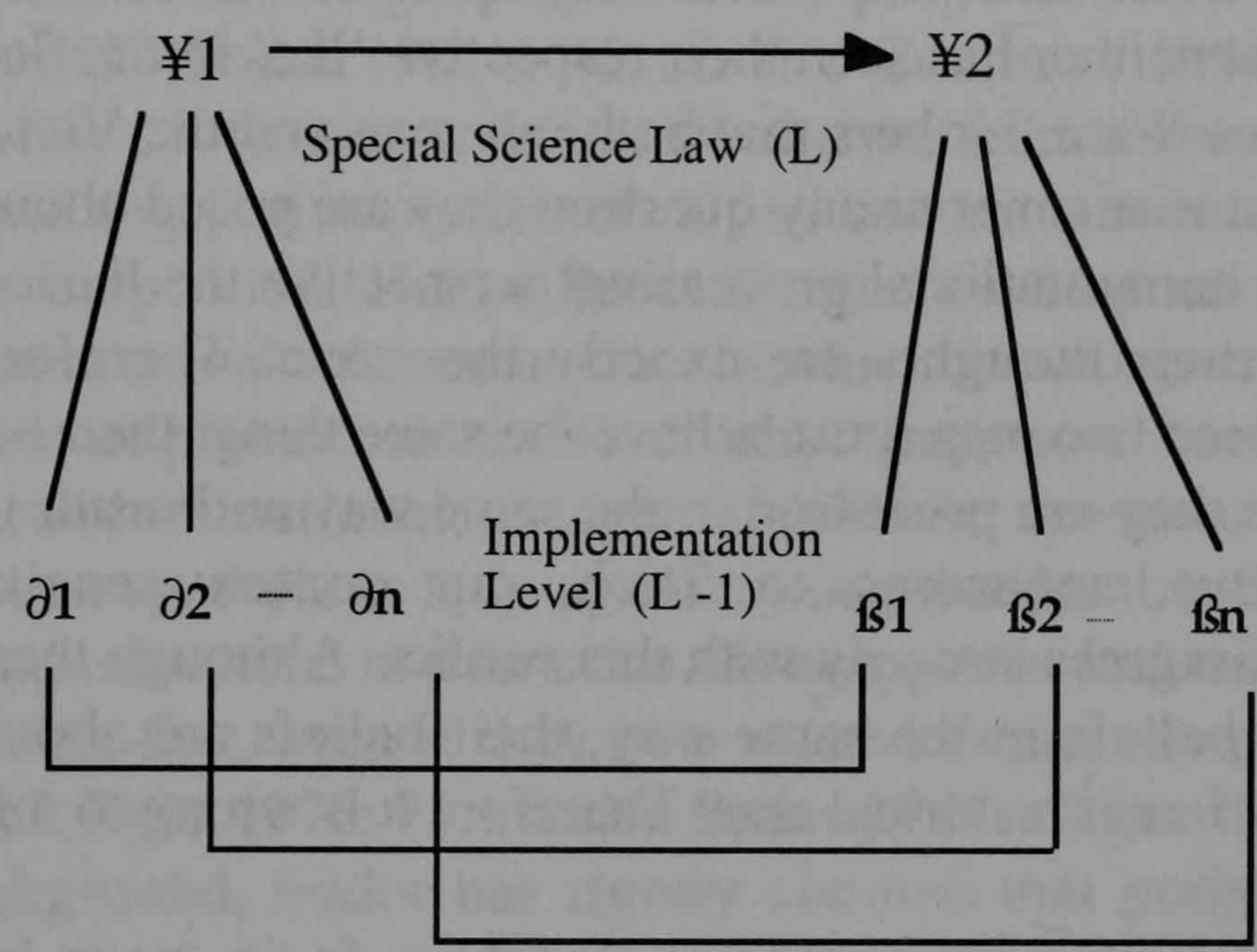

Figure 4. *Fodor's implementation levels*

The problem with this reply is that content is unique among higher-order phenomena in a way which makes it incapable of strongly supervening on internal states of any kind. The tension is illustrated by Twin Earth thought experiments. The standard Twin Earth scenario invites us to imagine a situation wherein two physically identical twins are confronted by a body of water and caused to form the indexical belief that would, if translated into English, appear in their respective 'belief boxes' as the mentalese sentence "This is water" (see Putnam 1975a). Unsurprisingly, given their identical internal make-up and assuming their cognitive processes strongly supervene on this make-up, all of their water-oriented behaviour is exactly the same. They are each as inclined or disinclined as the other to bathe in it, drink it, and so on. The twist is that one of the twins originates from Earth and his 'water-thoughts' have been formed in relation to genuine H_2O, whereas his döppelganger hails from Twin Earth where the chemical make-up of the 'water' is XYZ. The problem for the language of thought theorist is this; it would seem that the content of the twins' so-called 'water-thoughts' is not the same since one is thinking about H_2O while the other is thinking about XYZ. Put in the same Twin Earth environment, the Earth-born subject will have a 'false' thought about water, while the other will have a 'true' thought. This is because the contents of their 'water' thoughts differ. But from an internal point of view they are exactly the same, as are all their outward responses to the liquid.

Another example from Stich makes the same point in marginally less fantastical form. Consider the case of a contemporary of ours and a Victorian chap who are both associated with two different politicians from their own historical eras. Furthermore imagine that these historically distinct thinkers associate the name 'Ike' with these politicians who happen to have exactly the same limited set of

characteristics. For this reason both subjects think "Ike is tall", "Ike likes cigars" and so on. But neither has seen their respective 'Ike' in the flesh. So perfect is the match between these thinkers that both our man and the Victorian will answer in exactly the same manner to any question they are posed about 'Ike'. The point is that from the computational processing perspective the liaisons generated by the structure of their thoughts are exactly the same. Therefore from an internal perspective these two men must believe the same thing: their beliefs are exactly the same because they are processed in the same way and result in the same outward responses. But, according to Stich, our context-sensitive commonsense psychology disagrees strongly with this verdict. Although these two thinkers may process their beliefs in the same way, their beliefs are about politicians of two completely different historical eras. Therefore it is wrong to think that they believe the same thing.

By considering such scenarios a *prima facie* case is made against the possibility of the strong supervenience of semantic content upon the internal properties of the agent. This has persuaded many to adopt an externalist view of content. Externalism is the thesis that content must be, at least in part, determined by its extension. This creates a dilemma for Fodor's approach. For if one accepts that content, is even in part, determined by extension, then it cannot strongly supervene on the internal states of organisms. But if content does not strongly supervene on internal states then it cannot be causally efficacious. This is bad news for the possibility of solving the causal impotence problem by appeal to computational implementation. How might one respond to this? An obvious move would be to abandon externalism. For example, one might advocate a wholly internalist account of content. But monsters lie in wait on this path too since it is demonstrable that content externalism is a necessary ingredient in a sound propositional attitude psychology.

2.2. Why Commonsense Psychology is Externalist

First and foremost, any satisfactory internalist view of content must be more than a merely syntactical account. As Stich points out, with reference to Fodor's metaphor of shape, "...having an inscription of the right shape on his belief-CRT can't be all there is to Otto's believing that it's raining." (Stich 1983: 44). If mere syntax determined content then two creatures with tokens of the same syntactic type in their belief boxes would, for this reason alone, have identical contents. This would be so even if their syntactic structures had different inferential connections and caused the execution of different actions.

A seemingly more plausible, and reasonably popular, approach is to endorse an account which identifies content with functional or procedural roles. But advanced as a serious theory of content this position has tremendous problems. The most

grievous is that, "...the semantic properties of the propositions themselves are assumed, not *explained*" (Fodor 1991b: 46, cf. also Cummins 1991a: 122, Loar 1982: 274). In my view confusion on this score derives from a sloppy use of the all too convenient and suggestive picture of reified content sentences 'filling the head', which Stich lampooned. Hence Putnam rightly warns us to beware of implicitly equating conceptual role semantics, in which meanings are assumed to be in place (and not explained in terms of anything more basic), with the scientifically respectable functional, or procedural, role semantics which only treats relations between symbols as a, "...syntactic or procedural notion" (Putnam 1988: 54). Conceptual roles, read in a functionally acceptable fashion, do not yield an account of believing-that, even if they help us understand the processes that underlie believing (cf. Field 1978: 79, cf. also Fodor 1981c: 26).

Against this background, Fodor has rightly claimed that purist versions of functional role semantics fail in a straightforward way because it is a mistake to confuse the inferential roles of symbols with their content. For example, he says of theorists in artificial intelligence who support this approach that, "...they talk as though the semantics of [machine language] constituted a theory of language-and-the-world, whereas, in fact, it provides only a theory of language-and-the-insides-of-the-machine" (Fodor 1981c: 209). They fail to match our workaday intuitions that content is world-directed. According to propositional attitude psychology, to have a belief is to have a belief that *something*, where the *something* is expressed by a sentence or a proposition. Thus it is better to regard these proposals as advocating a 'no-content' syntactic theory of mind as Stich explicitly does. For good or ill purely internalist accounts of content are not compatible with commonsense psychology.

2.2.1. Two-Factor Theories

It is not surprising that those who have learned their lessons about pure forms of internalism have, in response, tended to favour the view that content has more than one aspect. In its simplest form this thinking inspires holistic two-factor theories of content. Two-factor theories attempt to account for the nature of content in terms of both internal functional roles and referential relations to the world (cf. Loar 1981, Block 1986b).[13] They recognise that, "...conceptual organization does not *on its own* explain, in general, our assignment to [X]'s sentences of truth-conditions or references" (Loar 1982: 275). In fact, although he condemns the current fashion, Fodor admits that the, "...two-factor story is currently the standard version of semantic functionalism" (Fodor 1987: 81).

Nevertheless there are reasons to doubt that any such theory can provide a naturalistic vindication of propositional attitude psychology. As Fodor notes they are, "...inescapably infected with holism and are therefore incompatible with the

working assumption that the laws of psychology are intentional" (Fodor 1994a: 6). He rails against such accounts because they are unable to support the required predictive generalisations. His worry concerns the way they construe the identity conditions of content. For according to such accounts two contents are identical if and only if they share all the same functional and inferential connections *as well as* the same extensions. The upshot of this requirement is that cases of identity are very, very rare. Bearing this in mind, Fodor highlights two illustrative problems for such a view.

First, such accounts are bad news because attitudes get fused with content whereas propositional attitude psychology requires that contents be logically distinct from the functionally defined psychological attitudes. For example, if McX acts because of his beliefs and desires about p, then the content of p-tokens must be distinct from their functional roles. If not we would have no principled means for thinking that a belief-that-p had anything more in common with the desire-that-p, than it would have with the belief-that-q. Rampant holism results in each psychological state having its own, unique conditions of individuation. It is because propositional content can be isolated that different psychological attitudes can be directed at the same propositional content. This underwrites the possibility of believing, hoping, fearing, desiring, or wishing that *p*.

Similarly holistic accounts of meaning have obstructive drawbacks, if read as an extreme forms of the meaning-is-use thesis. That is to say, in a good deal of linguistic communication truth-conditions must play a quasi-independent role in the determination of meaning. We need to disambiguate between the sense in which our words mean something and sense in which we mean something by our words (cf. Davidson 1984: 274, Searle 1969: 18–19). Recognition of this distinction is vital because, "There is a real danger that talk of meaning being plain and manifest in 'use', so interpreted, may make it appear that meaning is *not* a genuine abstraction from use." (Bilgrami 1986: 119). This is important, as most philosophers of language stress, because it is only if meanings are autonomous with respect to use that the same sentence, or proposition, can be employed to give a command, ask a question, tell a joke, make an assertion or warn a friend. It is precisely because, "...the same proposition can be common to different kinds of illocutionary acts [that] we can separate our analysis of the proposition from our analysis of kinds of illocutionary acts" (Searle 1969: 31).

Moreover, if contents were fused with attitudes, decision theory and rationalistic psychology would be impossible. The practical syllogism only works if the *same* content is related to different kinds of psychological attitude (e.g. beliefs and desires). Since two factor theories threaten even this degree of autonomy with respect to content, Fodor insists that the identity of propositional contents cannot depend, even in part, on their functional roles. He advises that we take seriously a functional theory of *believing* but not a functional (as opposed to a

denotational) theory of *believing that p*. In his own words, "... one can, of course, combine *accepting* the Functionalist ontology with *rejecting* the reduction of belief/desire explanations to functional ones" (Fodor 1990b: 10, cf. also 1987: 70, 71, 76, 105).

Secondly, it must be possible for a single subject to relate to tokens of p, with the same content, on more than one occasion – even if their other inferentially related thoughts about p have changed. But given a two-factor theory of content it would be unlikely that anyone would think *precisely* the same thought twice or that two people would ever entertain *exactly* the same proposition. In order to preserve a rationalistic realistic propositional attitude psychology, contents should not be thought of as fused with functionally defined psychological attitudes nor with other contents in their doxastic surround. This is crucial to note because, "For a Realist, the interesting thing about intentional content is its ability to survive variations in functional role. And this relative independence of content from functional role provides a prima facie reason for doubting that the former notion will reduce to the latter" (Fodor 1987: 71). Thus Fodor concludes that two-factor theories of content must be rejected.

Conversely he proposes that we should only countenance a single-factor, purely denotational theory of content. In this regard, he has been given to say such things as, "...there is nothing at all to meaning except denotation" (Fodor 1990b: 162). According to such a theory, all content is determined solely by extension and is to be assessed in bivalent terms. Unlike holistic two-factor theories of content, "...[Denotational] theories are, on the face of them, *atomistic* about content" (Fodor 1994a: 6).[14] Being atomistic, they open up the logical possibility of specifying the nature of such contents in principled and non-sceptical terms.

However any move in this direction openly raises the spectre of the causal impotence problem again. To take the denotational line is to be clearly committed to externalism about content and hence the problem of explaining the strong supervenience of content upon putative internal events. But to abandon the idea of such supervenience is to abandon the prospect of a naturalistic vindication of propositional attitude psychology (assuming some version of the principle of autonomy remains in force). So an attempt must be made to resolve this tension rather than avoid it.

2.2.2. Narrow Content

Initially Fodor hoped to do just this by appeal to narrow content. At one stage he claimed that the laws of a scientific psychology could be construed narrowly but not so narrowly that they would simply be computational laws. Commitment to computationalism, as we have seen, drives one to favour the idea that the causal-explanatory laws of a scientific psychology must be developed at the level of

procedural role. This is unacceptable to Fodor and defenders of a scientific version of commonsense psychology because it means that such laws would not involve contents at all. Such an empty scientific psychology would miss valuable psychological generalisations. Hence he and others introduced the idea that a scientific version of content-ascribing psychology ought to employ narrow contents (cf. Fodor 1987, 1990b, Block 1986b). This proposal was meant to keep some form of content-based psychology in the frame while remaining faithful to methodological solipsism and externalism about content *proper*. Note that the introduction of this construct was itself a revisionist manoeuvre. It raised the possibility that a scientific version of commonsense psychology, "...may well diverge from folk psychology in its criteria of individuation of psychological states" (Bernier 1993: 330). Fodor describes the intellectual climate which first led him to propose the idea in the following terms:

> It is obscure how externalist intentional laws could be computationally implemented. Very well, then, let there be *another kind* of intentionality – let there be, as one says, 'narrow content' as well as 'broad' content – such that narrow content is ipso facto *not* externalist (Fodor 1994a: 17).

But, for all its promise, there were always difficulties in giving a positive clarification of the nature of narrow content. It is normally defined negatively as the type of content which is independent of the way in which subjects are related to their environment. It does not concern reference or truth conditions. On the other hand, it isn't reducible to a mere syntactic or functional role. Fodor's technical and slightly more positive way of defining it has been to say that narrow content is a function (mathematically conceived) from contexts to truth conditions. But, however it is defined, it was supposed to be 'another kind of intentionality'. It was introduced to enable us to class behaviours as psychologically identical even when the propositional attitudes of subjects initiating those behaviours had different wide contents.[15] Thus the psychological similarity between physical twins allegedly lay in the fact that their internal states enter into the same partial function when their thoughts map onto the environment, even though the truth values of their thoughts may differ.[16]

For our purposes, it is important to note that the only way to characterise a narrow content as a partial function in relation to a context, is by first tying it to a context and specifying a denotational content. Fodor openly admits this when he talks of 'sneaking up' on narrow contents (Fodor 1987: 50). Notoriously, he pronounced that, "Narrow content is radically inexpressible, because it's only content potentially; it's what gets to be content when – and only when – it gets to be anchored" (Fodor 1987: 50). That it is only possible to speak about narrow contents indirectly reveals that one cannot fully explicate the nature of such content without consideration of the subject's environment at some stage. Thus even if

narrow contents turned out be useful to a scientific psychology our understanding of them would depend on having an account of denotational content.

2.3. All Roads Lead to Externalism

In the end it makes no difference to the overall success of Fodor's project whether or not he chooses to defend the idea that content-ascribing scientific psychology requires narrow contents. What is important is that if he wants to make legitimate use of content-based laws he must provide a naturalistic account of denotational content. As Bernier notes, "...in order to properly understand the notion of narrow content we must be able to say what is that relation between a mental type and an environment, such that it constitutes a context of a thought" (Bernier 1993: 332). But this is unavoidable in any case. In order to defend *any* scientifically respectable version of propositional attitude psychology we are inexorably led to accept some form of externalism. Consider that:

(a) Two-factor theorists regard content as partly determined by external factors.

(b) Narrow contents must be externally anchored in order to be content at all.

(c) Denotational accounts regard extension as the only factor which determines content.

Of late, Fodor's preferred strategy has been to accept (c) and attempt to marry it directly with a classical account of computational processing. He remarks:

> I used to think that the problem of constructing property theories to connect intentional laws with their computational implementations simply *demands* that intentional laws be construed as narrow. But I am now inclined to think that it doesn't and that a purely informational semantics can be reconciled with a purely computational theory of mental processes after all (Fodor 1994a: 17).

But, again, whichever path one chooses it is clear that, "...a serious intentional psychology must presuppose the *naturalizability* of content." (Fodor 1994a: 5). In order to deal adequately with the causal impotency problem, the big question which must be answered is: Can representational content be naturalistically explained? This is the real hurdle.

3. The Problem of Content

Naturalistic defenders of propositional attitude psychology are committed, in Field's words, to providing a, "...materialistically acceptable account of what it is for a sentence or sentence-analog in a system of internal representation to mean that p" (Field 1978: 85, Fodor 1991b: 44). Without a convincing semantic story computationalists will have only succeeded in drawing mere analogies between our reasoning processes and computational processes: they will have produced genuine as opposed to scientific metaphors.

The onus is on them to explain how internal states can be about objects in the external world using only naturalistic tools. Fodor has claimed that only two sorts of naturalistic theories have been advanced to explain representation: resemblance and causal theories (cf. Fodor 1990b: 33, Cummins: 1991). Resemblance theories are inadequate because they wrongly presuppose that the relation between represented and representation is symmetrical. Moreover they cannot accommodate the distinction between universals and particulars (Fodor 1990b: 33). Causal theories, on the other hand, appear better placed to provide a naturalistic account of representation. For instance, the distinction between universals and particulars can be explained by the fact that causal laws trade in types, but only tokens of those types actually enter into particular causal relations. Interestingly, in this respect, it is an advantage of the causalist approach that naturalists are initially concerned with pre-linguistic mental representations. For such an account has, "...a better chance of working for mental representations than it does for (e.g.) English words" (Fodor 1987: 100). This is because we do not regularly utter token English sentences such as 'There is a tree' whenever our visual systems are tickled by the presence of trees. Nevertheless it is possible that a token mental representation of a tree is produced in such circumstances even if we do not report it.

3.1. Causal Theories of Content

Fodor is convinced that *some* version of causal theory will yield a naturalistic account of representational content. The crudest version relies on the notion of co-variance which is explained by appeal to causal laws. These can be understood as conditional statements of the following kind: if, in the right conditions, subject X perceives an object (or feature) of type Φ, then X's mental machinery will produce a token of type ¥ (*ceteris paribus*). This strategy is reminiscent of Grice's project. Consider his examples of natural meaning; e.g. smoke means fire, spots mean measles, etc. (cf. Grice 1989). The hope is that at least basic perceptual contents could be explained in terms of such reliable co-variance. If the crude causal theory worked then mental representation could be explained by an unmysterious

naturalistic relation. For, as Fodor claims, "Causal relations are natural relations if anything is" (Fodor 1990b: 33).[17]

Before considering this proposal in more detail, it is important to make a caveat about the crudity of such accounts. When causal theorists discuss their ideas they openly confess that they are using, "...tinker toy example[s]" (Fodor 1994a: 13). For instance, they standardly gesture at cases of simple visual stimulation. Hence we find Fodor pointing to the inner tokening of a mental representation /dog/ when confronted by an instance of 'doghood' – say, a golden retriever. It is not that causal theorists don't recognise that there are many aspects of content which will not fit the simple model, rather it is their hope to show that, minimally, these primitive kinds of representation can be naturalistically explained (Fodor 1991c: 256, Fodor: 1998: 121, Akins 1996: 340-341). If this can be achieved they will have broken the back of the problem of getting content into the world-order. Assessing these cases is important because if causalism falls at this hurdle then its aspiration of naturalising more sophisticated contents is completely forlorn.

But, even at this level, causal theories are riddled with problems. Indeed it is generally acknowledged that crude causal accounts are simply non-starters. One fatal question is: How can such approaches designate the 'right' point in the causal chain, in a principled way, so as to fix the referent? How can this be accomplished using only causal criteria? If we attempt to fix this point by appeal to causal laws alone then mental representations will always denote the closest events in the causal chain. They will stand for things like retinal stimuli or neural excitations. No one thinks that this would be adequate as a theory of representational content (Jacob 1997: 97–98, 66).[18] In reply, the right kind of question to ask is, "Why do we see the elephant, not the firing of our retina?" (Jacob 1997: 66). The problem of fixing the right point in the causal chain is variously known as: 'the horizontal axis problem' (Godfrey-Smith 1989: 536) and the 'distality problem' (Jacob 1997: 66).

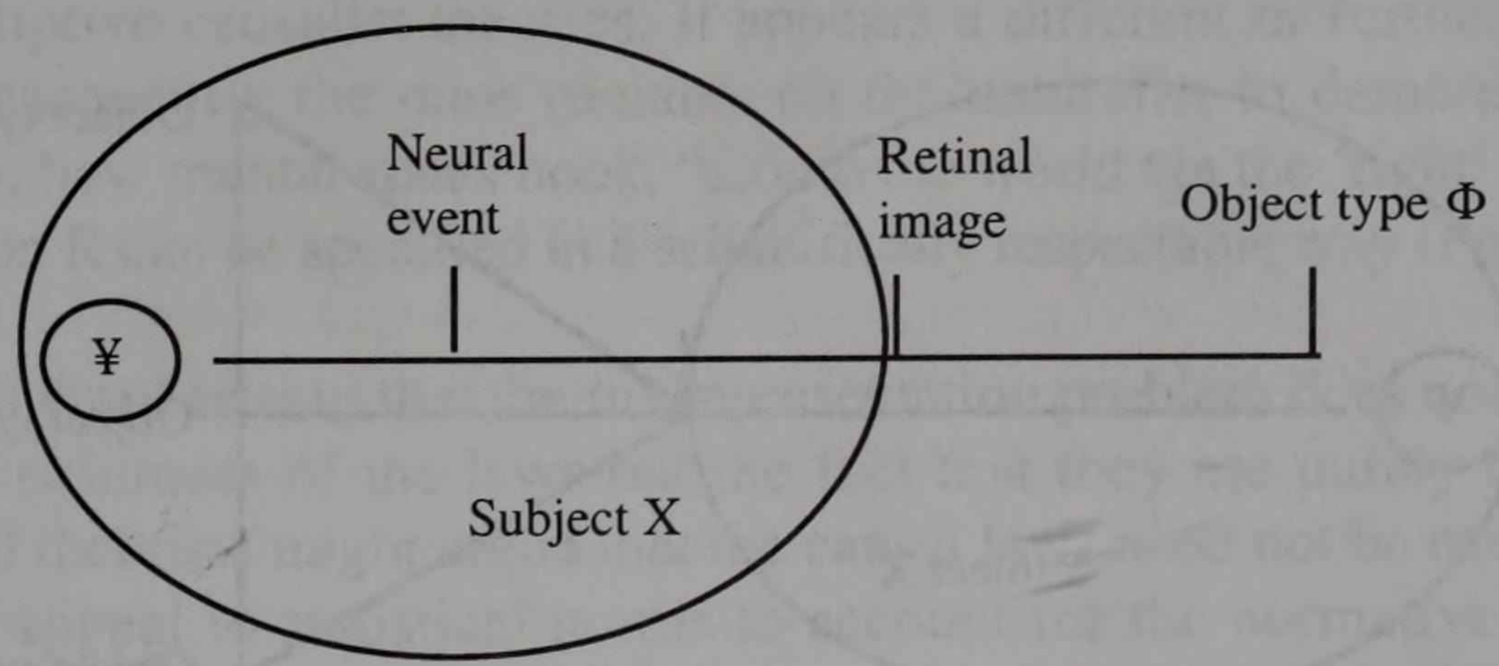

Figure 5. *The distality problem*

Advocates of causal theories content are really interested in distal stimuli, the non-local causes of representation, not proximate stimuli. Consequently they need to get outside the subject's head by respectable means. The standard move is called a 'broad' or 'wide' strategy because it explicitly relies on causal chains which extend beyond the skin. On this approach the causal relations of concern exist between the cogniser's representations and the external environment. In a nutshell, if certain features of the world are causally related to a subject 'in the right way' then its representations will be true, and if they are related 'in the wrong way' its representations will be false. The hope is that even if, "...internal manipulations do not determine reference....causal relations with the environment might" (McGinn 1991: 211, cf. also Fodor 1991b: 35, Jackson & Pettit 1988: 382). Sadly, for causal theorists, even if we ignore the distality problem, this account is unable to explain the doubled-edged nature of the representational relation. Causal relations lack an essential duality. In order to make the story work one needs a naturalistic means of specifying how causal relations can underpin the bivalent character of content which allows for the possibility of correct from incorrect mappings. This has become known as the problem of 'misrepresentation' or the 'vertical axis problem' in the literature. It is illustrated by figure 6.

The misrepresentation problem is embodied in the question: What establishes that a tokening of symbol type ¥ is true rather than false, in a given instance? It is demonstrable that any causal account which makes appeal to strict causal laws rules out the possibility of misrepresentation, and thereby by the possibility of representation. An easy and sufficient way to make this point is by appeal to the illustrative problem of disjunction. There are other complaints to be raised against the crude causal theory but I agree with Baker when she claims that, "...the disjunction problem, first identified by Fodor himself, is both deep and pervasive" (Baker 1991: 29).[19]

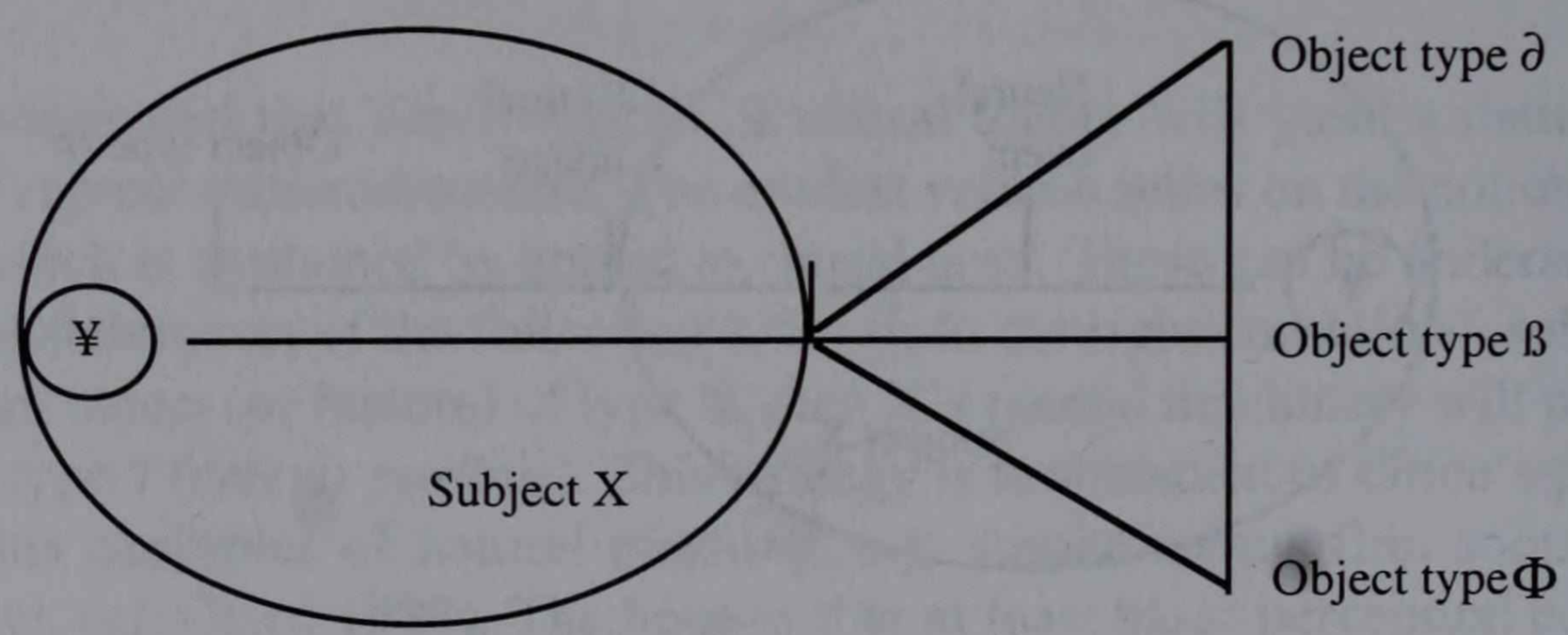

Figure 6. *The misrepresentation problem*

Consider the case in which a token of the symbol ¥ appears in my head because my perceptual apparatus is causally tickled by the presence of a dog. The crude causal theorist may wish to say that, in this case, my mental representation is true. But what makes it so? For what is different about the causal relation in this case as compared to one in which the token is incorrectly caused? This is pressing since a tokening of the symbol ¥ can be caused by many other things than the presence of *bona fide* canines. Imagine a case in which my symbol ¥ is causally connected *via* visual perception not to a real dog but to a dog-like sheep. If, *ex hypothesi*, it is only causal connection (as defined by strict laws) that determines how a symbol of type ¥ hooks onto the world, and it is the causal connections which also give ¥ its representational content, reference, and truth-conditions – then the symbol ¥ cannot represent *dog*; it must represent *dog or dog-like sheep*. This is the more accurate description of the psychophysical law in question. Unless the account is qualified or supplemented, ¥ must represent the presence of any and all of those other things. That is badly put. Strictly speaking, it is better to say it represents nothing at all. Thus a fatal consequence of crude causal theories is that they leave no room for error. There is no logical space for normative notions like right or wrong on a purely descriptive, causalist account.

Compare this case with another in the naturalistic domain which the causal theorist must treat as equivalent. Suppose that we have a law that predicts planetary orbits and we discover that Venus' orbit has, in one instance, deviated from that law. Would it make better sense to say that Venus has made a mistake or that our law was imperfect (or incomplete) and that a more precise law would account for the irregularity? As good law abiding naturalists we must prefer the latter option. Scientists cannot, and do not, blame the world for the shortcomings of their theories. It should be evident when reviewing these two cases that we cannot explain representation by a straightforward appeal to causal laws of any kind since such an analysis leaves no room for the normative notions of error/success and truth/falsity. Simply put, such notions cannot be milked from purely descriptive causalist theories. It appears a different or further relation is needed. Consequently the onus remains on the naturalist to demonstrate, non-semantically, how mental states hook, "...onto the world via the 'right' relation R" where relation R can be specified in a scientifically respectable way (Putnam 1990: 84).

It is worth emphasising that the misrepresentation problem does not essentially concern the strictness of the laws but the fact that they are purely descriptive. Crude causal theorists might argue that the causal laws need not be exceptionless. They might appeal to statistical norms to account for the normative dimension needed to explain semantics. On such a statistical causal account the token ¥ would be associated with whichever object was most frequently responsible for its tokening. But this amendment does not solve the disjunction problem. Statistically

speaking, the class of disjunctive stimuli are more frequently co-present with the tokening of ¥ and therefore more reliably co-varied with it (see Papineau 1984, 1987: 60–62, 1990: 36, cf. also Dretske 1986a). A principled statistical account will only be forthcoming if the problem of misrepresentation has already been solved.

3.2. Asymmetric Dependency Theory

It is recognition of this problem which led Fodor to propose his asymmetric dependency theory of content (Fodor 1987: 1990b). He advances the thesis that one class of distal stimuli, say type Φ, is dominant because, counterfactually speaking, if objects in this class did not cause tokens of ¥ then no other class of stimuli would. Thus if an item in another class of stimuli, type ß, prompts tokens of type ¥ in a subject it will only do so *in virtue* of the fact that items in the dominant class, type Φ, have that tendency. The asymmetric character of the dependency stems from the fact that the relation of dominance is one-way. Even though objects from the wild class can cause the appropriate inner state tokens, they would not do so expect for the fact that the objects from the dominant class do so. Hence, Loewer and Rey suggest that, "A useful analogy might be made with the phenomenon of imprinting: goslings are evidently born with a disposition to become bonded to the first animal they encounter....Once the imprinting occurs, misidentifications of an animal as the imprintee depend on the original correct identifications, but not *vice versa*." (Loewer and Rey 1991: xxvii, cf. also M. Williams 1998: 134). Fodor gives the following tongue-twisting example of the relationship:

> Cows cause /cow/ tokens, and (let's suppose) cats cause /cow/ tokens. But /cow/ means *cow* and not *cat or cow* because *there being cat-caused /cow/ tokens depends on there being cow-caused tokens, but not the other way around.* /Cow/ means *cow* because, as I shall henceforth put it, noncow-caused /cow/ tokens are *asymmetrically dependent upon* cow-caused tokens. /Cow/ means *cow* because *but that /cow/ tokens carry information about cows, they wouldn't carry information about anything* (Fodor 1990b: 91).

It is only because of this asymmetric relation of dominance that any given inner state has representational content. Misrepresentation is possible, in specific circumstances, due to the fact that some objects may cause the occurrence of particular types of inner states even though they do not belong to the dominant class of stimuli. Thus the crucial duality is explained. If the object which causes the tokening is a member of the dominant class then the representation is 'correct'; if not then it is 'false'. Notice that the asymmetric dependency relation must be characterised non-lawfully. Consequently it cannot be identified with a Humean

causal relation, even though they are necessary conditions for the existence of such relations. Figure 7, which is a modification of Cram's diagram, illustrates the fact that we are dealing with a different kind of relation. The asymmetric dependency relation is designated by the line labelled ADR and causal relations by the lines labelled CR (cf. Cram 1992: 59).

Although Fodor's account is not, in and of itself, a solution to the naturalisation problem, it does highlight its essential nature. It spells out the minimal conditions which any such theory must meet in order to be successful overcome the misrepresentation problem. To an extent Fodor acknowledges this when he writes, "The treatment of error I've proposed is, in a certain sense, purely formal...it looks like any theory of error will have to provide for the asymmetric dependence of false tokenings on true ones" (Fodor 1987: 110, cf. Jacob 1997: 99). It is purely formal because it is silent about how the asymmetric dependencies are, in fact, established. In absence of this vital detail it is of no use to the naturalist. For example, Kripkian baptismal accounts also fit this frame. According to Kripke, names gain their denotations through the practices of individuals or a community of speakers who engage in dubbing ceremonies. Names refer to the objects which speakers intend them to refer to. It is by appeal to such initial naming ceremonies that we can evaluate whether or not the names are being used properly or not (cf. Kripke 1972, Putnam 1975a).

There are two standard accounts of how such referential capacity gets established. One suggests that intentions fix the naming connections to the worldly objects by means of the underlying intentionality of an individual's thoughts. Another account suggests that it is by appeal to a social background that we are able to determine, in a culturally relative fashion, what glues the name to its object. The point is that on either construal a Kripkian-style causal account is non-naturalistic because it depends on the pre-existent intentionality of the speaker or that of a semantically accomplished community of speakers. Both versions presuppose what the naturalist seeks to explain because they, "...presuppose a conceptual backdrop" (Sterelny 1990: 118). In contrast, naturalism requires the reduction of intentionality to the non-semantic and the non-intentional (cf. Fodor 1990b: 32).[20]

Thus if one endorses the asymmetric dependency theory questions for the naturalist remain: How does one class of distal stimuli become dominant? What establishes and maintains this dominance? These questions must be answered without appeal to phenomena which are already infused with intentionality. At one point Fodor advances a two-stage account of how the asymmetric dependencies are established which makes *indirect* appeal to linguistic policies. He writes that, "...it's the linguistic policies of speakers that give rise to the asymmetric causal dependencies" (Fodor 1990b: 99). He also suggests that such, "Linguistic policies

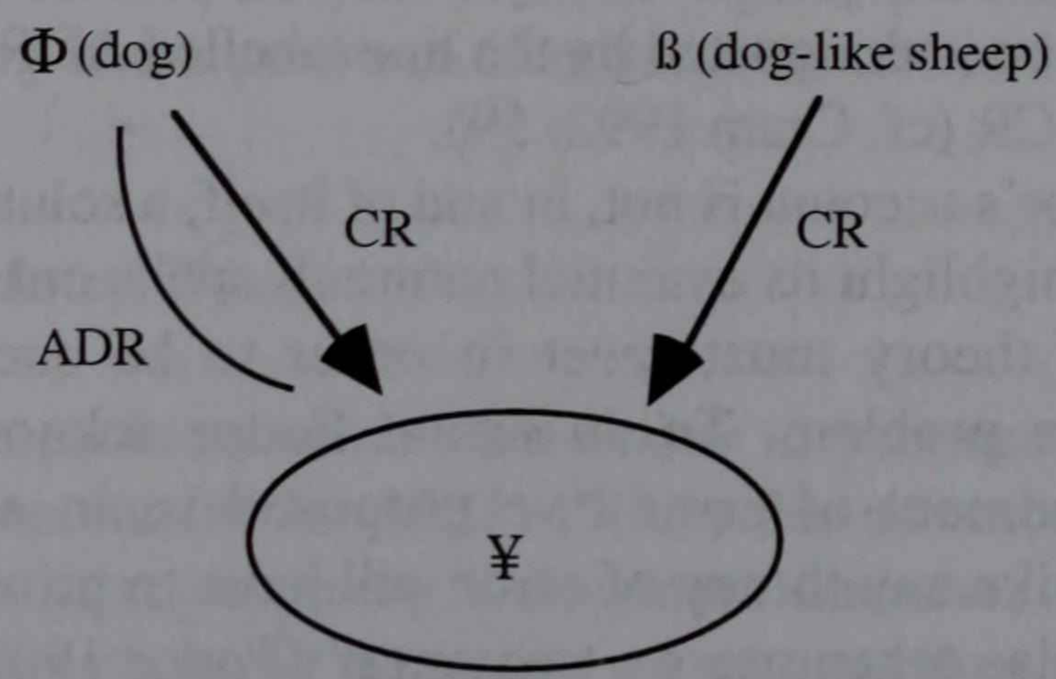

Figure 7. *The anomic nature of the asymmetric dependency relation*

don't make semantic relations; but maybe they make causal relations, and maybe causal relations make semantic relations. This, anyhow, is the hope by which informational semantics lives" (Fodor 1990b: 99). The idea is that pursuing a certain linguistic policy such as that of using the term 'slab' to make requests for such items depends on other linguistic policies such as referring to slabs as instances of slabhood. This latter policy sets up a number of standardised, regular causal patterns between mental representations in the minds of speakers and the objects they denote. Finally, Fodor maintains that it is only the causal patterns that matter to content (they just happen to be established by linguistic activity). It is not important, metaphysically speaking, how they were set up.

According to this 'one step back' approach, he believes the asymmetric dependency account might provide a solution to the naturalisation problem because it describes such relations without employing, "...intentional or semantic idiom" (Fodor 1990b: 92). He makes much of the fact that it is not necessary to *mention* the linguistic policies of speakers since, "...the conditions for robustness quantify over the mediating mechanisms, and so can be stated without referring to the policies; hence their compatibility with naturalism" (Fodor 1990b: 99). Nonetheless, although Fodor can successfully avoid directly employing the semantic idiom, he still fails to give a scientifically respectable explanation of the dependency relationship. It is not enough for a naturalist to be offered an account which is simply compatible with naturalism. What is required is an account which is not explanatorily hollow; one that provides a full account of the nature and origin of content (cf. Cram 1992: 62, Jacob 1997: 99, Cummins 1991a: 65). Given that his account is silent on these matters, at best what his asymmetric dependency proposal demonstrates is that purely nomological causal theories are unable to explain content. They lack at least one crucial relation which has the formal features of the kind Fodor illustrates.

Furthermore, in light of Fodor's larger project to explain central cognitive activities such as concept learning by appeal to mental representations, the asymmetric dependency account becomes viciously circular. It does matter how the causal patterns that give rise to the asymmetry were originally set up since, as was discussed in chapter one, section 3.3, he believes that our linguistic practices and abilities themselves depend on the pre-existence of a system of mental representation. But if our mental representations gain their content, in the way Fodor's new theory suggests, then they only do so *after* linguistic policies are in place. Thus given his other views Fodor, unless he relies on something other than linguistic policies to set up his asymmetric relations, he must face the problem of explaining which came first; the chicken or the egg.

3.3. The Rejection of Causal Theories

We have now come full circle. Fodor's recent work correctly describes what is required for a naturalistic account of representation. But, in doing so, it also dramatically reveals that it is not possible to give such an account of content in purely causalist terms (Boghossian 1991: 83, cf. also Loar 1991: 131, Cummins 1991a: ch. 4–6, Anthony & Levine 1991: 4). The moral is that causal theories of content cannot account for the normative aspects of content. We can only talk of deviance once we have a principally established norm. Despite the hopes of some naturalists, there is no 'metaphysical glue' (Putnam's term) strong enough to do the job of attaching symbols to reality in a way that makes them meaningful. This justifies Putnam's derogatory tone in the following passage: "...[the reference relation] is attached to R by causal connection not by metaphysical glue, they write. But this is in fact, just to say that R (causal connection) is self-identifying. This is to repeat the claim that a relation can at one and the same time be a physical relation and have dignity (the built-in intentionality, in other words) of choosing its own name" (Putnam 1990: 85). The problem is that it is not 'glue' that is wanted at all. What we require is an analysis of normative relations which is both illuminating and not question-begging.[21]

4. The Information-Theoretic Approach

Before moving on, it is worth considering another, *prima facie* more promising, naturalistic account. In his book *Knowledge and the Flow of Information*, Dretske acknowledges that the problem of misrepresentation also arises for information-theoretic accounts of content. To deal with it he draws an important distinction between informational and representational content.

4.1. Informational Content

Informational content can be understood in terms of the following technical definition (based upon a modified version of Shannon's theory of information, cf. Dretske 1981: 65).[22]

> A signal r carries the information that s if F = The conditional probability of s's being F, given r (and k), is 1 (but, given k alone, less than 1).

There are several things to note about this definition. First, the transmission of information depends upon narrowing down the possible ways of characterising a source so that it can be uniquely specified by a signal. This is what is meant by 's's being F having the conditional probability of 1'. The usual method of pairing down the possibilities, favoured by engineers and information scientists, is to encode information in binary fashion using a series of 0s and 1s. Hence if there are eight different possibilities, assuming the information receiver has no background knowledge, these can be uniquely coded by using a minimum of three binary strings. The possibilities can be divided into natural classes or grouped arbitrarily. Using either method, each possibility is given a unique designation which a signal can carry. For example, to classify eight possibilities one must use symbol-strings with at least three units in them. For example, the encoded possibilities might be identified as follows: 1 = 000, 2 = 001, 3 = 010, 4 = 011, 5 = 100, 6 = 101, 7 = 110, 8 = 111. The quantity of information carried by any one of these codes is said to be three bits given that it would take eight symbol strings with at least three symbols, corresponding to each possible state of affairs, in order to encode them unequivocally.

The same point can be made by considering a modified version of an example taken from Dretske. If there are eight possible places I might go on holiday and I wish to choose between them I could toss a coin in order to help me decide. Using this method, I would begin by dividing the eight possible destinations into two groups of four. Then, having selected one of those groups with the first coin toss, I would split the 'winning group' into two further groups of two. Finally, I would be left with only two possibilities. A final coin toss decides the matter.[23] Given the initial number of possibilities this process of 'binary decision making' would take a minimum of three moves to complete.

Vitally, it is this understanding of information that lies at the heart of Dretske's original notion of indication. This will be important to the upcoming discussion. He defines indication, or natural meaning, in the following way:

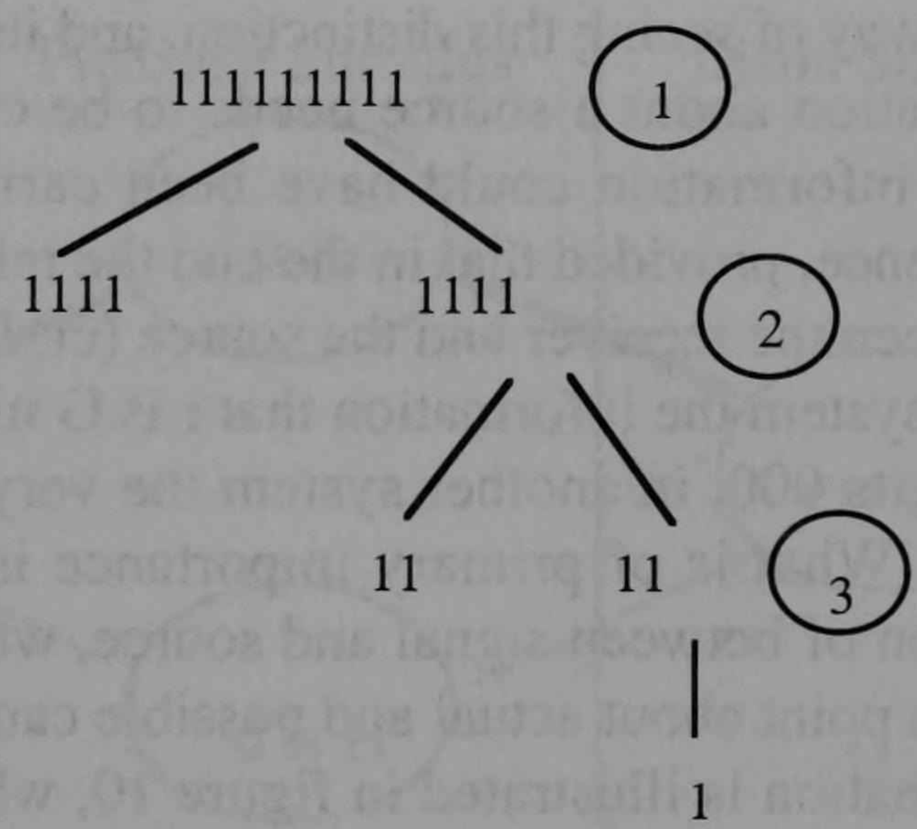

Figure 8. *The branches of binary decision making*

> ...if S (sign, signal) by being a, indicates or means that O is A, then S (or, more precisely, S's being a) carries information that O is A. What an event or condition (whether we think of it as a signal or not is irrelevant) indicates or means about another situation is the information it carries about that other situation (Dretske 1988: 59).

Second, information-theoretic approaches are distinct from causal theories in a way that enables them to deal neatly with the distality problem. Dretske is careful to distinguish informational from causal links, although he notes that the former depend on the latter. Consider that signals themselves are nothing but, "...physical process[es] which propagate from one place to another (light, electric current, sound waves, etc.) or a physical object suitably inscribed, that is transported from here to there (letters, books, pictures, etc.)" (Dretske 1981: 37). Nonetheless, while it is true that the flow of information relies on causal links, "According to this view, signals are one thing, the information they carry, their semantic content, another" (Dretske 1981: 40, cf. also 26, 38–39, 986b: 111). Dretske uses the example of a ghost channel in order to demonstrate the logical gap between these two kinds of relation. This is illustrated by figure 9, which is an adaptation of his own diagram (cf. Dretske 1981: 38).

The distinction can also be illustrated if we consider printed books. The information found in various copies of a particular book will be the same, given that they share the same source (e.g. the printing plates, proof copy). Each copy has a unique causal link to this source. Yet even though there is no interesting, direct physical or causal relation existing between them, both your copy and my copy share the same information. Hence it is possible to interrogate copy A for information about copy B, without going directly to the source. The distinction is important because it allows us to recognise the difference between the causal

vehicle of information, the signal, and the information itself which the signal carries. Another way of seeing this distinction, and its importance, is to realise that although information about a source needs to be carried by a particular causal route, the same information could have been carried by another causal route without consequence, provided that in the end the relation of perfect correlation is maintained between the receiver and the source (cf. Dennett 1997: 91-92, 95). For example, in one system the information that r is G might be unequivocal encoded by use of the digits 000, in another system the very same information might be encoded as 001. What is of primary importance is not the causal link but the perfect correlation of between signal and source, which underpins the indication relationship. This point about actual and possible causal means of transmitting the very same information is illustrated in figure 10, which is adapted from Dretske (cf. Dretske: 1981: 157, Jacob 1997: 69).

It is precisely because the interesting connection of perfectly reliable co-variance between informational content and its source is not causal *per se* that, "...the informational approach has an explanation of why the object of perception is a distal stimulus, not a proximal event" (Jacob 1997: 67, 69). This matters because it provides a ready solution to the distality problem. We must seek the points in the horizontal causal chain that enables us to establish that there is a relation of indication.

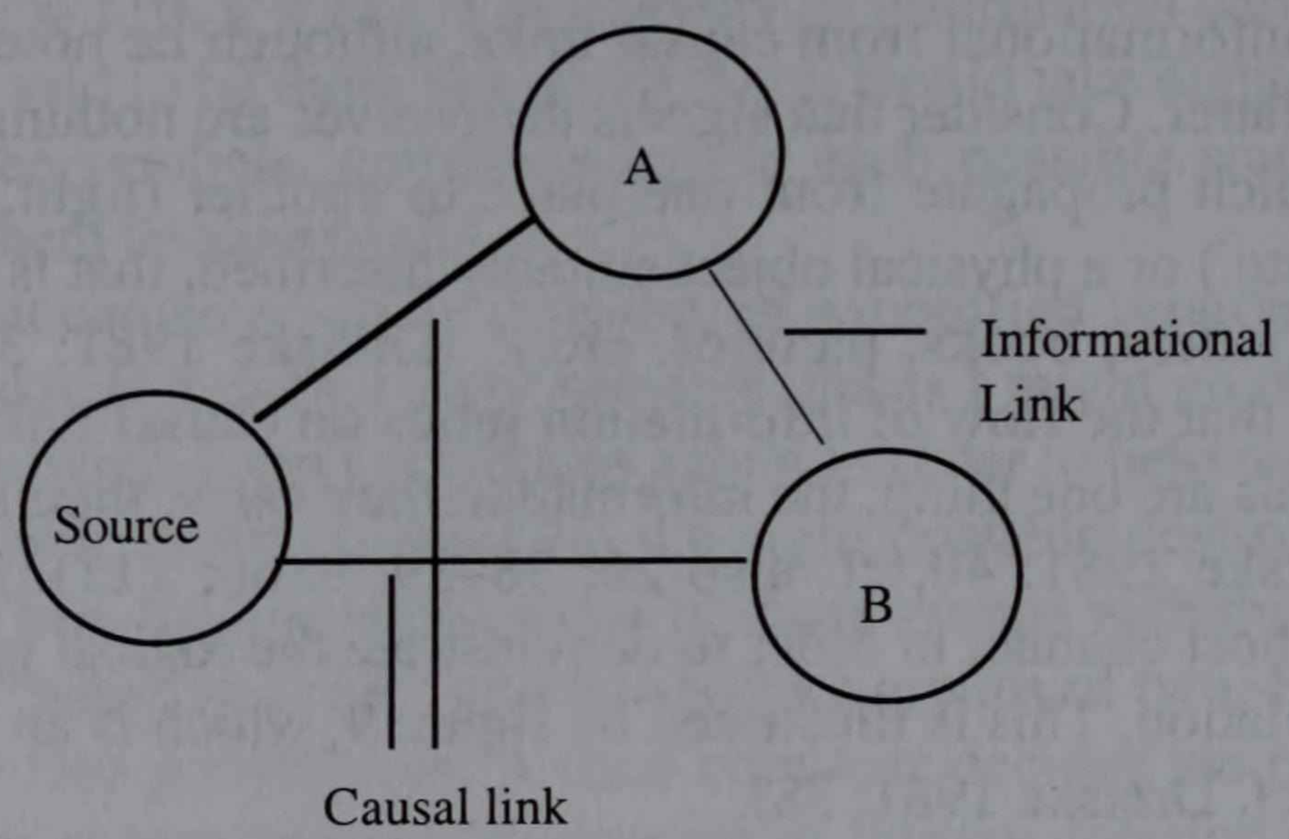

Figure 9. *Dretske's distinction between causal and informational links.*

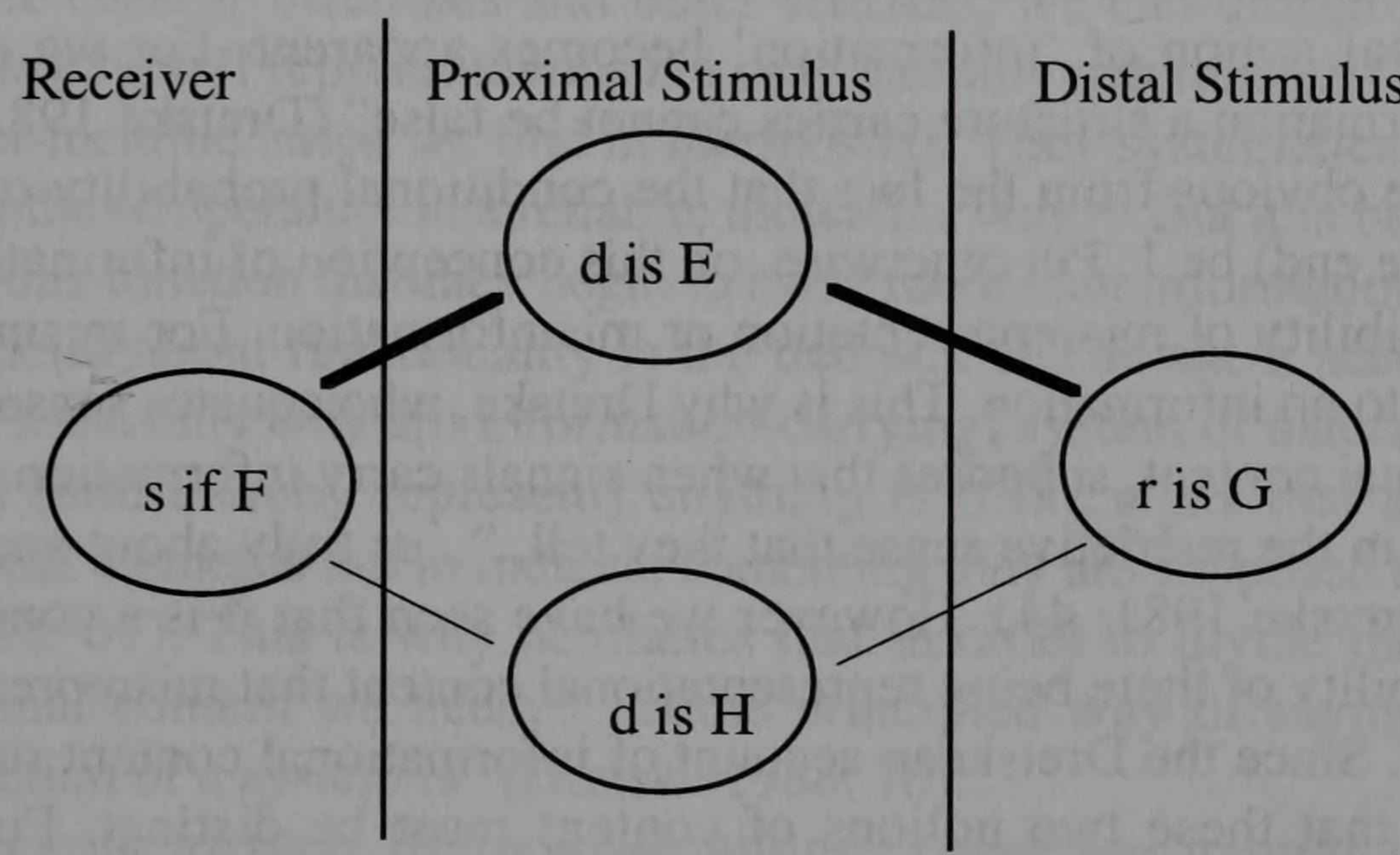

Figure 10. *Actual and possible causal links*

A third thing to note concerning Dretske's definition is that the possibilities need not be narrowed by the character of the signal itself. Hence he invokes the notion of 'background knowledge' of the information receiver by introducing k into his equation. If the information-receiver is suitably primed then a signal, which would otherwise contain no information, can unequivocally specify one state of affairs as opposed to another. For example, imagine we set up a code so that I know that the wave of your hand indicates that one of your two flatmates, Brendan or John, are at home. But, suppose I know independently that one of them, Brendan, is currently with me. In this case, barring deception and error, your wave conveys the information that John is at home. Dretske illustrates this same point by appeal to a game of duplicate bridge. He writes:

> Recently....I was asked by an opponent at a duplicate bridge tournament what my partner's bid meant. My partner had bid '5 clubs' in response to my '4 no trump' (Blackwood convention). I replied that it meant that he had either 0 or 4 aces. This is the conventional meaning of that bid, and the meaning of all bids must be revealed to one's opponent. Puzzled, however, by our somewhat erratic bidding (and doubtless, my incredulous stare at my partner), my opponent persisted: 'I know what it means. What I want to know is what his bid told you, what information you got from it?' Now this, obviously, is a completely different question, one that (given the rules of duplicate bridge) I am not obliged to answer. For what information was conveyed (to me) by my partner's bid was that he had no aces. This information was communicated to me because I had three aces in my hand; hence he could not have all four aces (Dretske 1981: 42–43).

It is here that the crucial distinction between the broad sense of 'meaning' and the technical notion of 'information' becomes apparent. For we can see that, "...the information a structure carries cannot be false" (Dretske 1981: 190). This ought to be obvious from the fact that the conditional probability of s's being F must (in the end) be 1. Put otherwise, on this conception of information there can be no possibility of misrepresentation or misinformation. For misinformation is equivalent to no information. This is why Dretske, who equates sensory content to informational content, supposes that when signals carry information to the senses they do so in the restrictive sense that they tell, "...us truly about another state of affairs" (Dretske 1981: 44). However we have seen that it is a condition on the very possibility of there being representational content that misrepresentation is a live option. Since the Dretskean account of informational content rules it out, he concludes that these two notions of content must be distinct. Put otherwise, information carrying systems can only represent external states of affairs if they are *supposed to* or are *meant to* represent them.

4.2. Teleofunctional Representation

It is with this in mind that Dretske developed a teleofunctionalist account representational content in his later book *Explaining Behaviour*. He draws a distinction between three main types of representational systems. First, there are type I representational systems. These are symbolic representations which have no intrinsic powers of representation. They gain their representational capacity solely by conventional agreement or fiat. Type I representations putatively include such things as; maps, diagrams, certain road signs, musical notations, gestures, codes, and the words of natural language (Dretske 1988: 53). Consider the use of the sign ∀ to stand for the logical quantifier 'For all x', as opposed to ∃ which represents 'There exists some x'. Treated simply as marks there is nothing about these two signs which makes them intrinsically representational. Similarly, it is easy to see that the use of a two-lantern code to warn the American revolutionaries of the British invasion by signalling 'One by sea, two by land', gains all of its representational powers through an established convention. The content of signs of this kind depends wholly on the pre-existing intentionality of the code-makers and, hence, is of no use to a naturalist.

Dretske also introduces the notion of type II representational systems which he defines as natural sign indicators. If a natural sign, F, reliably co-varies with the presence of Gs, then it can be used to carry information about the presence of Gs. This is his original notion of indication described in the previous section. We can exploit such pre-existing relations for various purposes. Dretske's examples include; using tree rings to tell the age of a tree, using tracks in the snow to tell which kinds of creatures have passed by, and using fingerprints to identify

people. In the case of machines and other artifacts, we can design them to do likewise. Hence type II representations are only partially naturalistic. Consider the function of bi-metallic strips we find in thermostats. They systematically respond to changes in the temperature in a reliable, indicative way.[24] But it is not until they are assigned this function that they begin to carry more than informational content. Indeed this point about functionality is the bedrock of Dretske's account for he claims that, "...the only way a[n information-carrying] system of natural signs can misrepresent [and thereby represent] anything is if the signs that serve as its representational elements fail to indicate something they are supposed to indicate" (Dretske 1988: 67). This is why he claims that in order to divine the nature of representational content we need, "...some principled way of saying what the indicator function of a system is" (Dretske 1988: 70).

This brings us to type III representations. These are natural systems of representation which have their own intrinsic indicator functions. These functions derive from the way the indicators have been developed and used by the system in which they are a part: e.g. animal sensory systems. Only representations of type III are of immediate interest to the naturalist, for only they have intrinsic indicator functions which could underwrite all other forms of representation. However, although type III representations are naturally occurring, there is scope for disagreement about what fixes their indicator functions.

One way an indicator system might gain a natural function is by having a biological function. For example, in a limited range of cases Dretske endorses a natural selectionist account of teleofunctionalism. He writes, "If we suppose that, through selection, an internal indicator acquired (over many generations) a biological function, the function to indicate something about the animal's surroundings, then we can say that this internal structure represents (or misrepresents, as the case may be) external affairs" (Dretske 1988: 94). In the simplest cases, an indicator system is supposed to represent a particular state of affairs, and not another, because the representation of that specific state of affairs is biologically useful to the organism. In this way, the representational capacities of biological creatures are shaped, at least in part, by evolutionary pressures. Consider Dretske's example of the natural selection of a particular representational system in stickleback. He writes:

> The fish exploit rather crude indicators (a bright red underside, for instance) to recognise one another. Males use the bright red underside to identify male intruders, females use it to identify interested males. The fish react similarly to a variety of objects of similar coloration: painted bits of wood elicit aggressive behaviour in the males and sexual interest in the females. But in the fish's natural habitat the correlation is good enough. By and large, *only* stickleback have this coloration (Dretske 1988: 103).

Ultimately, however, Dretske favours the idea that only internal indicators which are recruited to a given task by a process of operant conditioning can explain how reasons play an explanatory role in behaviour. Accordingly the stimulus-driven responses of a creature are reinforced through reward and punishment so that certain responses are selected over others (Dretske 1988: 94–96, 1991: 200–209, Cummins 1991b: 104, Horgan 1991: 86–87). He maintains that this account is necessary in order to explain the connection between reason-explanations and voluntary behaviour because voluntary behaviour is essentially learned.

4.3. The Property -Specific Character of Digital Information

Before I critically assess Dretske's teleofunctional proposal, it is important to highlight another distinction he draws between analogue and digital forms of information. He writes:

> ...a signal (structure, event, state) carries the information that *s is F* in *digital* form if and only if the signal carries no additional information about *s*, no information that is not already nested in *s's* being *F*. If the signal *does* carry additional information about *s*, information that is *not* nested in s's being F, then I shall say that the signal carries this information in analog form (Dretske 1981: 137).

This distinction is necessary if we are to understand a signal's property-specific character which enables it to serve a unique purpose in a control mechanism. Information in analogue form is unsuitable to serve a control function because it is too diffuse. For example, it is because sensory information is received in analogue form that Dretske claims that, "Typically, the sensory systems overload the information-handling capacity of our cognitive mechanisms so that not all that is given to us in perception can be digested" (Dretske 1981: 148). Consequently, "Until information is lost....an information-processing system....has failed to classify or categorise, failed to generalise, failed to *recognise* the input as being an instance (token) of a more general type" (Dretske 1981: 141, cf. Jacob 1997: 71). The process by which information is lost and through which a signal's information is selected or specified, is known as digitalisation. Akins' neurophysiological description of the way in which information is processed in hunting bats provides a vivid illustration of this kind of information loss. She writes:

> To say that a (sensory) neuron is informationally 'specific' is to say that when it processes an incoming signal there is information loss; because the neuron responds to only particular aspects of a complex stimulus, its signal

contains information about only those specific properties. For example, think of the response properties of neurons in the FM/FM areas of bat auditory cortex. These neurons will fire whenever there is a pair of stimuli with a distinct profile – two stimuli separated by a certain time delay with a particular frequency 'mismatch' between a fundamental tone and its harmonic, each stimulus failing within a certain amplitude range....the response of an FM neuron does not tell us, say, how many harmonics each signal contained, or their pattern of frequency modulation, or the amplitude of the signals except a certain point in the signal sweep and so on. Much information about the acoustic properties of the original stimuli....has simply been lost (Akins 1993: 149).

Dretske is attracted to the idea that some information is digital for several reasons. Firstly, he thinks it is digitalisation that enables a system to represent a specific state of affairs as being of *this* kind rather than of *that* kind. This is important because it explains why he thinks that all representations must indicate just one thing in order to perform a specific control function (see McLaughlin 1991: 161).

Furthermore, the property-specific character of digital information promises to provide answers to some difficult questions concerning our ability to learn. For example, if information is property-specific it can be nested in a signal. Signals can nest information either analytically or nomically. Thus Dretske writes:

> So, for example, I will say that the information s is a rectangle (or not a circle, or either a square or circle) is analytically nested in s's being a square. On the other hand the fact that I weigh over 160 pounds is nomically nested (if it is nested at all) in the reading of my bathroom scale (Dretske 1981: 71).

But such nesting is not necessarily apparent. Therefore, "...we are sometimes in a position to see that (hence, know that) s if F without being able to tell whether s is G despite the fact that every F is G" (Dretske 1981: 74). In this light, when a system recognises such connections we can say it has learned something (perhaps something important) about its environment. Finally, property-specificity might also account for a rudimentary form of opacity which is a hallmark of semantic content. It may explain the, "...fine–grainedness that is characteristic of intentional systems" (Dretske 1988: 75, Jacob 1997: 75).

4.4. The Indeterminacy Problem

For all its virtues there is a lethal chink in the armour of Dretske's account. It is this: there is no principled way to uniquely specify an indicator's control function which is compatible with the idea that the property-specific character of content rests on a digital informational base. It is best to consider this problem in stages. First of all, recall that the digital information carried by a particular signal can be nested such that a signal S which carries information about Fs will also carries information about Gs. Given this, it might be asked: How can it be claimed that some organism, O, has exploited a pre-existing relation of indication between S and F, such that its indicators serve to detect Fs, rather than saying, on similar grounds, that they function to detect Gs, or Fs *and* Gs? (cf. McLaughlin 1991: 158, 164–165). To the extent that Dretske rests his teleofunctional account on an information-theoretic account of indication his natural indicators will be infected with at least this level of functional indeterminacy. To prevent confusion with the problem of misrepresentation, as characterised by disjunction, Jacob helpfully labels this nesting problem: the transitivity problem. It is illustrated by figure 11.

Even though the problem of transitivity only applies to cases of analytically or nomically nested information it reveals the structural instability that lies at the base of Dretske's account. A tension emerges between his information-theoretic notion of indication and his teleofunctional notion of representation. This is commonly illustrated by appeal to his example of magnetic sensitivities of marine bacteria. He writes:

> Some marine bacteria have internal magnets (called magnetosomes) that function like compass needles, aligning themselves (and, as a result, the bacteria) parallel to the earth's magnetic field. Since these magnetic lines incline downwards (towards geomagnetic north) in the northern hemisphere (upwards in the southern hemisphere), bacteria in the northern hemisphere, oriented by their internal magnetosomes, propel themselves towards geomagnetic north. The survival value of magnetotaxis (as this sensory mechanism is called) is not obvious, but it is reasonable to suppose that it functions so as to enable the bacteria to avoid surface water. Since these creatures are capable of living only in the absence of oxygen, movement towards geomagnetic north will take the bacteria away from oxygen-rich surface water and towards the comparatively oxygen-free sediment at the bottom (Dretske 1986a: 26).

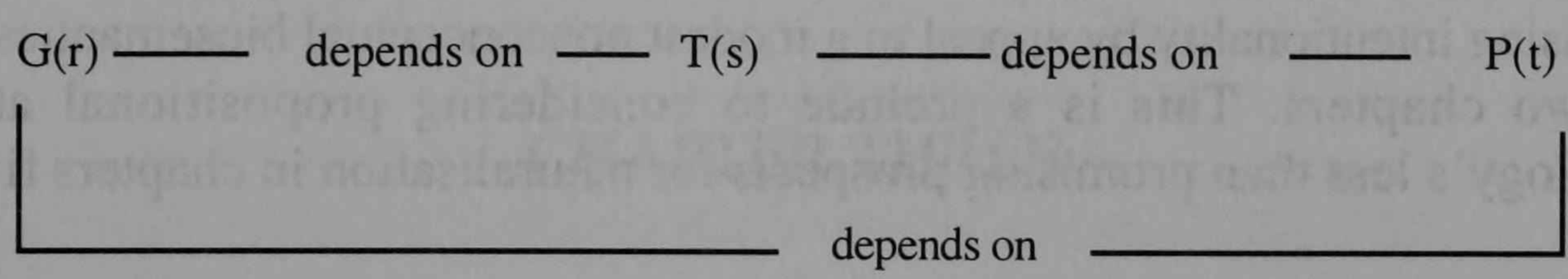

Figure 11. *The transitivity problem (adapted from Jacob 1997: 101).*

The magnetosomes of northern hemisphere bacteria carry information about geomagnetic north as well as, given transitivity, the presence of oxygen-free water. But which is the correct way to describe their representational indicator function? If we rely on an information-theoretic notion of indication we have no principle means to decide this issue. It appears to carry information about both. But this is not the verdict returned if we consider the ends served by the magnetosomes as they relate to the needs of the organism in question. The problem is that what a system needs is not necessarily what its indicators reliably detect in the original Dretskean sense of indication. Dretske must choose what fixes the content of the indicator: Is it what is reliably indicated or what the organism needs? Like the magnetosome, Dretske is pulled towards two targets at once (although both pull in the same direction on a good day). On the one hand, using Ockham's razor, if we can get by with an indeterminate informationally-based description of the indicator function why would we need the higher-level description at all? Contrawise, why should we bother with informational notion of indication if a device's teleofunction determines its representational content?

We have already seen good reasons to prefer the teleofunctional notion of an indicator over that of the information-theoretic version, since only teleofunctionalism allows for misrepresentation. Nevertheless, it must be realised that in taking this road the informational theory of content must be relinquished as the means of fixing content. Furthermore, another version of the transitivity problem arises at the teleofunctional level. This will be discussed in the next chapter when a need-based teleofunctionalism is more fully explored in the guise of biosemantics. Only after that will the role that information sensitivity plays in understanding basic intentionality be considered.

5. Conclusion

We have now reviewed the reasons for scepticism about a straightforward naturalistic vindication of propositional attitude psychology by appeal to an expansive classical cognitivism. This has usefully revealed the problems which must be overcome if commonsense psychology is to be naturalised. Armed with this analysis, I begin a more positive two stage exploration of its naturalistic

prospects in the next four chapters. I outline reasons for being optimistic about naturalising intentionality by appeal to a modest nonconceptual biosemantics in the next two chapters. This is a prelude to considering propositional attitude psychology's less than promising prospects for naturalisation in chapters five and six.

CHAPTER THREE

Black Dots and Red Herrings

"...the only hope for intentional psychology
is to embrace its biological roots."

– Millikan, *White Queen Psychology and Other Essays for Alice,* 1993

1. Introduction

In the wake of the failure of causal and information-theoretic accounts, this chapter investigates the prospects of the increasingly popular biological strategy of naturalising content. My aim is to introduce and defend a modest, biosemantic theory of content. According to Rowlands:

> A modest version of the teleological account might see itself as restricted in scope to fairly basic representational mechanisms, probably innate, probably mechanisms of perceptual representation, and probably possessed by creatures whose representational capacities are fairly fixed and limited (Rowlands 1997: 279, cf. Jacob 1997: 106).

While the account I advocate is modest in all these ways, it is also modest in the further important respect that it limits the ambitions of the biosemanticist to providing an understanding of the most rudimentary forms of intentionality (with a 't'). In particular, I argue that a non-truth conditional, non-referential theory of content is the best we can get from biosemantics. This has implications both for the possibility of naturalising commonsense psychology and, ultimately, the coherence of the eliminativist position. These issues will be discussed in chapters four and seven, respectively. Adopting this account also invites questions about how intentional content can be best understood in relation to full blown propositional content. This issue will be addressed in chapters four and five. However the first step is to introduce and provide a general defence of the biosemantic programme before offering reasons for thinking that modesty really is a virtue in this domain.

2. Biosemantics: The Natural Choice

Biosemanticists claim that if we appeal to a principled notion of biological function it is possible to give a genuine explanatory account of the nature of intentional phenomena. Millikan allows herself the following flourish when giving a crude rendition of her theory: she writes, "...what makes a thing into an inner representation is, near enough, that its function is to represent." (Millikan 1989a: 283). Yet, as she notes, talk of representational function in this loose way, as is the wont of many biosemanticists, does not sufficiently capture the unique features of her approach to biosemantics. However, before outlining these, it is important to be clear about her understanding of biological, or as she calls them proper, functions.

Firstly, Millikan's theory must be distinguished from the kind of account of teleological functions advanced by Wright. According to Wright a thing's function must be understood in terms of some benefit it confers and its existence is explained by citing this benefit. The primary appeal of this definition of functions is that it distinguishes between, "...a function and a fortuitous benefit" (Godfrey-Smith 1993: 198). Making this distinction is also a central feature of Millikan's approach but unlike Wright her theory is not a conceptual analysis of purposive behaviour. Consequently it does not mistakenly regard successful occasions of performance as paradigm cases of such behaviour. This is important because if one regards such cases as paradigmatic it invites a vision of purposive behaviour that can be measured along a spectrum from success to complete failure. In turn this encourages one to think of purposive behaviour in dispositional terms; as behaviour that 'tends to (successfully) bring about a certain effect' (cf. Millikan 1993: 24–25). But, as Millikan notes, under this schema ascriptions of purposiveness are arbitrary unless some formula is provided for saying how often a behaviour must succeed in order to earn the accolade 'successful' (cf. chapter two, section 3.1). Furthermore Wright is silent about the origins of his teleological functions. For this reason the scope of his definition is implausibly wide. Godfrey-Smith illustrates this with the following example:

> One might see a small, smooth rock supporting a larger rock in a fast-flowing creek, and note that if it did not hold up that larger rock, it would be washed away, and no longer 'be there'. But it is not the function of the small rock to support the larger one. (Godfrey-Smith 1994a: 345)

The point is that to truly make the distinction between a function and a merely fortuitous benefit we need more qualifications on the notion of a function than Wright's account provides. Not only do we need an explanation of X's 'being there' because of what it does, we need an account of how this effect was itself non-accidentally put into place. Millikan's theoretically grounded definition of

proper functions is superior to Wright's approach in precisely this respect. Her original definition of a proper function is:

> A function F is a direct proper function of x if x exists having a character C because by having C it can perform F. (Millikan 1984: 26).

This definition introduces an historical element since she bids us to, "...interpret 'because by having C it can perform F' to mean 'because there were things that performed F in the past due to having C." (Millikan 1984: 26, cf. also Godfrey-Smith 1993: 198-199). In order to understand this proposal fully we must distinguish (i) from (ii):

(i) The nature of a device's proper function

(ii) The character of a device which underwrites that proper function.

In the literature there are misunderstandings and objections to biosemantic theory which concern both these aspects. I discuss these in turn before returning to the issue of how biosemantics can provide a naturalistic account of intentional content. This is a crucial prelude to explaining the basis and character of such content.

2.1. Etiological versus Propensity Accounts of Selection

Proper functions are essentially normative. Importantly, they are normative in the strong sense that a device can have a proper function even if it fails to perform it. In this regard, Millikan never tires of observing that abnormality and dysfunction only make sense against the background of a non-arbitrary understanding of proper functioning. One of her favourite ways to illustrate this point is by appeal to the case of sperm and the ovum. She writes, "It is the biological purpose of the sperm to swim until it reaches an ovum. That is what its tail is for. But very few sperm actually achieve this biological end because ova are in such short supply." (Millikan 1993: 223, cf. also Millikan 1984: 4, 29, 34). If we accept that sperm have this biological function then we must also accept that it cannot be understood by sole reference to their mechanical dispositions. For one thing there are too many things, neutrally described, which devices are disposed to do. Sperm may be disposed to wriggle their tails in various mathematically describable angles that stand in relation to one another or in relation to the location of New York City. Do these dispositions constitute a biological functions of sperm? Of course not. What we need is a principled means of determining what the device is 'supposed to' do. It is, "...the task is to give a naturalist account of 'supposed to' in this context"

(Millikan 1984: 88, cf. Millikan 1993: 142). Millikan's suggestion is that this understanding can be provided if we rely on natural selection as our guide.

Her theory is fleshed out by appeal to the interlinked notions of 'normal explanations' and 'normal conditions'. The former, "...explains the performance of a particular function, telling how it was (typically) historically performed on those (perhaps rare) occasions when it was properly performed" (Millikan 1993: 86). With respect to normal conditions her position is that they must, "...be mentioned when giving a full normal explanation for the performance of that function" (Millikan 1993: 86). To understand proper functions therefore requires reference to a device's normal conditions of operation. The understanding of these norms requires appeal to natural selection and there are good reasons to understand this in a decidedly etiological, or historically-based fashion. Thus Millikan says that normal conditions could also be described as "...historically optimal conditions" (Millikan 1993: 87, cf. also 28).

Yet not all those who appeal to natural selection in order to understand proper functions agree that we ought to think of such selection in an historical fashion. For example, proponents of the propensity theory seek to explain the norms of proper functionality in the forward looking terms of *aptness* for future selection (cf. Bigelow and Pargetter 1987, cf. Coates 1997: 176-177). The thought is that we can say whether or not a device is doing what it ought by appealing to the good effects this has on its current and future chances for reproductive success. This approach is tailored to fit with the fact that, "Natural selection does not slack after the emergence of a structure but actually preserves it by acting against the emergence of less fit structures" (Millikan 1993: 86). While this proposal is well intentioned, the propensity theory appears inadequate because in order to make sense of the idea of 'aptness for future selection' it is necessary to appeal to an organism's past history and to mention real competitors and environmental features (Neander 1991: 464, Millikan 1993: 39-40, Godfrey-Smith 1994a: 352-353). Without reference to these factors, the propensity theorist's counterfactuals about a trait's (or device's) tendencies or dispositions that make it apt for future selection are under-determined, if not wholly empty.[25] How can we say what function a device is meant to serve if we don't already know whether or not, or in what respect it will, in fact, help the organism? Although it is not vital to the biosemantic project to decide the future prospects of propensity theory here, in what follows I will assume, along with Millikan and others, that an historically-based, etiological account is the right way to explicate the kind of functionality that underwrites proper functions.

Even so, the etiological approach can be qualified by taking into consideration recent episodes of selection as opposed to ancient ones. That is to say, it can tolerate the idea that while certain devices have been initially selected for a given purpose they can be re-selected or co-opted for new ones. Thus it is plausible that

natural selection can find different, perhaps more urgent, jobs for pre-selected biological artefacts to perform. For example, Dennett chronicles the fact that, "...what started out as a wristbone can become a panda's thumb" (Dennett 1987: 307). For another, slightly more controversial, example, it is at least hypothetically possible that the wing feathers of certain species of birds may have initially been selected for warmth, re-selected for hunting insects as they lengthened, and eventually re-selected for flight (cf. Millikan 1993: 44, Godfrey-Smith 1994a: 357-358). If we imagine these progressive deployments took place then, in line with the recent history thesis, the most recent instance of selection would determine the current proper function of wing feathers. It is worth emphasising that, given these functions are products of natural selection and not merely spandrels, the recent history account still preserves a principled means of distinguishing between a mere beneficial effect and proper function (cf. Dennett 1995. ch. 10). In contrast, it is precisely the distinction between merely, "...serving a function and having a function" which the propensity theorists seem unable to draw (Neander 1991: 465, cf. also Millikan 1993: 38). Of course, even on the recent history account we must be aware that it will take longer than a single generation in order for trait to be naturally selected. We must not misunderstand what is meant by recent history in evolutionary terms. This brings us back to the issue of the underlying reproducible character of devices which is another crucial component in the definition of proper functions.

2.2. The Reproducible Character of Biological Devices

Millikan claims that items with proper functions must share a set of common characteristics with earlier items. Furthermore, she holds that these characteristics are gained through systematic reproduction or copying. The idea is that such items form a family, which she calls a reproductively established family. The first item in this chain of reproduction is 'the model'. This way of understanding reproduction is often illustrated by analogy with the products of a Xerox machine (cf. Millikan 1984: 20, Davies 1994: 369-370). For example, if I make a dozen photocopies of page twenty of *Language, Thought and Other Biological Categories* each copy will all share a certain range of features (e.g. letter shapes and sizes, line spacing, and so on). Moreover they will share these common features because they all derive from the same original (the model) and because they have been generated by a reliable copying process.[26] This notion of reproduction is vital to Millikan's account. It is the means by which we can make sense of the idea that current devices have proper functions *because* they belong to a certain family with a common historical origin.

The category of things belonging to reproductively established families is quite large. For example, it includes items such as tokens of English words and

gestures all of which have systematically reproduced features (Millikan 1989b: 293). Any of these things may turn out to have proper functions. However, given my agenda to assess the prospects of naturalism, I will restrict attention to those things which are members of biologically-based reproductively established families – specifically those which have been naturally selected and reproduced by genetic mechanisms. And, as is well known, the general structure of Millikan's account is designed to be compatible with an account of trait selection and maintenance according to which reproduction operates at the genetic level.

But there are complications. In order to apply to phenotypes Millikan's account must encompass devices which are members of *higher-order* reproductively established families, as well as members of reproductively established families *simpliciter*. This distinction is best illustrated with the help of an example. Consider hearts. Hearts can only have the direct proper function of circulating blood, if and only if they are members of some kind of reproductively established family (cf. Elder 1994: 167). But given that no heart is directly copied from any other, the reproduction process must rely on intermediaries to secure their membership in such a family. Hearts, or their essential characteristics, must be systematically produced by some lower-order items which are themselves directly copied from earlier lower order items (in this case certain genes). Furthermore these lower level items must have the 'production of hearts' as their direct proper function (cf. Rosenberg 1985a: 51). That is to say, such items will have been copied from earlier items because they had the very effect of 'producing hearts'. To sum up, in order to understand how high-order devices come to have proper functions we must look to the benefits provided by historically earlier items from which they are indirectly copied (cf. Millikan 1993: 41). Genes are allegedly paradigm examples of a lower-order reproductively established family of this kind. Neander emphasises this in her most basic formulation of the definition of proper functions. She writes:

> Def: Some effect (Z) is the proper function of some trait (X) in organism (O) iff the genotype responsible for X was selected for doing Z because doing Z was adaptive for O's ancestors (Neander 1995: 111).

Davies has objected that genetic transmission in sexually reproducing organisms cannot be useful compared to Xerox copying and, for this reason, natural selectionist accounts of proper functions are doomed. With reference to the case of hearts, he argues that there is no identifiable set of genes which are both part of a reproductively established family, and yet have the characteristic function of 'producing hearts'. The objection rests on the fact that gamete cells are produced in sexually reproducing creatures *via* the process of meiosis, which as Davies notes, "...reduces by half the number of chromosomes and hence the number of genes" (Davies 1994: 370). The point is that two parents contributed to

the set of genes which are responsible for the development of my heart – just as four parents contributed to the development of the set of genes responsible for my parents' hearts, and so on. Since sexual contributions have an inherent variability we cannot identify any precise set of physio-chemical features which are copied from one generation to another. Consequently Davies is right to pose the question, "How are we to type-identify heart-producing (liver-producing, kidney-producing, etc.) genes?" (Davies 1994: 372). Certainly we should not look for identities at the physio-chemical level. But, despite conceding that there is a necessary level of variation in the devices and traits of sexually reproducing organisms (e.g. different heart sizes, shapes, etc.), it is clear that what remains constant from generation to generation is the overall characteristic effect that hearts produce. Hearts were selected because of the role they play. If we type-identify 'what is reproduced' at this level of functionality, as opposed to some lower level, then Davies' objection loses its teeth (cf. Elder 1994: 168–169). There is nothing circular about this move as long as we understand this kind of functional type-identification in Cumminsesque terms. According to Cummins:

> Functional analysis consists in analyzing a disposition into a number of less problematic dispositions such that programmed manifestation of these analyzing dispositions amounts to a manifestation of the analyzed disposition (Cummins 1983: 28).

Essentially, Cummins-functions should be understood simply as effect-producers which contribute to the overall response of a larger system. In this respect, they should not to be confused with Millikan's proper functions, although they are compatible with them (Godfrey-Smith 1993: 200, cf. also Millikan 1993: 20).

2.3. Biosemantics and Intentionality

Having outlined and defended the central features of Millikan's account of proper functions we are now in a position to consider the claim that the distinction, "...between proper functioning and malfunctioning might somehow underwrite the normative notion of content – and its distinction between representation and misrepresentation" (Neander 1995: 112). In this regard, Millikan rightly insists that the first place that it is appropriate to speak of content, with respect to biological devices, is in relation to what she calls 'intentional icons'.[27] As the name implies such icons are intentional (not intensional) in Brentano's sense that they are directed at features, objects or states of affairs which may or may not exist (cf. Millikan 1993: 109, 1989: 96). Intentional icons have three essential features. These are:

(a) They are relationally adapted to some feature of the world.

(b) The relation described in (a) can be characterised by means of a mathematical mapping rule.

(c) They have the direct proper function to guide a co-operating (consumer) device in the performance of its direct proper function(s).

Millikan's paradigm example of a simple intentional icon is the bee dance. Such dances are meant to generate an appropriate response in a co-operating consumer mechanism(s); the watching bee or bees. In this case, a patterned flight response which takes the bee(s) to the location of nectar. A bee dance, like other intentional icons, has both indicative and imperative aspects. If the relational conditions for this characteristic type of dance are normal then it will successfully map, *via* some mathematically describable projection rule. This is the icon's indicative aspect. If this is satisfied then it can direct the icon consuming bee(s) to the nectar's location.[28] This is the icon's imperative aspect (cf. Millikan 1993: 107).

Although the example of co-operating mechanisms in the bee dance is one of two separate devices, in this case two or more different organisms, the account works just as easily by appeal to distinct mechanisms within a single organism. Millikan is explicit about this intended use. She writes:

> Put (an analogue of) the bee dance inside the body so that it mediates between two parts of the same organism and you have...an inner representation (Millikan 1993: 164).[29]

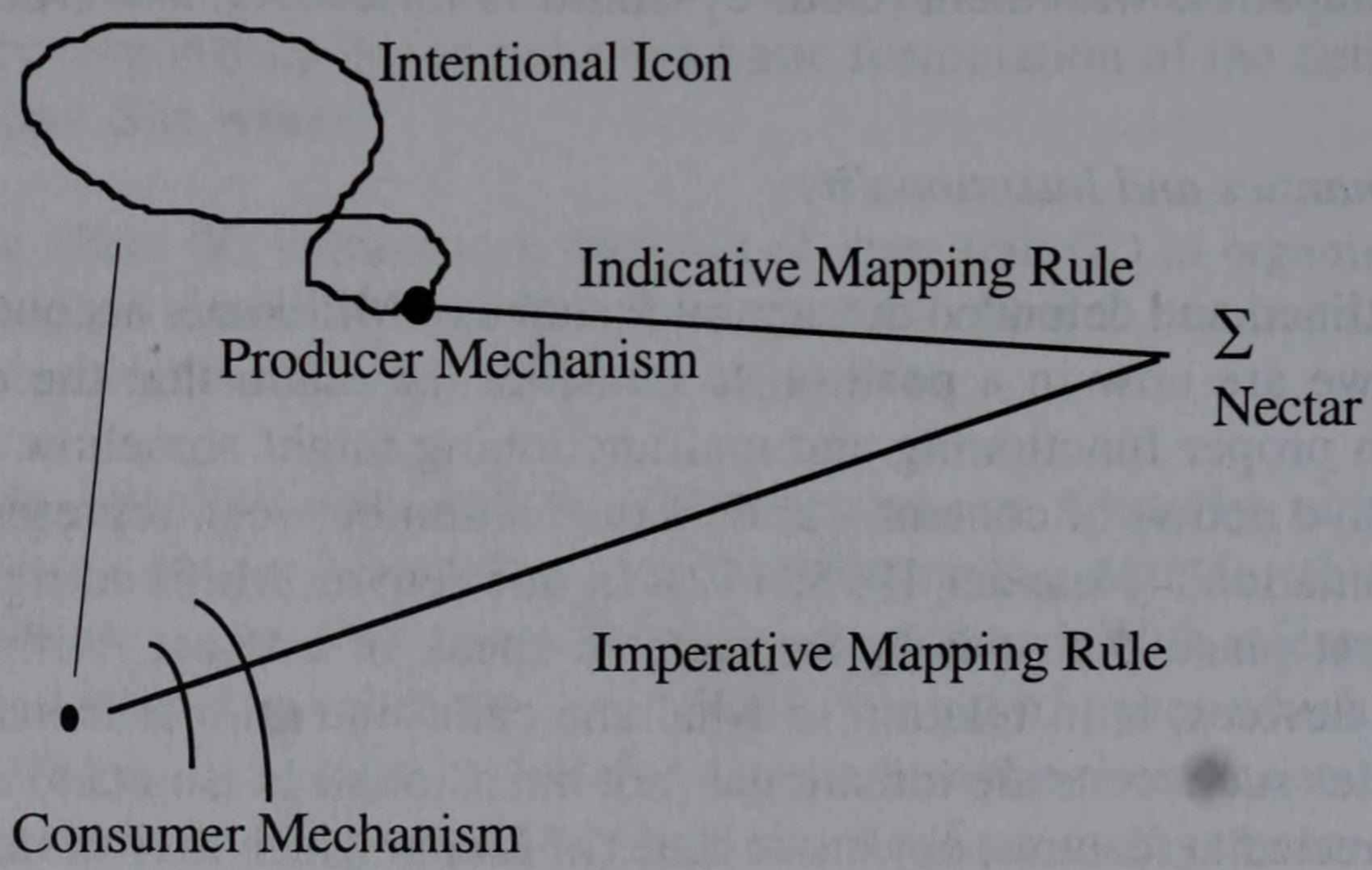

Figure 12. *The bee dance: triangulation at the intentional level.*

The important thing to note is the emphasis that Millikan places on icon consumption in determining what is being represented (cf. Elder 1998: 351-352). This is why she insists there is a strong contrast between her biosemantic approach and causal-informational theories of content. This contrast exists even with teleofunctional theories of indicator functions, such as Dretske's, to the extent that they seek to define indication in information-theoretic terms. In this regard, Millikan stresses that:

> ...intentional icons do not, as such or in general, carry 'natural information'. Nor do they 'covary' with or 'track' what they icon. Their definition makes no reference to how likely or unlikely they are actually to correspond to their designated environmental features, nor how likely these features are to get mapped by them (Millikan 1993: 107).

We cannot understand the direct proper function of intentional icons solely by reference to what they regularly map unto, statistically understood. Millikan's approach sharply distinguishes proper, or teleofunctions, from mechano-functions. As we have observed, a device can have a proper function due to its history even if it rarely, or never actually succeeds, in performing it. Similarly the direct proper functions of icons need to be explicated with reference to what effect they ought to produce. This is determined by their historical pedigree – not their current dispositions. To understand these functions we must look to the normal conditions for the operation of their producer-consumer devices. For this reason, the norms introduced by such an account must not be confused with statistical norms. They appeal to a device's selectionist history in which success is measured, not by achieving a certain result x number of times above some arbitrary statistical threshold, but by a principled appeal to what was originally 'good enough' for proliferation of the trait or response.

In this light, it is easy to see how biosemantic theories avoid the problems of distality (along the horizontal axis), misrepresentation (along the vertical axis), and transitivity which, taken together, are the bane of purely causal and informational theories of content. This is because the biosemantic theory provides a principled, yet naturalistic, means of assessing error. In Millikan's own words, "Some of us have argued [we should] ground the needed norms in evolutionary biology – to let Darwinian natural processes set the standards against which failures, untruths, incorrectness, etc. are measured" (Millikan 1991: 151). Appeals to proper functions, which describe the effects a device is *supposed to* or *meant to* produce, provide the guide by which we are to assess whether or not any given 'intentional state' is mapped appropriately or inappropriately. Nevertheless unlike other naturalistic theories of content the biosemantic theory does not depend on locating a special kind of 'relation' that exists between representation and that which is represented (cf. Akins 1996: 339-340). It attempts to understand the

dynamics and norms involved in icon consumption. We get the 'representational' relation *gratis* once this is done. It is, as it were, fallout. However adoption of this view has consequences for the causal efficacy of content, which I shall discuss in the next chapter.

3. Defending High Church Biosemantics

Biosemantic theory has encountered the strong criticism that, appearances to the contrary, it fails to provide a stable means of determining, and hence naturalising, content. This is despite some of its obvious virtues in solving the triad of problems which plague other theories of content as discussed in chapter two, sections 3 and 4.

3.1. The Intensional Indeterminacy Argument

Fodor has been the most vigilant opponent of this class of views and his remarks are the *locus classicus* of the critique. Referring to the time-worn example of the frog that can't distinguish between flies, black dots, and bee-bees, he writes:

> You can say why snapping is a good thing for frogs to do given their situation, whichever way you describe what they snap *at*. All that's required for frog snaps to be functional is that they normally succeed in getting flies into the frogs; and so long as the little black dots in the frog's Normal environment *are* flies, the snaps do this equally well on either account of their intentional objects (Fodor 1990b: 106).

It is important to be clear about this objection. It is well known that frogs tend to react to a whole range of small, dark, moving objects in the same manner (i.e. shooting out their tongues). We have already been told that appeal to device's proper function provides the logical space for a normative assessment of misrepresentation. But such assessment requires that we look to, "...the *most proximate* Normal explanation for full proper performance" (Millikan, 1984: 100, emphasis original). Millikan's stipulation leads us to favour the view that the function of the tongue-snapping behaviour is not disjunctive at all; rather it is directed at flies and flies alone, since ingesting these served the ancestors of this type of frog in the evolutionary history of these creatures. It alone enters into a full explanation of why the response mechanism proliferated (Millikan 1991: 156).[30] Her version of biosemantics focuses on the fact that direct proper function of intentional icons must be defined in relation to the benefits which accrue to the consumer device(s) in the organism. Thus the frog's tongue-snapping behaviour is designed to get its dinner. Given this, its tongue-action *should* be directed at the

range of edible objects which are compatible with the effective working of its stomach in normal conditions of the kind which benefited its ancestors.

Bearing these points in mind, Fodor's critique has the following form: He agrees that from the selectionist perspective, a frog's inner icon only misrepresents in those cases where its snapping behaviour is not in line with the normal conditions that are part of an explanation of why the device was originally selected. Nevertheless he suggests that if in the historical environment of the frogs' ancestors black dots of such and such a size, speed and shape were in fact coextensive with flies (or were often enough flies), then it doesn't matter which way we now choose to describe the frog's detection device. In giving our explanation we can call it a fly or black-dot detector without consequence. All that matters is that a device with the propensity to detect small, dark moving things, in fact, aided this type of frog (or at least did not competitively disadvantage them) in the normal habitat of its ancestors and hence it was reproduced. His claim is that historical explanations of this kind do not provide a principled basis for making distinctions between intensionally alternative, but explanatorily adequate, descriptions. He writes, "The context '– happened' is transparent for the '–' position, so it would be sort of surprising if contexts defined in terms of it weren't transparent too." (Fodor 1991c: 296, cf. also Rowlands 1997: 280). Consequently it doesn't matter how we choose to describe the function of the frog's detector mechanism because all that mattered to its selection was the neutrally described capacity to detect small, moving black dots. This is consistent with the fact that what were detected were often enough flies and that these alone provided the ancestor frogs with (enough) nutrition to get by.

There is something right and wrong in both these responses. On the one hand, Fodor is right in thinking that how we choose to describe what-it-was that the ancestor frogs were meant to detect when now offering selectionist explanations couldn't have mattered, metaphysically speaking, to what they actually detected. The only thing that could have made a difference to the selection of such detection devices was that they enabled frogs equipped with them to respond well enough to proliferate. On the other hand, Millikan is certainly right to think that it is the getting of nutrition, not merely coloured dots of a certain size and speed, which is responsible for proliferation of the frogs (and their detection devices). Without a top-down teleological account we would not rightly understand why the mechanisms were selected at all. In this light, Neander's suppositions about what Darwin's likely interests are apposite. She writes:

> [Darwin]...might wonder which feature of the A/B situation was beneficial to the organism, and why it was beneficial, and he might wonder how the organism detects that it is an A/B situation, and which feature of such situations the organism is responding to. And the answers to these questions are not description insensitive. Perhaps A was the benefit and B was the

stimulus, or perhaps B was the benefit and A was the stimulus, or perhaps A was both the benefit and the stimulus, or perhaps B was (Neander 1995: 122).

The point is that given the organism's needs describing the detection-response devices of frogs as fly-detectors has explanatory priority over describing them as black-dot detectors. For this reason, Fodor's outright rejection of teleology misunderstands the order of explanation.

3.2. The Capacity Argument

Having said this, looking from the bottom up, Millikan's account has appeared explanatorily insufficient to some. Her critics suggest that if we concentrate wholly on what has actually benefited organisms as the basis for determining what an icon ought to point to then we are in danger of counter-intuitively demanding too much of the representational capacities of these devices (Neander 1995: 126-129, Jacob 1997: 123-124). Such descriptions can be quite abstract and general. Neander illustrates the problem with the example of a male hoverfly. According to an exclusively benefit-based account, given his ultimate needs a male hoverfly, "...misrepresents if he chases an infertile female or one who is soon to be the dinner of some passing bat" (Neander 1997: 127). Likewise the frog misrepresents if the fly it detects happens to be carrying a frog-killing virus or if it isn't in fact nutritious. Accordingly the correct descriptions of the proper functions of such devices is to lead hoverflies to 'fertile mates' and to direct frogs to 'nutritious protein'. This in turn fixes the representational content of icons at a higher level of grain than described earlier.

On this construal, the main issue is whether or not we can seriously credit organisms with the just-so capacity to represent that, and only that, which is 'good for them'. Hence if we consider figure 13 we might wonder: Which level, and its associated proper function, matters to intentional content? In response, Neander suggests that when offering a biosemantic account we ought only to give attention to the, "...lowest level at which the trait in question is an unanalysed component of the [proper] functional analysis" (Neander 1995: 129). It is her contention that this reflects sound biological practice. She claims that, "...with respect to a given part (P) of an organism, biologists should (and do) give priority to that level of description which is most specific to P – the lowest level of description in the functional analysis before we go sub-P" (Neander 1995: 128).

And the fact is that we can deploy the terminology of proper functions to a device's lowest level of operation. Such devices can malfunction in a way that demands a normative understanding.[31] For example, if we think of the frog as directed, not at flies, but at 'small dark, moving things' it is still possible for its responses to go awry. Neander writes:

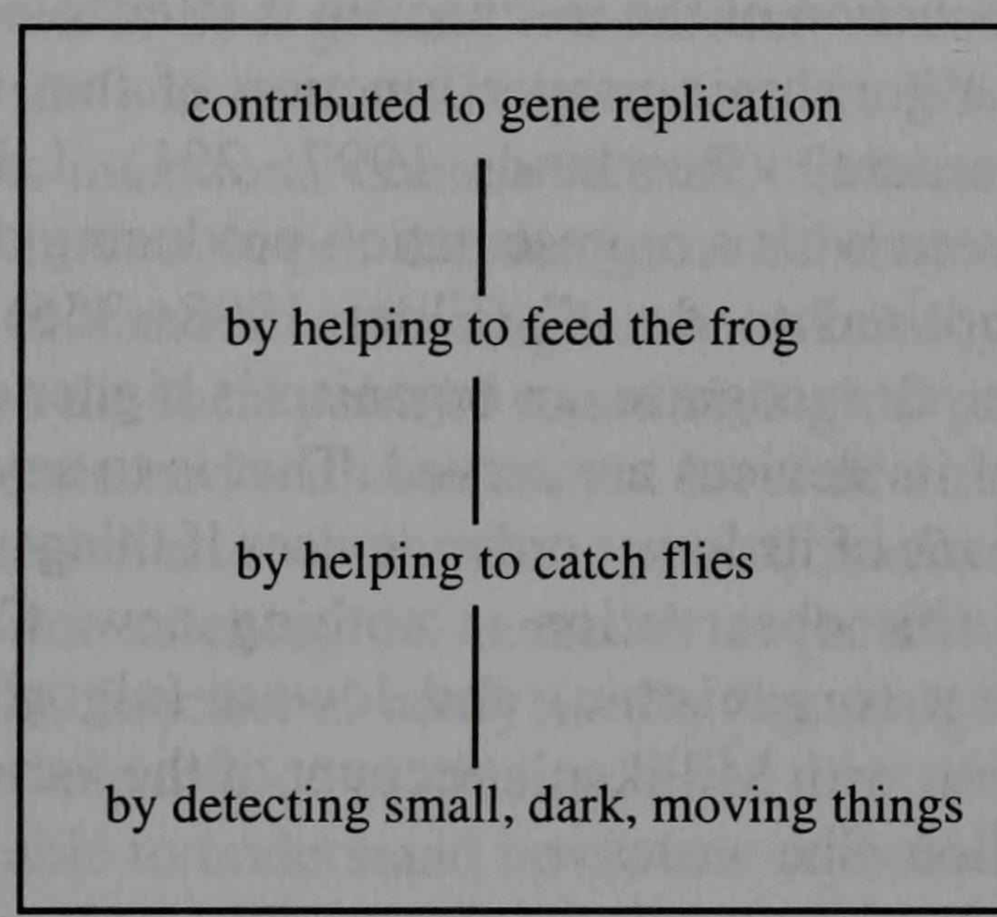

Figure 13. *Neander's low church biosemantics (from Neander 1995: 125)*

> A sick frog might R-token at a snail if it was dysfunctional in the right way. Damaging the frog's neurology, interfering in its embryological development, tinkering with its genes, giving it a virus, all of these could introduce malfunction and error. Therefore, the theory I am defending does not reduce content to the non-normative notion of indication or natural meaning (Neander 1995: 131, cf. Jacob 1997: 118, 134).

Such is true even of the brain's, normally reliable, opioid receptors which are meant to interact with endorphin molecules. As Dennett notes, "Both can be tricked – that is opened by an impostor. Morphine molecules are artifactual skeleton keys that have been fashioned to open the opioid-receptor doors too" (Dennett 1997: 47).

These facts inspire a view of content which locates it at the owest possible teleofunctional level. But this move is ill-motivated. It is wholly consistent with Millikan's benefit-based account of the direct proper function of intentional icons that there exist logically stacked proper functions of the kind Neander describes. What her diagram reveals is simply how various higher level ends are served by the successful performance of lower level devices or mechanisms. Organisms have devices (traits, responses, etc.) with different, multiple proper functions that are related to one another in a means-end fashion. The higher ends are served by the operation of mechanisms which can be functionally described in various ways depending on which end of the spectrum we wish to study. Rowlands recognises this kind of divide between higher and lower ends by insisting that we distinguish between, "...two importantly distinct levels of functional specification" which he respectively calls, "...the organismic and the algorithmic levels of description" (Rowlands 1997: 288). For example, in the case of the rattlesnake, "...the

organismic proper function of the mechanism is to enable the rattlesnake to detect eatability, but the algorithmic proper function of that mechanism is to detect warmth and movement" (Rowlands 1997: 291). Likewise Elder draws a distinction, "...between what a representation-producing device is supposed to do, and how it is supposed to do it" (Elder 1998: 356). Clearly, as Neander, Rowlands and Elder all recognise, an organism's higher ends will only be served if the lower ends of its devices are served. That is to say, an organism will only reap benefits by means of its lower order devices if things are functioning properly on all fronts. But this observation is nothing new. Compare the distinction between the highest (organismic) and lowest (algorithmic) level of proper functional description with Millikan's account of the various ways in which a bee dance can malfunction. She writes,

> It is a function of the bee dance to lead the watching bees to the indicated nectar, *even if it is poisoned.* Prior to that, it is a function to lead them to a certain location, *even if someone has taken the nectar away or replaced it with a trap.* Suppose that a bee makes a mistake and dances a dance that's just wrong, either because it is not normal or because environmental conditions are not as required for its accurate functioning. In either case, *a function of the dance is still to lead the watching bees to where the dance says the nectar is* (Millikan 1993: 167-168, emphases mine).

In this context, we can see the importance of Millikan's distinction between a device's direct proper function and its derived, adapted proper function. For example, a token bee dance will point the consumer bee(s) towards the current, possibly unique, location of nectar. In this sense, it has the derived function to point *thusly*. But it only has this function in virtue of the fact that bee dances, as a class, have the stable direct proper function to send watching bees toward nectar. This is comparable to the way in which photocopier has both the general proper function to copy 'that which is placed on the glass' and the supporting adapted, derived proper function to produce copies of the particular items placed on it. The point is that the function to point in particular direction is inherited from its direct proper function to get the consumer bee to nectar. This is why Millikan assigns predominance to a device's direct proper function when determining the content of an icon. It is also why she is able to ask, rhetorically, "Is it the heart's function to pump, or to pump blood? Both, I should think...And so to with the magnetosome and the frog's detecting mechanisms." (Millikan 1991: 160, cf. Jacob 1997: 120). Although these proper functions standardly complement one another nothing guarantees that they always do. Malfunctions can occur at various levels, and for a number of reasons. But our concern is with the intentionality of the complex responses, not the possibility of malfunction in the mechanisms that underlie them.

In light of what appears to be a general consensus, we can now return to the real question in this private war between biosemanticists: Which level, or which proper function, matters to intentional content? Recall that Neander's proposes that we should look to the lowest level, on the grounds that this reflects sound biological practice. But this should encourage us to ask: Why is attention to the lowest level of description the focal point of sound biological practice? The answer is that the biologist is concerned with the lowest level of a mechanism's proper functions because it is at such a level that the explanation of a device's capacities can be 'discharged' in non-teleological terms. It is the hand over point for a different, Cummins-style, functional analysis. By focusing on the effects that mechanisms are, in fact, capable of producing when contributing to an overall system response we are able to understand how such capacities can be ultimately broken down into purely physio-chemical (merely causal) processes. Rosenberg supplies some detailed examples of the way this kind of homuncular discharge takes place, "...at the border of molecular biology and organic chemistry" (Rosenberg 1985a: 59). However, despite bidding us to look low, Rosenberg rightly recognises that attention to the lowest level is wholly consistent with the idea that at the higher level these mechanisms should be accorded functions with respect to their, "...evolutionary or adaptive pedigree" (Rosenberg 1985a: 59). Thus he writes:

> The function of the liver is to maintain blood-glucose levels because (1) the liver is *capable of doing so*, and (2) the body that contains it *meets the needs* to supply energy to the muscles, in part *through* the capacity of the liver (Rosenberg 1985a: 58, first and second emphasis mine).

Nothing in Millikan's benefit-based account of the direct proper function of intentional icons breaks faith with this. All Rosenberg's example reveals is that various higher level ends are served by the successful performance of lower level devices. Looking at matters in this light demonstrates why there is no need to make a choice between high and low biosemantics. If we accept that the various teleofunctional levels are complementary then, to return to Neander's example, it is surely the case that the correct selectionist explanation of a hoverfly's target is simply 'female hoverfly', while the frog's target is 'fly'. This gives us the means of properly addressing the concerns raised about perceptual capacities (cf. Elder 1998: 359). Consider the frogs. In the normal environment of their ancestors, it was the (perhaps rare) encounters with flies that accounted for the proliferation of their less than perfect sensory systems. However limited, their responses to flies were good enough when all was well, given their needs and their competition. Hence the mechanisms of their descendants ought to point to flies. And ought implies *can*.

In my view, the Neander objection about capacities is confused because the notion of a capacity is equivocal. Sometimes it is important to talk of greater and lesser capacities. However, in this case, it is wrong to understand the notion comparatively. Of course, it is true that frogs respond more frequently to 'small moving black dots' than to 'flies'. But this is no surprise since such responses are the means by which they are able to get flies. The capacity argument would only be a genuine cause for worry if we defined proper functions in statistical terms – which we have not. In the right conditions, the frogs' icons do have the capacity to point to flies. What they lack is the capacity to discriminate flies from all non-flies.

Nonetheless it may be argued that this was not the point of the objection. What of the claim that taking the high road demands too much of the perceptual capacities of such creatures and assigns too much content to their icons? Why not say it is the more abstract end of spectrum that defines the intentional object since it defines the organism's ultimate target? Why say the frog is chasing 'flies' instead of chasing 'nutritious things', or simply 'nutrition'? Once again, the answer concerns explanation. Given the historical competition the ingesting of 'flies' was *good enough* to get nutrition. If it wasn't we presume that frogs of this type would have adapted more precise sensory mechanisms to detect flies (which are co-extensive with the nutritious things) or they would have failed to proliferate. In either case, neither their target was fixed nor were the conditions which contributed to its mechanistic arms race developed in a vacuum or in an abstract design space. In all, taking the 'too high' road results in loss of explanatory purchase. If we travel too far in the direction of an abstract description of the organism's needs then every creature will be described as targeting the same things. For all creatures, great and small, success in the wild depends on making appropriate responses to 'fertile mates', 'predators' and 'nutritious objects'. Despite this, not every creature in the same biological niche are genuine competitors for the same resources, even though their targets fall under these general categories. The fly's potential mate is the frog's potential dinner. Creature needs are particular and, thus, must be distinguished more finely than the top-level description allows.

3.3. The Reply from on High

Apart from being ill-motivated, there are other more serious problems in taking Neander's recommended low road when it comes to fixing intentional content which drive me to think that her proposed, "...philosophical marriage of Fodor and Millikan" must end in divorce (cf. Neander 1995: 137). First, her account re-introduces the distality problem which Millikan's version of biosemantics laid to rest. She notes this herself by telling us that low church biosemantics, "...seems

to drive us to proximal content...[for example,] it is, after all, by responding to a retinal pattern of a particular kind that the frog responds to small dark moving things" (Neander 1995: 136).[32] This is not trivial because if so low church biosemantics violates one of the minimal conditions for a device to count as an intentional icon. The mere fact that a biological device has a proper function and, hence, can malfunction is not sufficient to regard it as having a representational capacity. For if this was all that was required then all naturally selected biological devices would have content of some kind.

Millikan demonstrates that, generally, it is unsafe to infer intentionality from biological functionality. She points out that the devices in chameleons, which enable them to alter their skin coloration in relation to the background of their particular environments, have proper functions but do not thereby belong to the class of intentional icons. Such devices have this capacity as a relational, derived adapted proper function. As such they produce an appropriate mapping rule, but they lack a co-operating consumer device. For this reason, they are not intentional icons (Millikan 1993: 107). But the situation is not the same with respect to Neander's proposal. Unlike the case of pigment adjusters in chameleons, the problem with low-level 'black dot' detectors is not that they lack a consumer device for the retinal pattern they produce.

What exactly, then, is wrong with the idea the frog is directed at certain characteristic patterns on its retina as opposed to flies? The problem is that describing the function of the producer device only in terms of 'generating retinal patterns', as opposed to any of a number of other proximal causal descriptions, does not explain why the response, and its underlying mechanisms, proliferated. Only mention of the distal object of concern, its 'final' cause, – be it internal or external – can do that. We must tread carefully here in order to understand the logic of this point. The argument is not that it is an essential feature of biosemantic theory that the functions of organismic systems reach out to the external world. This may not be obvious given that Millikan has no compunction in supporting a very broad vision of what intentional icons can be directed at (cf. Millikan 1984: 100). Godfrey-Smith provides an illustrative example:

> Sand scorpions detect prey by picking up combinations of mechanical waves in the sand....When an intruding biologist disturbs the ground with a ruler, and elicits a strike, the scorpion could be functioning normally, but the environment is abnormal (Godfrey-Smith 1989: 546).

The point is that the scorpion's co-ordinated response is meant for a particular environment. This is in harmony with the idea that the behaviour of interest to the biopsychologist must be classified in, "...accordance with [proper] function" which is defined with reference to a loop into....an environment (Millikan 1993: 135, 136). Nonetheless, while talk of environments is important, the internal-

external dichotomy is largely artificial. The only interesting difference between an internal and external environment, in the biosemantic context, is that the former tend to be homestatically regulated and, hence, more stable (cf. Millikan 1993: 161, Dennett 1997: 98). What is important is that the organism ought to be directed at a distal object – the one that would satisfy its ends. Consider again Millikan's distinction between a device's direct proper function and its derived, adapted proper function. This distinction provides a means of clearly demarcating proximal and distal projection rules. She uses it in just this way when she describes hoverfly mating responses. Of the former, she writes:

> Rather than turning toward the target in order to track it, the hoverfly turns away from the target and accelerates in a straight line so as to intercept it. Given that (1) female hoverflies are of uniform size, hence are first detected at a roughly uniform distance (about .7 m), (2) females cruise at a standard velocity (about 8m/sec), and (3) males accelerate at a constant rate (about 30-35 m/sec^2), the geometry of motion dictates that to intercept the female, the male must make a turn that is 180 degrees away from the target minus 1/10 of the vector angular velocity (measured in degrees per second) of the target's image across his retina....Taking note that this rule is not about how the hoverfly should behave in relation to distal objects but rather how he should react to a proximal stimulus, to a moving spot on his retina, let us call this rule 'the proximal hoverfly rule' (Millikan 1993: 219).

The point is that the lower-level, algorithmic projection rule (which is a product of the icon's derived, adapted proper function) enables us to understand how the hoverfly should respond to proximal stimuli. This is contrasted with the 'distal hoverfly rule' (which is a product of the icon's direct proper function). Millikan describes that rule as 'if you see a female catch it' (Millikan 1993: 222). The point is that if we are to speak of the content appropriate to an intentional icon we need to focus on the icon's direct proper function, or what Rowlands has called the organismic proper function since this alone reveals the icon's end target.

3.4. A Modest Proposal

Having now defended Millikan's high church version of biosemantics from some recent criticisms, I want to encourage adoption of it in a modest form. Ambitious biosemantic accounts suffer because they attempt to unpack the notion of basic representation in terms of truth-evaluable content.[33] Consider these remarks of Papineau and McGinn.

The biological function of any given type of belief is to be present when a certain condition obtains: that then is the belief's truth condition (Papineau 1987: 64).[34]

...teleology turns into truth conditions....[because a] sensory state fulfils its function of indicating Fs by representing the world as containing Fs; and it fails in discharging that function when what it represents is not how things environmentally are (McGinn 1989: 148, 151).

I am unhappy with these remarks and with Millikan's claim that biosemantic theory provides a 'non-vacuous' ground for a correspondence theory of truth. By such lights all representations, whatever other features they exhibit or fail to exhibit, have truth-conditional content. While the scope of this claim may give us pause to wonder, the biosemanticist re-assures us that only humans really have beliefs with propositional content; lesser creatures have less sophisticated representations (i.e. proto-beliefs, sub-doxastic states, and the like). Even so, these crude representations can still be true or false. Millikan's examples of simple organisms are specifically meant to, "...make it clear how very local and minimal may be the mirroring of the environment is accomplished by an intentional icon" (Millikan 1993: 106). The thought is that such content enters into our natural history at a very early phase and becomes tied up with more and more complex cognitive dynamics as we travel up the phylogenetic tree and progress up the ontogenetic ladder. It is because we can describe systems of representation of graded complexity that we can explain theemergence of propositional content as a late development. For instance, we can mark the differences between creatures which are hard-wired for a particular environment from those which display plasticity (i.e. the ability to learn to cope with new environments). This point is crucial to note lest we be led astray by talk of bees and frogs into thinking that there are no differences between their forms of representation and ours.

Millikan lists six fundamental differences between human and animal representations which are meant to, "...secure our superiority, [and] make us feel comfortably more endowed with mind" (Millikan 1989a: 297).[35] The most important one on her list is the fact that we are able to make logical inferences by means of propositional content. Thus only representations of the kind which respect the law of non-contradiction can be deemed to have propositional content. In a nutshell, she holds that there are distinct types and levels of representation and that not all representations have the kind of content appropriate to full-fledged beliefs or desires. What this means is that biosemanticists need not, and should not, hold that content of the frog's intentional icon is captured by the conceptual content of the English sentence "There is an edible bug" or any other near equivalent. Millikan is explicit about this. With reference to bees she writes:

> Bee dances, though (as I will argue) these are intentional items, do not contain denotative elements, because interpreter bees (presumably) do not identify the referents of these devices but merely react to them appropriately (Millikan 1984: 71).

What I take from this remark is that we identify the object that the bee is directed at as nectar using our own conceptual scheme. Indeed we settle on this description because it is explanatorily relevant when giving a full, selectionist explanation of the proper function of bee dances. This much is incontestable. Yet due consideration of this fact reveals that even though Fodor's critique concerning the intensional indeterminacy of our selectionist explanations fails to undermine the biosemantic project, it is apposite because it reminds us that such descriptions are intensional in a way that the content of the icons is not. Thus any attempt to state the content of intentional icons in conceptual terms is an inappropriate attempt to deploy our own standard scheme of reference.

But if one is willing to concede this then it is difficult to see what could motivate thinking that basic representations have truth conditions. If icons lack intensional content then it is surely misguided to think of their mappings to the world in such terms. If we accept that they are not propositions, and that sense determines reference, we might ask: What is true? How can we have a truth relation if one of the crucial *relata* is absent. The situation is rather like that which inflamed the argument, in *Alice in Wonderland*, over whether or not the Cheshire Cat could be beheaded. It went as such:

> The executioner's argument was, that you couldn't cut off a head unless there was a body to cut it off from: that he had never had to do such a thing before, and he wasn't going to begin at his time of life.
>
> The King's argument was, that anything that had a head could be beheaded, and that you weren't to talk nonsense (Carroll, *Alice in Wonderland*: 116-117).

I am with the executioner on this one. Hence, even if a modest version of biosemantics gives us a handle on the bivalent nature of the intentional content, it is a mistake to think of such content as truth-conditional. Minimally to speak of truth requires that the subject in question has a capacity for propositional judgement. As Dummett notes, "In order to say anything illuminating about the concept of truth, then, we must link it with that of judgement or assertion" (Dummett 1993: 157). How to understand the conditions that make talk and the assessment of truth possible is a complicated business. Hence Dummett makes the further claim that, "A philosophical account of judgements lies, however, in their having a further significance than merely being right or wrong: they build up our

picture of the world" (Dummett 1993: 157). I will return to this issue and explore it from a Davidsonian perspective in chapter five, section 2.2.

In contrast to truth-conditional versions of biosemantics, my alternative proposal is much more conservative. I suggest that organisms are informationally sensitive to features of their target objects in ways that enable them to engage with them successfully. Information sensitivity, as I am using it, should not be understood in Dretskean form. It does not require that there be unequivocal, perfect correlations between the source and receiver, although it is a move in this direction. We can take on board Neander's point about low-level teleofunctions and admit that in some circumstances there might be malfunctions even at the lowest level of performance. An organism's perceptual mechanisms are only supposed to respond to certain features of things, but even the most reliable devices can be tricked and tampered with. In this regard, it is useful to recall Akins' description of the phase-locked character of the FM neuron in the bat's auditory cortex (as discussed in chapter two, section 4.3). Recall that the neurons fire in response two auditory stimuli which, in the bats' normal environment, are reliably co-varied with the beat of wings of certain type of insects. This reliability is good but not perfect. But if we know that the creature is in a historically normal environment (regard this as the background knowledge, k, in Dretske's original definition) then we can say the its neural response carries information about the beat of the insects wing. Of course, to respond appropriately the bat does not, and does not need to, extract this information.

The big issue in understanding this aright is not to determine what bats are informationally sensitive to (in terms of the mechanics of its response), but what they are intentionally directed at. To understand intentionality aright, it proves useful to revive its original, medieval image of, "...aiming an arrow at something (*intendere acrum in*)" (Dennett 1997: 48). With this in mind, we can ask: Should we say the bats' neural responses are directed at the beat of insect wings? As we have seen, one reason to think so is because their perceptual or neural systems may appear to have this as their proper function. That is to say, we note that they can malfunction in the performance of their duty. But as I have already argued, noting this is not sufficient for the ascription of intentional content. Although responding to the wing beat of certain insects is part of the bats' co-ordinated response to their usual food source, we must remember that it is still only *part* of a co-evolved package. Informational sensitivity to a certain feature or features is rarely an end itself for most creatures. In the bats' case, their FM neurons would not be responding thusly unless they competitively enabled bats of this type to get enough food. They do that by getting insects. Hence it is towards insects that they are directed. These are the focus of their intentionality. Anything short of the whole insect will fail to meet the bats' minimal nutritional needs. A response to anything less would be insufficient to explain why the mechanism proliferated.

Conversely bats are not directed at the protein an insect contains for the obverse reason. In their home environment targeting anything more refined than insects is more than they require to meet their basic needs.

In summation, while organisms are informationally sensitive to certain features of the world they are not (usually) directed at those features. Rather they are directed at objects in the world which enable them to meet their basic needs. Their informationally sensitive perceptual systems provide *the means* of detecting, and thus competing, for what they require. This is illustrated by figure 14. Two organisms of the same kind, ∂ and ß, can be intentionally directed at the same type of objects, Xs, because these objects provide a basic resource which they both require (e.g. nutrition). But how they detect Xs will vary if their, perceptual and neural mechanisms are informationally sensitive to different features of Xs, as genetic variation ensures. Of course, which features are better to be sensitive to depends on the context. Naturally this is not the whole story. The contest between the two organisms can be played out on other fronts as well. For example, although ∂ may be a poor hunter-gatherer, he may be a better lover. Nevertheless even this crude rendering of the fight for survival is sufficient to enable us to draw the distinction between intentional direction and informational sensitivity.

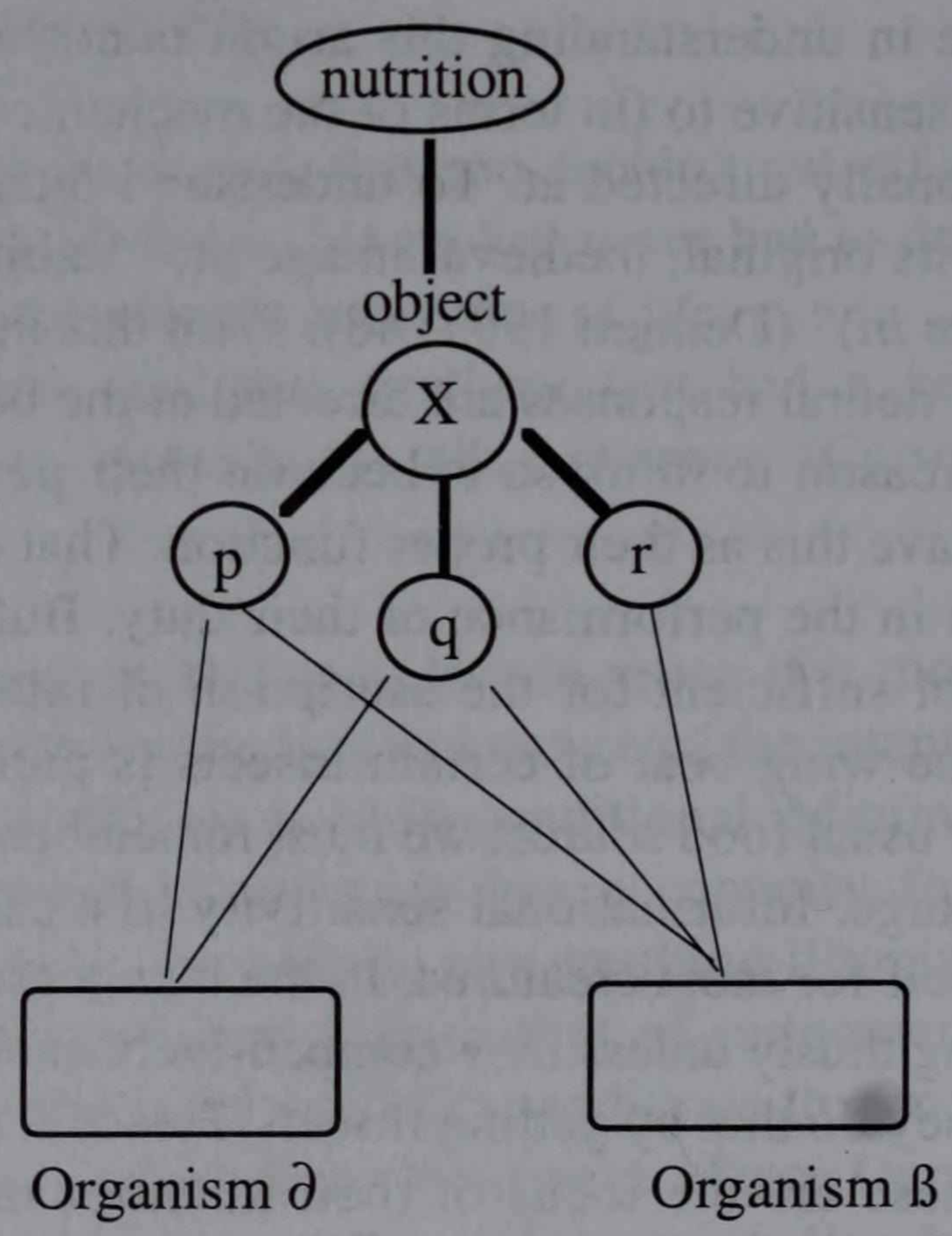

Figure 14. *Intentional direction and informational sensitivity*

A good way to make sense of this suggestion is to compare it with Akins' neurophysiological challenge to the traditional view of perception. Armstrong speaks for the tradition when he says, "...the senses inform us about what is going on in our environment (that is the evolutionary reason for their existence) and, sometimes deceive us" (Armstrong 1973: 27). To make us reconsider this standard philosophical assumption, Akins describes the way in which our thermoreceptors respond to changes in temperature. She reports that, "...our sensations are the result of the action of four different types of receptors: two thermoreceptors 'warm spots' and 'cold spots', and two pain receptors (nociceptors) that fire only in the extreme conditions of very high or very low temperature" (Akins 1996: 346). Each thermoreceptor has a static as well as a dynamic function. The static function of both cold and warm spots is to respond to a constant temperature, but they do so in different ways. For example, warm spots respond only to a narrow temperature range with an increase in activity as they reach the top of the range, then they quickly cease responding. In contrast, cold spots respond to a wider range of temperatures and their maximal response comes at the centre-point of this range. In light of this, Akins notes that, "The static functions of neither the warm spots nor the cold spots are thermometer-like with a certain set increase in firing rate per degree of temperature change" (Akins 1996: 347). Things get even more complicated when we consider the dynamic functions of these thermoreceptors. Warm spots respond to temperature increases by increasing activity until they obtain a stable higher base rate, but the degree of activity varies in relation to the initial temperature. With respect to temperature decreases the rate of firing simply tampers off dropping from one plateau to another. The dynamic function of cold spots is the reverse of this – firing increases as temperature decreases and *vice versa*.

Having outlined the various mechanisms that underpin thermoreception, Akins then invites us to:

> ...reconsider the old illusion created by placing one hand in cold water, the other in hot, and then, after a few minutes, placing both hands simultaneously in some tepid water. Stupid sensors. They tell you that the tepid water is two different temperatures. But the sensors seem less dull-witted if you think of them as telling you how a stimulus is affecting your skin – that one hand is rapidly cooling while the other is rapidly warming. Skin damage can occur from extremes of temperature (being burnt or frozen) but also from rapid temperature changes alone, even if the changes occur within the range for healthy skin (Akins 1996: 349).

She concludes that, "What the organism is worried about, in the best of narcissistic traditions, is its own comfort. The system is not asking 'What is it like out there?' – a question about the objective temperature states of the body's skin" (Akins 1996: 349). Although thermoreception is but one case of sensory perception it is sufficient to cast doubt on the traditional view of the function of senses because it reveals that veridical representation would, in many cases, be evolutionarily excessive and expensive (cf. Clark 1989: 64). This sits well with Stich's observation that, "...natural selection does not care about truth; it cares about reproductive success" (Stich 1990d: 62).[36]

In dropping the traditional assumption about the function of the senses, Akins' final analysis brings exactly the right level of sophistication to bear on the issue of how we should understand biologically directed responses. She writes:

> Of course, it is true that, as a whole, an animal's behaviour must be directed toward certain salient objects or properties of its environment. Objects (and their properties) are important to the survival of all creatures. But from this fact alone, one cannot infer that the system of sensory encoding, used to produce that behaviour, uses a veridical encoding. That is, it does not follow from the fact that the owl's behaviour is *directed toward* the mouse or that the brain states bear systematic relations to stimuli, that there are any states of the owl's sensory system that are *about* or serve to detect the property of being a mouse (Akins 1996: 364).[37]

Standardly in thinking about representations there is a tendency to reify. On the one hand, there is the inner representation. On the other, is the external object it represents when suitably related. But I have argued that intentional icons do not represent objects *per se*, even though they are directed at them. Nevertheless, even if one takes seriously the modest biosemantics I have advocated, one could attempt to preserve this kind of picture is by thinking of iconic contents as representing Gibsonian affordances. Affordances are defined as:

> ...relational properties of things; they have to be specified relative to the creature in question....Thus, the surface of a lake affords neither support nor easy locomotion for a horse, but it offers both of these things for a water bug. To speak of an affordance is to speak elliptically; an affordance exists only in relation to particular organisms (Rowlands 1997: 287).

Armed with this notion, Rowlands suggests that the organism must be able to detect the affordances of its environment (as they relate to it) but not necessarily the objects of the environment *per se* (as we might describe them from our perspective). From this angle, he describes the direct proper function of the

rattlesnake's detection-response mechanism as designed, "...to detect a certain affordance of the environment, namely eatability. This allows the attribution of content such as 'eatability!' or 'eatability, there!' to the rattlesnake" (Rowlands 1997: 291). The fact that we describe the proper function of the snake's detection device as one of locating 'mice', and can do so on principled explanatory grounds, is incidental. Although I am sympathetic to the spirit of this proposal, I think it is wrongheaded. It must be asked: What is the value of thinking that organisms have internal representations about creature-relative affordances? None, I think. They drive us in the direction of odd representational contents as well as bizarre metaphysical entities. I will make the second complaint first.

Affordances are, at best, explanatorily superfluous and metaphysically extravagant. If we consider figure 14 again instead of simply talking in terms of organism ∂ and ß having different means of detecting a common object of concern, we must consider that if they have different discriminatory capacities they have different representations. But, since these representations necessarily only relate to subjective-relative properties, they will be representations of different 'things'. This would constitute a dramatic increase in our ontological economy. Every subtle difference in discriminatory capacity would need to be matched by a detectable creature-relative property. Before we endorse such ontological overcrowding we ought to be certain it is necessary. And, as long as we do not misleadingly reify representations, it is not. Even putting to one side the peculiar metaphysics the affordances approach inspires, it is clearly obstructive to biological explanation. This is because organisms must be oriented towards a common world if they are to compete for its resources. This problem would need to be addressed if an affordance account were to be made viable.

Furthermore we must be wary the suggestion that we ought to positively designate the content of intentional icons in terms of a creature-relative concepts such as 'eatability'. Consider Elder's embarrassing attempt to provide such a designation in the case of the marine bacteria. He writes:

> The bacteria have not a single thought about oxygen, and could not recognize it if it were right in front of them. So it is misleading to suggest that the content of a given tug is 'oxygen-free water thither'; it would be better to say, 'safe travel that-away' (Elder 1998: 360).

Better, but not good enough. The problem with these awkward descriptions is that intentional icons simply do not have Fregean sense of any kind. In an important sense, even though they can be directed at things and features of things, they are not about anything, if aboutness requires us to say how the thing is thought about (cf. Dennett 1997: 49). We must distinguish between having contentful thoughts about things and being merely directed at them.

For all these reasons, I am suspicious of the idea that the biological norms which underwrite intentional content could be straightforwardly deployed in, "...flatfooted correspondence views of representation and truth" (Millikan 1993: 12).[38] Contra Millikan, I maintain that although creatures are normatively directed at features of the world the correctness conditions which underpin intentionality are not best understood as truth conditions. Biosemanticists should not assume that natural selection is sufficient to explain veridical responses, even if intentionality plays a role in underpinning propositional judgements, which can be true or false. This might appear to be a major concession to its critics who have argued that, at best, appeal to evolutionary accounts results in an understanding of proto-concepts and sub-doxastic states (cf. Putnam 1993: 31-33). For the most part, I am prepared to make this concession – even though I balk at the labels, but I explain why it is not damning in the next chapter.

4. Conclusion

My claim is that a modest version of biosemantics is well placed to explain the bivalent character of intentional content without sacrificing scientific respectability. It does so because the norms it appeals to rest on nothing over and above the integrity of the notion of proper functions understood in terms of natural selection. I reserve chapter five for a discussion of how this naturalistic account of intentional content might plausibly play a part in underpinning truth-evaluable, propositional content. However, before turning my attention to this issue, I want to further clarify the nonconceptual character of intentional content. For it is with such an analysis in hand that it becomes apparent both just how well this account sits with respect to cutting-edge work in today's cognitive science and how distant it is from providing the means of naturalising commonsense psychology proper.

CHAPTER FOUR

Seeing without Believing

"...much of psychological life can only be captured
by means of, and should, therefore be modelled
in terms of non-conceptual content."

– Cussins, *"The Connectionist Construction of Concepts"*, 1990

1. Introduction

Having now introduced and defended an appropriately modest biosemantic theory of intentionality this chapter attempts to augment it, by clarifying how it complements and is complemented by an account of basic nonconceptual content and its appropriate vehicles. I claim that a modest biosemantic theory provides just the right nonconceptualist approach to content. Crucially what is needed is an account of content that allows for correctness conditions which are not truth conditions *per se*. And, if this case can be made, I will have the beginnings of an argument which demonstrates why commonsense psychology is not threatened by the revolutionary developments of cognitive science; in particular, connectionist approaches to information processing.

2. Nonconceptual Content Introduced

In recent times the question of whether or not there is such a thing as nonconceptual content has been the object of much serious attention. For analytical philosophers, the *locus classicus* of the claim that there is such a phenomena is found in Evans' remarks about perceptual experience in *Varieties of Reference*. He wrote:

> In general, we may regard a perceptual experience as an informational state of the subject: it has a certain *content* – the world is represented a certain way – and hence it permits of a non-derivative classification as *true* or *false*. For an internal state to be so regarded, it must have appropriate connections

> with behaviour – it must have a certain motive force upon the actions of the subject....The informational states which a subject acquires through perception are *non-conceptual*, or *nonconceptualised*. Judgements *based upon* such states necessarily involve conceptualisation (Evans 1982: 226-227).

Reading this passage alone one might wonder exactly what a nonconceptual informational state is supposed to be. In 1990 Cussins helpfully distinguished conceptual from nonconceptual content in the following, formal way:

> Conceptual properties are those cannonically characterized by means of concepts that a creature must possess for that property to be applicable to it.
>
> Nonconceptual properties are those cannonically characterised in terms of concepts which a creature need not possess for the property to apply.

An easy and useful illustration of this distinction is provided by the well-known Müller-Lyer perceptual illusion (cf. also Crane 1992: 150-151, Jacob 1997: 67–69). A version of the illusion is presented in figure 13. When confronted with this visual illusion we see one thing but believe another. We know the lines are the same length, but even so one of them still looks longer. It is clear that a judgement with a propositional content such as: 'The lines are the same length' must be characterised in terms of concepts possessed by the subject (or as some would say, it is composed of them). In contrast, the purely perceptual response to the illusion is distinguished by the fact that those same concepts are inappropriate in a principled description of its content. This example is also useful in emphasising a primary motivation for drawing the distinction between conceptual and nonconceptual content, which is to demonstrate that nonconceptual responses can be behaviourally influential and relevant. For example, it is not hard to imagine situations in which perceptual illusions influence actions, although in the standard setting their effects are generally limited to perceptual experiences alone.

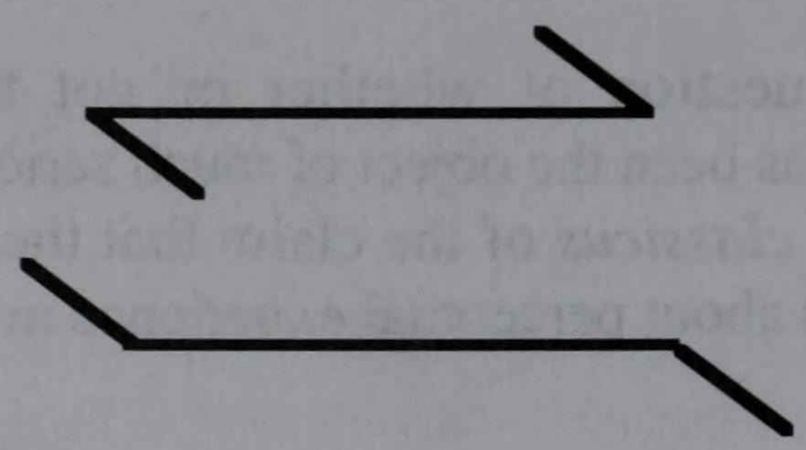

Figure 15. *A version of the Müller-Lyer illusion.*

This is clear if we consider the plausible hypothesis that the Müller-Lyer illusion works on us because in a normal 3D setting the kind of perceptual responses it inspires help us in the detection of edges.

2.1. The Virtues of Nonconceptual Content

The idea that nonconceptual content might be a genuine psychological phenomena has excited the imagination of both philosophers and psychologists alike. This for several prominent reasons. It promises to provide: (i) a working insight into the mentality of pre-verbals and animals which is useful to comparative psychology, (ii) a means of representing the fine-grained character of perceptual experience that concepts are allegedly ill-suited to capture, (iii) answers to certain puzzles about the nature of perceptual relativism and visual illusions, and possibly (iv) a means of characterising demonstrative and indexical mental states which cannot be understood with reference to conceptual content alone.

Take the situation in comparative psychology, sometimes regarded by its practitioners as animal cognitive science. After a comprehensive review of the field, Chater and Heyes have complained that there are no theories or methods by means of which to pin down the content of animal concepts in a rigorous way. Their analysis suggests that the problem is widespread and applies both to standard theories of concepts, taken from cognitive psychology, as well as the experimental methodologies of comparative psychologists in the laboratory and the field. The basis of their investigation is the assumption that, "...studies of discrimination in animals cannot be understood to reveal anything about animal concepts unless there is a sense of 'concept' that (1) applies to humans, (2) applies to nonlinguistic animals, and (3) allows the investigation of concepts in animals" (Chater and Heyes 1994: 235).

The three main theories of concepts in cognitive psychology are definitional, exemplar and prototype theories. Definitional theory characterises concept possession in terms of a subject's grasp of the necessary and sufficient conditions that give the concept its definitional boundaries and sense. Thus each higher level concept can be decomposed into simpler ones. To borrow the most common example, to posses the concept of 'bachelor' is to be able to define it by appeal to the constitutive concepts, 'unmarried' and 'man'. Or, as Fodor expresses it, the thesis is that, "...having a concept is being prepared to draw (or otherwise acknowledge) the inferences that define it" (Fodor 1998: 43, cf. also Clark 1993b: 89). According to exemplar theory, concept possession is essentially the ability to classify items as falling under a concept by comparing their similarity to a concrete example which acts as a standard. Consequently if a swallow is the subject's paradigm example of a bird then whether or not particular items will be judged to be birds is fixed by their relative similarity to this standard. In this case, we can expect that the subject will be more likely to judge that eagles, rather than

ostriches, fall under the concept 'bird'. Prototype theory is akin to exemplar theory in that it also relies on idea of similarity judgements or discriminations. The main difference between these approaches is that, according to prototype theory, judgements do not rest on a comparative appeal to a single standard-setting item. Instead, they depend on identifying the target item as exhibiting the statistically frequent, prototypical properties of the type. On this view, concept possession is just the having of a discriminatory ability to judge whether a given item has features which are statistically associated with the class of things which give the concept its boundaries.

It is important to note that all of these theories can be advanced in a conceptualist form. This is most obvious in the case of the definitional theory, since the definition which provides conceptual boundaries is supposed to be composed of simpler concepts. Yet it is also possible to understand exemplar and prototype theories in a comparable fashion. It might be thought that basic conceptual judgements of some kind underwrite the judgements that some entity exhibits features similar enough to the exemplar or which are prototypical of the class and hence warrant inclusion in it. For example, such judgements might rest on the conceptually based beliefs that 'birds are usually feathered' or that 'birds usually fly' (cf. Fodor 1998: 92). However in the domain of comparative psychology, a conceptual reading of these theories will only provide useful parameters in understanding the nature of concepts if we can comfortably assume that subjects in question have mastery of some basic concepts to begin with. But the only way we can make sense of this idea is if we presuppose, "...that animal concepts correspond closely to the terms of human language" (Chater and Heyes 1994: 214). Without that unwarranted assumption in play, conceptualist versions of these theories fail to meet criteria (2) to (3).

Malcolm provides a philosophically motivated illustration of the problem this generates by imagining a dog who is barking up the wrong tree in pursuit of a cat (cf. Malcolm 1997: 49-50). Although, initially, it seems quite unproblematic to say of the dog that it believes or thinks that 'There is a cat up the tree', there is no sound basis for thinking that the concepts 'tree' or 'cat', as we understand them, play any role in the psychology of dogs. Dogs do not systematically distinguish propositions such as 'There is a cat up the tree' from a host of other closely related propositions, such as: 'There is a four-legged feline in the oak' or 'There is something worth chasing in this arboreous giant'. Without other constraints, we might equally ascribe such propositional contents to the dog in similar circumstances. But there are other constraints, and the reason we find it unsatisfactory to make such ascriptions seriously is that non-verbals do not discriminate between different descriptions of states of affairs in the way that language users do. In light of these facts, we mis-describe the intentional object of the dog's pursuit by embedding it in a proposition, however general.

Nevertheless, for some, this is not the natural conclusion. There are those who maintain that dogs have their own concepts which we are simply unable to describe. But a great pressure is put on those who adopt this line, as Bernard Williams reveals. He writes, "We do say about animals that they take things for things, and that they recognise things as things, and in at least certain cases even that they believe that something is something. The puzzle about this is: where does this concept come from which, as it were, we pretend the animal is using?" (Williams 1973: 138). Or in Armstrong's words, "We want to say the dog believes something – but we do not seem able to say what" (Armstrong 1973: 25, cf. also Stich 1979: 18).[39]

A similar problem arises in the case of pre-verbal humans. Marcus remarks:

> The pre-verbal child hears familiar footsteps and believes a person known to her is approaching, a person who perhaps elicits behaviour anticipatory of pleasure. It may not be the anticipated person, and when the child sees this, her behaviour will mark the mistake. But must there be some linguistic *obligato* in the *child* to attribute to her a mistaken belief, or a disappointment? (Marcus 1990: 135)

In response to Marcus' question, the tendency of many cognitive scientists and philosophers is to say: "Yes"! As Carruthers suggests, "...although we may find ourselves forced, implausibly, to describe animals and infant thoughts using adult humans concepts and categories, this is *our* problem not theirs" (cf. Carruthers 1998: 220). But, if one follows this line of thought to its natural end then, given that there is no adequate translation into our concepts, we must presuppose that the child has something like a Fodorian *lingua mentis* (such that the propositions of its inner language are fleshed out by a stock of unique *child* concepts). Clearly, this violates Chater and Heyes' first criterion. For if we remove the assumption that we can freely deploy our concepts to describe animal concepts, we must wonder afresh, " ...how, if at all, the animal defines or conceives of the stimuli to which it is exposed" (Chater and Heyes 1994: 232). Advocates of this position are destined to be victims of the embarrassing question: How can you give a principled account of the conceptual content of the beliefs of non-verbal subjects?

This question is even more pressing since we cannot rely on the general, explanatory stability of the language of thought hypothesis to underwrite alien contents. *A fortiori*, it does not, on its own, underwrite representational content at all. If we simply rely on the resources of the causal-informational accounts, there are no facts about the environment, a non-verbal's internal mechanisms, or its outward responses which entitle us to think that is capable of harbouring propositional contents. Nor are they provided *via* biosemantics. Such content only appears on the scene after an advance has been made on the kind of intentionality that can be understood by appeal to a modest biosemantic theory, as discussed in

chapter three. If ever there was a legitimate use for Ockham's razor it is here. We can neither make sense of the conceptual contents of non-verbals nor could citing them do any useful explanatory work. This verdict tallies with the conclusion reached by Chater and Heyes when they write:

> It may seem that our conclusion that concepts can be ascribed only to language-users implies radical, and perhaps implausible cognitive discontinuities both in evolutionary and developmental terms. Certainly the thrust of much comparative research, across the behaviour analysis and animal cognition divide, has attempted to find continuities between the psychological processes of nonlinguistic animals and humans. Equally, developmental psychology is concerned to view the cognitive psychology of prelinguistic infants in the same theoretical terms as the cognitive psychology of linguistically competent adults. Hence, from both these perspectives, the postulation of such a discontinuity would go against the grain (Chater and Heyes 1994: 236).

Granting this, perhaps the problem is that we have looked for continuity in the wrong place. Contrary to the tradition, it may be that we are more animalistic and less intellectual in many of our responses (cf. Clark 1997). It is precisely in developing this proposal that the idea of nonconceptual content becomes interesting. Although the definitional theory of concepts gives way at this point, it is possible to imagine variants of the exemplar and prototype theories that do not. For such accounts need not treat categorical discriminations as resting on a capacity for conceptually based judgements. Rather continuity in response to prototypically similar things, or those resembling a given standard, might be achieved by nonconceptual discriminatory capacities which can be developed through training. This is an interesting proposal precisely because the standard argument that particular animals must be using concepts is normally based on nothing more than evidence of their capacities for stimulus discrimination and generalisation. Chater and Heyes find this ironic because, "...in the comparative field, stimulus generalization is commonly regarded as an alternative to concept-mediated accounts of categorisation" (Chater and Heyes 1994: 215). This fits with the fact that non-verbals simply do not manifest the kind of discriminatory capacities that are manifested in the speech reports of humans. Such manifestations are a minimal condition, but not the sole one, for belief or concept ascription. The other conditions will be discussed in detail in the next chapter.

For these reasons, it is a good question to ask: Why we ought to think of exemplar and prototype theories as theories of concepts, as opposed to being theories about the character of nonconceptual responses? It is in exploring this possibility that a nonconceptual account of intentionality can play its most

important role: that of providing at least one necessary, bottom-level pillar of the platform upon which conceptual development rests.

2.2. *The Role of Nonconceptual Content in Conceptual Development*

The good news is that if we opt for a nonconceptualist approach to non-verbal responses then *no* concepts need be innate, nor do I think any are. There is a plausible programme afoot, initiated by Cussins, which hopes explain, "...the construction of cognitive properties out of non-cognitive properties" (Cussins 1990: 374). He encourages us to suppose, "...that the mind/world distinction is a phylogenetic or ontogenetic achievement" (Cussins 1990: 409). If he is right nonconceptual content might provide the means for understanding the emergence of objective thought from non-objective responding.

There has been some confusion concerning this issue in the literature. For example, Bermúdez has recently claimed that Peacocke's rejection of the 'Autonomy Thesis' makes the latter's account of nonconceptual content useless in providing a phylogenetic theory of conceptual development. The autonomy thesis states that, "...it is possible for a creature to be in states with nonconceptual content, even though that creature possesses no concepts at all" (Bermúdez 1994: 405). Thus Bermúdez writes if such a thesis, "...is rejected, then it will be *ipso facto* impossible to give at least one type of developmental explanation of conceptual content in terms of nonconceptual content – namely, the type of developmental explanation that involves explaining how a creature in states with only nonconceptual content, such as for example a new-born human infant, can develop into a full-fledged concept user" (Bermúdez 1994: 405).

In reply, Peacocke makes it clear that Bermúdez's criticisms miss their mark. He concedes that there is potentially more than one form of nonconceptual content. In this regard, it is important to note that what he calls 'scenario content' – which is equivalent to the notion of a spatial type – is, "...arguably the most fundamental type of *representational* content" (Peacocke 1992b: 105, emphasis mine, Peacocke 1992a: 64, Peacocke 1994: 420). It is therefore clearly consistent for him to endorse the claim that scenario content requires a rejection of the autonomy thesis without being committed to the view that all forms of nonconceptual content require a similar rejection of that thesis. In particular, he tells us that his views are compatible with acceptance of Cussin's construction-theoretic account of nonconceptual content. He writes:

> Since CT [construction-theoretic] content is distinct from scenario content, and my rejection of the Autonomy Thesis is a rejection of a thesis about scenario content, that rejection poses no obstacle to those who see either a developmental or a phylogenetic progression from states with only CT content to states with conceptual content (Peacocke 1994: 424, cf. also 425).

This is an interesting response because in allowing for the possibility of what Peacocke calls CT content he is allowing for a kind of nonconceptual content that is both 'pre-objective' and which 'lacks correctness conditions' (cf. Peacocke 1994: 423, 424). But I want to argue that although the most basic form of nonconceptual content is pre-objective it nevertheless must have correctness conditions on the pain of not being content at all. It is important to keep these notions distinct.

2.2.1. Non-Objective Thought, the Generality Constraint and Connectionism

If nonconceptual content is to play any role in explaining, or more modestly underpinning, the development of concepts then it follows that it must be pre-objective in character. Cussins, Clark and Chrisley have laid the ground for understanding this aspect of nonconceptual content by excavating the work of Evans and Strawson (Evans 1982, Strawson 1959). They remind us of the important distinction between objective and pre-objective modes of thought. Chrisley writes:

> Truly objective thought is manifested in the possession and maintenance of a unified conceptual framework within which the subject can locate, and thus relate to any other arbitrary object of thought, the bit of the world thought about...
>
> Pre-objective representation involves contents that present the world, but not as the world, not as something that is or can be independent of the subject (Chrisley 1993: 331).

He also provides an illustration of a possible case of pre-objective thought with reference to the responses of infants before they have attained the stage of recognising object permanence. He suggests infants lack the ability to think of objects as existing unperceived and, hence, they clearly lack a conceptual capacity to represent objects *qua* objects in our sense. He writes:

> an infant which cannot....think of a particular object (a glass, say) as existing unseen, but it can represent its mother as being behind, out of view (on the basis of hearing her voice or feeling her arm, say). The contents of such an infant will violate the Generality Constraint, since the infant may be able to think (something like) *glass in front of me* and *mother behind me* but not *glass behind me* (Chrisley 1993: 332).

Whether or not this is the best way to understand the psychology of infants, we can agree that the imagined case is a violation of the 'generality constraint'. Evans

defines the constraint in the following terms, "...if a subject can be credited with the thought a is F then he must have the conceptual resources for entertaining the thought that a is G, for every property of G of which he has a conception." (Evans 1982: 104, cf. also Strawson 1959: ch. 3, Cussins 1992b: 661-663, Clark 1993b: 75–76). This is connected to the idea that creatures which are only capable of nonconceptual responses are limited in their capacity to represent, whereas, in contrast, those capable of more general modes of thought will have the potential to think of an object under several descriptions. For example, in the case of the hypothetical infant, it is a mark of progress that it can come to think of the glass as being potentially both 'in front of me' and 'behind me'.

An useful way to reveal the relevance of this discussion to the account of intentionality developed in chapter three is to consider Clark's description of the limited response capacities of frogs. He writes:

> The frog lacks the general idea of a fly as a part of an objective world that can be related to that world in a wide variety of ways (e.g. being dead, being alive, being nearby, being far away, being crunchy). It knows about flies only from a single and limited perspective: that of 'spotting and eating'. By contrast, a system which met the generality constraint with respect to its ideas of flies would have to be able to think about flies in a much wider variety of ways. This is seen as constituting the move towards objectivity (Clark 1993b: 76, cf. Akins 1996: 369).

It follows from consideration of these cases that the crucial feature of non-objective modes of thought is that they lack the kind of systematicity which is one hallmark of logico-linguistic thinking. Agents which only have a capacity for non-objective thought are simply not capable of making the kinds of systematic, formal substitutions which are bread and butter to conceptual thinkers. The tradition would have it that if an organism can think of some object, x, that it has a property, Fx, then it must also have the ability to think of some other object (of similar type) such that it could have this property as well (e.g. Fy). The same applies to relational forms, such that if a subject can represent xRy, then it must also be capable of representing yRx. To accept this is to accept that such substitutions, which are inherent in compositionality, inference making, and productivity, are also a pervasive criterion of cognition.

This is precisely the line taken by Fodor and Pylyshyn in their famous attack on the possibility of using connectionist architectures to describe cognitive level functions (cf. Fodor and Pylyshyn 1995: 116). Connectionist networks do not employ structured context-invariant symbols, hence they fail to exhibit systematicity . They respond to situations in a distributed, context-sensitive fashion. This is commonly illustrated by Smolensky's coffee cup which is a

paradigm example of what Clark calls a 'blended' response (Clark 1993b: 22). As far as the network's response is concerned, coffee-in-a-cup, coffee-spilt-on-the-table, coffee-in-the-form-of-granules-in-a-jar, are all greeted by distinct, internal patterns of response; having no structural components in common. In other words, for such networks, "...the object/property decomposition is not built-in" (Chrisley 1993: 344). The same goes for the subject-predicate dichotomy. Chrisley supplies a good example of how such blended responses are deployed in connectionist networks. He writes:

> Just consider a network that is simultaneously learning two routes that happen to intersect at a few points. There is nothing that necessitates that the network use the same code for the place in both contexts; as a matter of fact, it might facilitate learning the sequences to have the representations distinct (Chrisley 1993: 346, cf. also Cussins 1990: 426-429, Cussins 1992b).

In this light, there appears to be an important link between non-objective, nonconceptual modes of thought and the non-systematic character of PDP, connectionist processing. Indeed, Clark, Cussins and Chrisley have been at pains to point out that connectionist-style processing *needs* nonconceptual content (and *vice versa*). In doing so, they make a deft riposte to Fodor and Pylyshyn's assault on connectionist architectures, by exposing that argument's reliance on the implicit and unwarranted assumption that all content must be conceptual (cf. Chrisley 1993: 323, 324, Clark 1993b: 11, Cussins 1990: 435). The kind of content most appropriate for non-systematic, context-sensitive PDP-processing is both non-objective and nonconceptual.

If the connectionists are right, an organism's pre-objective responses to the environment are not achieved by means of the manipulation of internal representations in the way that many philosophers (and classical cognitivists) have claimed. This is because, although they may be crucially involved in guiding and controlling actions, the vehicles of content appropriate to such basic responses are not like those that the tradition has postulated. Furthermore nonconceptual contents do not map onto what we would recognise as objective features of the world. Strictly speaking, it would be a mistake to think that they are systematically representative of the objects and features of an external reality as defined by our usual scheme of reference.

It may be thought that this kind of account is, at best, limited to the rather simple kind of perceptually prompted reactions exhibited by creatures such as frogs. Frogs have very simple stimulus-response mechanisms. But surely other more complex creatures deserve the accolade of concept-user. However, by exploring the link between connectionist vehicles and nonconceptual content, it is clear that nonconceptual responses do not have to be as simple and reflexively hard-wired as those of frogs. By means of a connectionist account of information

processing we can account for the behaviour of organisms which are informationally sensitive*via* several modalities and which are thus able to produce very sophisticated and discriminating responses, without relying on concepts. Moreover we can even account for creatures which check the nature of their prey or situation before reacting and for those which can generalise, making decisions and remember things. If this can be demonstrated then there is no reason to conclude that these feats could not be performed wholly nonconceptually.

Making this case is crucial since such acts are often considered to be impossible without some degree of conceptualisation. For example, Millikan claims that, "Preverbal humans, indeed, any animal that collects practical knowledge over time of how to relate to specific stuffs, individuals, and real kinds *must have* concepts of them" (Millikan 1998: 56, emphasis mine). What it the force of this *must*? Millikan claims that such know-how amounts to mining the 'rich inductive potential' of certain relatively stable, re-identifiable environmental items. In effect, this equates to making generalisations about such items. Hence a creature that can re-identify mice will have many imperfectly reliable but nevertheless useful expectations about them and their hidden properties, as gleaned from earlier encounters. She tells a similar, persuasive story for re-identified individuals and stuffs and, deferring to Aristotle, labels them all 'substances' (cf. Millikan 1998: 56-58).

Something like this is broadly correct. But, in light of the above discussion, we ought to be careful not to jump to the conclusion that generalisations involving re-identifiable substances require that they be systematically represented or that concepts need to be, or ought to be, invoked to explain this capacity. This is especially pertinent given that Millikan, herself, recognises:

> Throughout the history of philosophy and psychology, the tendency has been to project into the mind itself the structure of the object grasp by thought. I will argue the contrary, namely that substances are grasped not by understanding the structures or principles that hold them together but by knowing how to exploit these substances for information gathering purposes. Knowing how to use a thing is not knowing facts about its structure (Millikan 1998: 58).

This is a strong move against the prevailing conceptualist tradition. However it is also a move against her own injunction about the need to invoke concepts when it comes to understanding practical knowledge. For in order to explain such knowledge at most what is needed is an explanation of the capacities for re-identification and association. Connectionist networks are renown for their pattern recognition abilities and, to borrow from Clark, they are associative engines *par excellence*. In this regard, he is wont to speak of 'prototype extraction and

generalization'. For example, over time a net will respond to the prototypical, statistically frequent, features of items it encounters during training. Moreover the statistically frequent features become 'highly marked' and 'mutually associated' (cf. Clark 1993b: 20). To return to Millikan's example, mice will be re-identified, not because they share a set of essential features, but because they are prototypically similar. Moreover if the organism has discovered, on several occasions, some relatively hidden properties of mice – those not immediately tied to how they appear; such as how they taste or react – these will be mutually associated and reinforced on subsequent encounters. This is because there will be an increase in the excitatory links, or connections, between these features and other statistically frequent but more manifest features. Such properties might be precisely those of use to the organism in question. For example, it may become sensitive to the way in which mice react when confronted face on rather than from behind. The point is that this kind of practical knowledge can be neatly accounted for by appeal to nothing more elaborate than the microcognitive architecture of connectionist networks.

All the same we might think that connectionism can at best account for the behaviour of Skinnerian as opposed to Popperian creatures (Dennett 1997: 112-123). According to Dennett, Skinnerian creatures are those which can learn by means of reinforcement and association. Indeed he makes the explicit link to connectionism by suggesting that such creatures deal effectively with the world by means of ABC learning - which is an acronym for 'Associationism, Behaviourism and Connectionism'. Certainly learning about mice in a trial-and-error way demands nothing more. In contrast he sees Popperian creatures as distinctly capable of trying out their hypotheses in advance instead of relying on risky, on the job, trial-and-error learning. Popperian creatures use stored information in order to respond to novel situations in a pre-selective fashion. On these grounds, it might be thought that Popperian creatures are necessarily concept users. For instance, Marian Stampe Dawkins favours this as a measure by which to determine whether or not a creature is really capable of conceptual, representational thought. Despite her own well-placed warnings about the dangers or anthropomorphism, she hopes to accumulate evidence that at least some animals are capable of using 'internal representations to work out what to do in a novel situation'. And she suggests they do this by means of deploying rudimentary 'concepts' (cf. Dawkins 1998: 14, 105-114).

Putting aside, for the moment, my earlier concerns about the content of such concepts, I want to question whether the jump from a Skinnerian to a Popperian mind is really a dramatic jump in kind (in every case) or if might merely be an increase in complexity of response along a continuum. Consider that it is often concluded, without further ado, that in order to make a complex 'decision' a creature will necessarily need to deploy concepts and information in the form of

internal representations about discrete features of its world. For example, Dretske once painted a useful picture of the complexities involved in a creature's decision making processes when confronted with a novel scenario (Dretske 1988: 138-141). He asked us to imagine a jackal lingering around the remains of an antelope which it wants but which does not approach directly because of the presence of a tiger, which it fears. Normally the jackal would simply go for the carcass. What could explain this sudden change in behaviour? Dretske claims:

> Part of the explanation for such creative responses must lie in the cognitive differences in these situations. After all, the representation of a tiger being near the food is, we are supposing, a new element in the situation. Furthermore, we can easily imagine that the jackal exploits previously acquired but not-yet-applied bits of knowledge (perhaps acquired by watching older, more experienced jackals in similar situations) in figuring out what to do (Dretske 1988: 141).

The question for the nonconceptually minded connectionist boils down to: Does the creature actually need to manipulate discrete representations of the various features of this novel situation in order to respond intelligently? It would seem not. Even a well trained connectionist network could become sensitive to microfeatures of situations involving antelope carcasses on their own (associated with pursuit) and those involving only tigers (associated with flight). And, given this kind of training, such a network would be able to make an intelligent response to a novel, 'blended' situation involving the mixed features of an antelope-carcass-tiger. Moreover it would do this without employing a structured representation that neatly separated the elements corresponding to predator and prey. Indeed it is just this kind of flexibility that is one of the great virtues of connectionist networks (cf. Clark 1993b: 21–22). Such a capacity is very useful but it is not like 'figuring out what to do' by conceptual means. For this reason, the fact that some creatures make clever responses to novel situations is not sufficient evidence of concept use. Furthermore if the creature's response really was structured the elements might be imagistic as opposed to conceptual. Such a creature might form images and associations which help to guide its actions without having any idea about what it was doing (cf. Carruthers 1998: 221).

In either case, the moves of such a nonconceptual creature, however impressive, would be very specialised. Its 'knowledge' would be context-bound. It would not be possible to re-locate it to a structurally similar, but otherwise completely new environment and expect it to retain its degree of competence (Clark 1993b: 83–86, cf. also 111-112, 224). This is because it would lack a grasp of the abstract description of its situation. For example, we would expect such a creature to have difficulty, at least initially, in adjusting to a new habitat in

which there were entirely different kinds of predators and prey. It would have no structured, representations of things belonging to these categories such that it could easily swap new items into their slots. Such a capacity demands context invariant, systematically recombinable symbols which are part and parcel of more general forms of structural thinking. My point is that such thinking is not necessary in order to meet the practical, cognitive needs of the great majority of creatures.

Note that in making a case for a nonconceptualist reading of animal behaviour I have not challenged the observable data about the capacities of animals. The issue is not empirical, but concerns how we ought to correctly describe how animals manage to do what they do. The moral is that we must not assume that these creatures must be using concepts in order to get by. This is akin to the moral that Clark draws with respect to the interpretation of connectionist networks. He warns us not to, "...be too quick to assume that a network, even an apparently successful one, has actually fixed on the features which you wanted it to fix. An up-and-running network is an opaque beast which requires further analysis if we are to understand what it is actually doing and why" (Clark 1993b: 41). For example, he describes the case of a network which was apparently sensitive to presence and absence of tanks in a series of photographs and performed very well at making this discrimination. However it was later discovered that the network wasn't responding to this difference at all but instead to the differences in light and density that were contrastingly exhibited in the photos with tanks as opposed to those without, due to the lighting conditions at the times they taken.

Nevertheless a final legitimate worry may rear its head at this juncture. It may be questioned whether or not it is possible to lump together the responses of pre-verbal infants, non-verbal animals, and the non-intellectual responses of language-using adult humans. I do not want to suggest that appeals to nonconceptual intentional content and connectionist processing alone will be adequate to explain all the important aspects, or even all the forms, of nonconceptual responding. For example, no mention has been made of the experiential aspect of nonconceptual perceptual responses.[40] So far I have only advanced two main claims. Firstly, that all such responses, whatever other differences that exist between them, fall on the nonconceptual side of the dividing line. Secondly, that the standard arguments to show that such responses *must be* conceptual are overstated. The principled way of drawing the dividing line beween the conceptual and the nonconceptual is to treat systematicity and veridicality among the necessary conditions for the making of conceptual judgements. It follows that conceiving, as opposed to mere perceiving and responding, comes into its own if and only if context-invariant symbols are in play and it is possible to err with respect to truth or falsity. Although the boundary cases will be undoubtedly fuzzy, in their full-blooded forms these are the genuine hallmarks of conceptual thinking.

2.2.2. The Need for Correctness Conditions

This brings us neatly back to a puzzle concerning Peacocke's second, more controversial point about correctness conditions. If we say that a purely perceptual, but potentially action-guiding, response is nonconceptual we still want to regard it as having content. For even in the most basic cases of nonconceptual content it would appear that things, features and situations are presented as being a certain way. It is for this reason that such content is regarded as representational. For example, Bermúdez writes:

> Conceptual and nonconceptual content are distinguished not by whether they are representational, but according to how they represent (Bermúdez 1995: 335).[41]

But it is also standardly supposed that to have any kind of representational content at all it is necessary that there are specifiable correctness conditions. In Crane's words:

> To say that any state has content is just to say that it *represents* the world as being a certain way. It thus has what Peacocke....calls a 'correctness condition' – the condition under which it represents correctly (Crane 1992: 139, emphasis mine).

Advocates of the view that nonconceptual content has correctness conditions generally believe that providing an explanation of conceptual development will be a clearly difficult but not intractable task. They think this precisely because they allow themselves the luxury of a postulated base of nonconceptual, informational states of the kind Evans envisioned. While it is my concern to defend Evans' distinction between the conceptual and the nonconceptual, and indeed to draw the dividing line precisely where he draws it – at the point where judgements come into play – I am alive to a serious charge that McDowell raises against him when he supposes nonconceptual contents to be truth-conditional, information states. McDowell claims that, although it is not immediately obvious, Evans' account is itself, "...a version of the Myth of the Given" (McDowell 1994: 51). It is not obvious because Evans clearly cuts the cords with judgement making when he advocates a nonconceptual account of experience. Nevertheless it is his traditional understanding of content that betrays him. In McDowell's words, "...the word 'content' plays just the role in Evans' account that is played in that position by the fraudulent word 'conceptual'..." (McDowell 1994: 53). This objection sticks. If nonconceptual content is to play any part at all in helping us to understand

conceptual development, even a non-reductive role, then it cannot be that such content is already thought of as truth-conditional.

If the nonconceptualists hope to aid an understanding of the development of full blown concepts by appeal to Evans' notion of content they will be caught on the horns of a dilemma. The first horn is that if nonconceptual content is to play this role, then it cannot be unpacked by means of pregiven notions of truth and reference (cf. Cussins 1992b). The second horn, is that if we do not appeal to truth and reference in order to clarify nonconceptual content then it is unclear that correctness conditions apply in this domain at all. But if we cannot speak of correctness conditions with respect to nonconceptual content then it is doubtful that it is a form of content at all. Consequently the nature of the content that putatively applies to nonconceptual responses stands in desperate need of clarification.

Recognising this, Cussins has recently re-formulated his account of nonconceptual content by appeal to the notion of the 'realm of embodiment' which he contrasts with the 'realm of reference'. He objects to accounts which, "...suppose that representation in embodied practice is to be explained in terms of the prior, and therefore *given*, notions of truth and the realm of reference" (Cussins 1992b: 652 emphasis mine). His is an attempt to prioritise our experiential responding over our capacity to refer when theorising about our most basic cognitive performances. His notion of nonconceptual content is meant to provide 'a genuine notion of significant representation' which does not require taking truth or reference as basic notions. While I agree that this is needed, I cannot see how an appeal to experience alone is sufficient to enable us to understand the kind of intentionality that could potentially underpin an objective perspective on the world.

The good news is that there is nothing wrong in thinking that even the most basic nonconceptual contents have correctness conditions provided we are careful in understanding their source and nature (cf. Brinck 1997: 125-126, 140). I argued in chapter three that a sufficiently modest biosemantics cannot be a truth-conditional account of representation. At best it is only an account of basic intentionality. Eschewing truth conditions enables us to understand the kind of correctness conditions which apply to nonconceptual contents. Non-verbals can, and regularly do, produce very sophisticated, appropriate responses without employing any concepts. Even our own acts of perception and responses to the world are not always (perhaps not even all that frequently) conceptual in character. But this does not mean that such nonconceptual, intentional responses are merely mechanical or dispositional. They have a normative dimension. Unlike Cussins, I do think that, "...human practice shorn of external norms is mere physiology and physics" (Cussins 1992b: 653). Nevertheless the fact is that in the case of basic intentionality the external norms do not derive from human practice. For this

reason, nonconceptual responses can be misdirected in a way that mere mechanical responses cannot. This happens, for example, when an organism responds appropriately to an abnormal environment. Nevertheless such misdirection is not equivalent to misjudgement.

Consider the reaction of a shark to the sound of a splashing paddle. Imagine it fails to distinguish the noise made by the paddle from that made by its usual food source (which we might identify as a struggling fish). If the shark moves to attack the paddle we might say that it 'thought' or 'believed' the paddle was a struggling fish. But making this kind of attribution is nothing more than making the standard, but defeasible, *first* move in interpretation. To seriously make the ascription stick would require, in part, that we find a more complex pattern in the shark's responses. Without further evidence it would be wrong to suggest that the shark harbours anything like the concept 'struggling fish'. If we accept this, then questions such as 'Does the shark see the world *as* it truly is?' are exposed as naive. Such 'seeing' requires sophistication of a kind that the shark doesn't require as a hunter calibrated to its environment by natural selection, even if it extends its practical abilities by means of Skinnerian or Popperian learning. We must not infer that a shark sees, or identifies, features of the world *as* anything or *judges* them to be of*this* or*that* kind, from the mere fact that it responds in a biologically purposive and flexible way to them. It would be improper to understand such responses to the world as veridically referential.

If we wanted to know, in more detail, why the shark reacts as it does to the noise of paddles, and what its response is supposed to be directed at, we must turn to facts about its historical and ancestral background and its individual learning history. Evolutionary biology can be called upon to tell us why different types of organisms are equipped with their particular sensory-response devices and at what these ought to be directed. It may, after all, offer an account of original intentionality: but only if such intentionality is thought of in terms of *directedness*.[42] Thus Fodor is wrong to say, "Darwin has nothing to say to Brentano" (Fodor 1990b: 79). If the basic forms of intentionality must be understood only in terms of nonconceptual content then Darwin may have quite a lot to say.[43] A modest biosemantic theory is the best way to understand the character of the most basic form of content which applies to nonconceptual responses.

3. The Failure of Eliminativist Plan A

We are now, finally, in a position to take stock and make a first reply to our initial question: Does cognitive science threaten the status of commonsense psychology? Recall that, according to the eliminativist strategy A, one reason to think it does is

because recent developments in cognitive science concerning connectionist theories of information processing threaten those of classical cognitivism. But even though it is true that classical cognitivism fails to defend commonsense psychology this fact is not sufficient to secure the success of strategy A. Dennett is right to remark, "It is rather curious to say 'I'm going to show you that folk psychology is false by showing you that Jerry Fodor is mistaken'... " (Dennett 1991b: 138).

For the sake of argument, let us say that connectionism will be the victorious psychotechtonic theory. Let us also grant that it is a good approximation of the way in which our neural machinery works. Furthermore let us agree that expansive classical cognitivism (epitomised as some form of *lingua mentis*) is a failure since it cannot explain representational content (cf. Horgan 1993: 200, Clark 1993a: 226-229).[44] Still this is not enough to incriminate commonsense psychology of a capital offence. This is because it is a mistake to link the fate of commonsense psychology to that of expansive classical cognitivism. Only confusion is sponsored when eliminativists say such things as:

> ...the crucial folk psychological tenets [which forge] the link between connectionism and eliminativism are the claims that propositional attitudes are *functionally discrete*, *semantically interpretable* states that play a *causal role* in the production of behaviour (Ramsey *et al.* 1991: 97).

The key factor that misleads certain types of realists (i.e. Fodor, Harman) and their anti-realist opponents (i.e. Ramsey, Stich and the Churchlands) is that they are confusing two importantly different claims. As Dennett notes these are:

(i) Commonsense psychological descriptions yield an objective, real pattern in the world.

(ii) That the real pattern of (i) is produced by another real pattern roughly isomorphic to it within the brains of intelligent creatures (thereby making beliefs, and the like, real entities in the head).[45]

We can see evidence of this confusion in Paul Churchland's remark that, "...neuroscience is unlikely to find 'sentences in the head'...on the strength of this assumption, I am willing to infer that folk psychology is false" (P.M. Churchland 1991: 65). Although I do not agree with Dennett in his views about the reality and determinacy of these patterns, I think he is right to say that the defender of commonsense psychology is only committed to claim (i). In this sense, Stich is correct in thinking that there are no such *things* as the contents of beliefs and desires – where *things* are paradigmatically defined as spatio-temporally items that can act as proximate causes. There are no physical *things* (or non-physical *things*)

identifiable in such terms to which the labels 'belief' or 'desire' refer. But this is not to say that the content of beliefs and desires are not genuine phenomena of interest. It is only to say they should not be reified. With this in mind, I will challenge the idea that commonsense psychologists are committed to the following claims:

(a) Contents are functionally discrete entities which have causal roles, and can in this way explain behaviour.

(b) Contents are identifiable with inner states which are semantically interpretable in the same way we make interpretations of a person's utterances.

In order to fully appreciate the dialectical situation we must note the distinction between intentional psychology, which is concerned only with nonconceptual, intentional responses, and propositional attitude psychology *proper*, which is concerned with a significantly different kind of content. Although I will return to the issue of how propositional content fares with respect to these claims in chapters five and six, I will limit this discussion of the remainder of this chapter to intentional content only.

3.1. Structuring Causes and Biosemantic Explanations

What of the claim that contents are functionally discrete and have causal roles? With respect to intentional content, the behaviour of interest to the biopsychologist, which is classified in terms of proper functions, must be contrasted with the more familiar mechanistic or dispositional understanding of behaviour which drives us to understand and explain it in terms of input-output laws. For this reason, Millikan encourages us to, "...distinguish behaviours from physiological processes" (Millikan 1993: 137). Despite the notable differences in his version of teleofunctionalism, Dretske reaches a similar conclusion. And his analysis has some features which are useful to consider. For example, he claims that there are at least two crucially different ways of understanding behaviour; one which sees it as concerned with bodily movements, the other which sees it as concerned with actions performed for reasons.[46] The first, more commonplace characterisation regards behaviour as a kind of product; a bodily movement produced by some internal state. This is the conception of behaviour at work in Fodor's theory outlined in chapter two, section 2.1. Moreover, because Fodor thinks of behaviour in this fashion, he feels compelled to describe the physically identical movements of the twins (of the Twin Earth scenario) as being the instances of same behaviour.

In contrast, Dretske notes that there is another way in which we can speak of behaviour. It can be identified with the whole series of events that lead up to, and result in, the production of a given bodily movement in particular circumstances. To think of behaviour in this way is to treat it as a process as opposed to a product. It is, therefore, distinct from a form of bodily movement. In this light, when we ask*why* a bodily movement occurred, as opposed to *how* it occurred, we can interpret this question as asking about what put the whole causal process in place. Or more precisely, we can inquire into the conditions that set it up such that the causal process would produce a result of X kind (as opposed to one of Y kind). With this distinction between the two understandings of behaviour in hand, Dretske respectively labels the target of these different inquiries 'triggering causes' and 'structuring causes' (Dretske 1988: ch. 1 & 2, Dretske 1993).

The difference is easily demonstrated by considering the questions we can ask about the way in which a bi-metallic strip enables a thermostat to control the level of a room's heat. Such strips are assigned to this job because they are naturally responsive to changes of temperature. Yet it is easy to imagine that they could be given a different control function if they were tied to some other output mechanism. For example, the strip could be enlisted to control the opening and closing of a garage door. In such a case when the room-temperature falls below 18° Celsius instead of switching on the boiler, the garage door would be rigged to open (Dretske 1988: 41, 1991: 212-213). The triggering cause is the drop in room temperature but the structuring cause is the practical joker who has tied the thermostat to the garage door mechanism. Another of Dretske's examples involves a situation in which the triggering cause of a bomb exploding is a car's reaching a speed of more than sixty-five miles per hour. In this case, the structuring cause is the terrorist's wiring up a semtex device to the speedometer. According to biosemantic theory, natural selection plays the role of the practical joker or the terrorist. In this respect, Millikan's normal explanations can be seen as specifying structuring causes by outlining the historical conditions in which whole stimulus-response devices were selected.

Kim has complained that this kind of account encounters the problem of explanatory exclusion. The thought is that if the physical domain is closed to non-physical influences then we cannot have two separate, properly causal explanations of one and the same event. Thus he holds that, "...if the causal role of reasons is to be preserved within a physicalist framework, the exclusion problem tells us that it must be physically reduced" (Kim 1991: 61). If not we are left with:

> ...phenomena in the world which cannot be explained physically; or, to speak in terms of causes, there are phenomena in the world that have non-physical causes. These things are either physical or nonphysical; if they are

> nonphysical, we have a dualism of entities, and dual systems of causes and explanations; if they are physical, some physical things have only nonphysical causes, a form of Cartesian causal dualism and a violation of the causal closure of the physical (Kim 1991: 60).[47]

But Kim's objection misses the mark. Understood biologically, structuring causes do not explain the same event but an historically distant one that explains why a current process gives the results it does. Yet this does indicate that only triggering causes can be genuinely efficacious and hence causal-explanatory in a Humean sense. Only purely mechanical features get mentioned in input-output laws and it is impossible to understand content in such terms. Millikan rightly says that, "...having a certain history is not, of course, an attribute that has 'causal powers'. Hence reasons cannot be, as such, causes. More generally, that a thing has a teleofunction is a causally impotent fact about it" (Millikan 1993: 186). Intentional contents are not things or events, they are relational properties that can only be understood by appeal to a historical standard which a structuring cause has put in place, *a fortiori*, they are not causes in the mechanistic sense. It is for this reason, not Kim's, that they are not proper candidates for physical reduction. Nor is citing them causal-explanatory. They tell us what a thing ought to have done, not what it has regularly done in the past nor what it will now do. Yet it is precisely because citing contents is not causal-explanatory that such explanations do not compete with those of neurology and physiology. This is what makes them appropriate to the explanation of Millikanesque behaviours. Moreover the biosemantic account is attractive in light of the problems Fodor had in trying to find an explanatory and metaphysical role for content which did not violate the principle of autonomy or conflict with a mechanistic vision of implementation levels.

Having said this, it is worth digressing to respond to those philosophers who are suspicious of the biosemantic approach because of the way it handles the 'ahistorical replica' problem. The challenge to biosemanticists is issued in this form: Imagine a creature, ∂, which happens to be an exact physical replica of some other creature, ß, such that all of their internal and external dispositional responses are the same. Imagine also that creature ß was freakishly created, not evolved. A creature like ß might be a Frankenstein monster, a random mutant thrown up by evolution in a non-selectionist fashion, or the product of AI laboratory. The point is that, "...on the selectionist view, [a proper] function can never be attributed to the random mutant" (G. MacDonald 1989: 198). Opponents of biosemantics claim that because creature ∂ exhibits the same complex patterns of dispositional response as creature ß we must return the same verdict with respect to their intentionality. Either they both have intentionality or neither does. Since a consequence of adopting biosemantic theory is that we must deny this

equivalence, Horgan has suggested that, "...a position this counterintuitive begins to look seriously misguided" (Horgan 1991: 88).

Yet the ahistorical replica scenario presents no real problem. It is not a genuine counterexample since it merely assumes that intentionality can, and must, be ascribed by appeal to behaviourist or functionalist criteria. But it is precisely because dispositions and internal mechanisms alone do not license such ascriptions that they are insufficient to underpin the intentionality of ahistorically generated beings. To say that this is counter-intuitive is not a sufficient reason for rejecting biosemantics. Thinking it is only advertises the character, not the basis, of one's intuition (cf. Millikan 1984: 93, 1993: 23).We must ask whether this intuition is well-grounded. If mechanical dispositions alone provided criteria enough to warrant content ascriptions then there would be no question about the replica. But, for the reasons set out in chapter two, the fact that we cannot rely on mechanics alone in this regard reveals this intuition to be philosophically unstable. As to its source I put this down to our initial, but defeasible, tendency to ascribe possible propositional contents to others on the basis of their situation or outward behaviour. But the replica is a standard case of our finding minds where they are not (cf. Dennett 1997: 43–44). However, unlike in the case of non-verbals, it is not the behaviour that lets the replica down it is its wholly mechanical nature. Once we discover its freakish origins we withdraw our ascription of content.

3.2. Domestic Anthropology and the Limits of Propositional Ascription

What of the claim that commonsense psychologists posit inner states which are semantically interpretable in the way utterances are? Certainly if we think this then there will be large classes of creatures that do not fall under the scope of commonsense psychology. The problem is that ascriptions of propositional content depend upon our capacity to ascribe subjects 'some possible propositional content of our own'. For this reason, Stich regards folk psychology as a species of domestic anthropology and claims it cannot make sense of the beliefs of exotic subjects; children, animals and confused folk (Stich 1982: 163). This is no surprise since, *a fortiori*, non-verbals don't have propositional contents to be so described. In attempting to make such ascriptions we anthropocentrically misrepresent what is going on in their minds.

As was observed in section 2.1, if we understand basic intentionality in nonconceptual fashion then any number of closely related propositional substitutions serve equally well as descriptions of a non-verbal's state of mind. Millikan makes this very point when discussing the content of a frog's perceptual 'representation'. She writes, "What a frog's eye tells a frog's brain may not imply any definite English sentence or sentences, but English sentences can be constructed that will imply the truth of what any particular frog's eye is currently

telling its brain" (Millikan 1993: 119). We can freely make such substitutions because the Fregean restraint on substitutions in belief contexts only applies to conceptual contents.[48] Hence we will have no principled grounds for making a particular propositional ascription.

By drawing the independently motivated distinction between intentional and propositional content, we are able to deal with this problem. Nevertheless this move may give us legitimately pause to wonder: How a nonconceptual account of intentionality fits with our everyday practice of ascribing beliefs, thoughts and desires to animals and very young children? The proposal is not as shocking as it might at first seem. When we ascribe such contents we start from our point of view, using our concepts. This is a legitimate and necessary initial move in any interpretation and it will be discussed more fully in section 4.2 of the next chapter when we give consideration to Davidson's principle of correspondence. For our current purposes, it is sufficient and helpful to note that there are important parallels between Davidson's principle and the fundamental maxim of the simulation theorists. Simulation theorists maintain that in real life cases when we wish to understand the action or behaviour of our fellows we engage in a kind of practical simulation; a species of pretend-play. Simulation is seen strongly linked to our ability to role-play or imaginatively adopt the persona of another (cf. Gordon 1992: 31-32). Gordon describes the process of 'good' simulation in the following way:

> This is not the same as deciding what I myself would do. One tries to make adjustments for relevant differences. In chess, for example, a player would make not only the imaginative shifts required for predicting 'what I would do in his shoes', but the further shifts required for predicting what he will do in his shoes. To this purpose the player might, e.g. simulate a lower level of play, trade one set of idiosyncrasies for another, and above all pretend ignorance of his own (actual) intentions. Army generals, salespeople, and detectives claim to do this sort of thing (Gordon 1986: 162).

Noting that serious simulation requires us to bear in mind relevant differences is absolutely crucial. Any serious attempt at interpretation is meant to probe and reveal those differences. A rigorous interpretative investigation of non-verbals reveals the ascription of propositional contents is not sustainable. Our willingness to defend such ascriptions stems from the fact that for non-philosophical purposes a more rigorous investigation is unnecessary. Non-verbals don't really have beliefs if beliefs must be identified by their propositional content. Any terminological trouble which may present itself at this juncture can be resolved by the stipulation that the term 'belief' can be used in different ways. The ordinary concept of belief ranges over cases which, from the philosophical point of view, we should distinguish as instances of beliefs-as-propositional-attitudes and

beliefs-as-intentional-attitudes.[49] The important thing to note is that when referring to purely intentional responses, we cannot ask if the creature represented the world truly or not. We ought to take a Davidsonian stance on this issue and hold that:

> A creature may react with the world in complex ways without entertaining any propositions. It may discriminate among colours, tastes, sounds and shapes. It may 'learn', that is, change its behaviour in ways that preserve its life or increase its food intake. It may 'generalize', in the sense of reacting to new stimuli. Yet none *of this*, no matter how successful by my standards, shows the creature commands the subjective-objective contrast, as required by belief (Davidson 1985c: 480, cf. also Davidson 1991b: 201, Davidson 1992: 263).

Davidson claims that having a 'concept of belief' is an essential ingredient in sponsoring a truth-evaluable perspective on things. To fully understand his reasons for supporting this position we will need to look at his views on interpretation in more detail. This is part of the task of the next chapter. The main point is that once we make the necessary concessions about the limited scope of propositional attitude psychology this aspect of Stich's critique of commonsense psychology loses its bite.

The key objection to commonsense psychology on this score is that it is inconsistent with the demands of a serious scientific psychology. But if we draw a distinction between intentional content and propositional content this conclusion is exposed as being premature. The reputation of commonsense psychology has yet to be tarnished. This for two reasons. Firstly, it has not been established that intentional psychology is at odds with a serious scientific psychology. Secondly, if it isn't, then to make Stich's argument successful it would have to be shown that intentional psychology is at odds with propositional attitude psychology. I attempt to establish that intentional psychology is, in fact, in harmony with cognitive science in the next section and to establish that it is also in harmony with propositional attitude psychology in the next chapter.

3.3. Connectionist Processing, Situated Robotics and Intentional Psychology

There are reasons for thinking that a harmonious relation exists between intentional psychology and recent developments in cognitive science. Consider that traditional views of cognitive processing describe systems whose transactions are made over atomic structures by means of logical rules. Yet such substitutions are not possible in the domain of nonconceptual, intentional thought because such thought is not symbolically structured. This is wholly consistent with a connectionist understanding of the vehicles of content. But, as we have already seen, such vehicles need nonconceptual as opposed to conceptual contents. And I

have argued that a modest theory of biosemantics is the best way to understand the nature of such content.

Furthermore the modest biosemantic theory of nonconceptual content is a good bedfellow for other cutting edge research being carried out in the field of situated robotics (or autonomous agents) by those interested in embedded and embodied cognition (cf. Chrisley 1995). Thus when describing this work Clark employs the euphemism of 'niche-dependent sensing' and reminds us that this notion originates with Von Uexkull who speaks of the *Umwelt*s of particular animals. Here is a quotation from Clark concerning the *Umwelt* of a tick which makes the connection transparent.

> Butyric acid, when detected, induces the tick to loose its hold on a branch and to fall on the animal. Tactile contact extinguishes the olfactory response and initiates a procedure of running about until heat is detected. Detection of heat initiates boring and burrowing (Clark 1997: 24).

Clark's deployment of Von Uexkull's vision is clearly Gibsonian in spirit if not letter. In particular, he only wishes to endorse Gibson once the account is 'sanitised' of its resistance to the postulation of inner mechanisms (cf. Clark 1997: 50). What he wants to retain is, "Von Uexkull's vision....of different animals inhabiting different *effective environments*. The effective environment is defined by the parameters that matter to an animal with a specific lifestyle....Biological cognition is highly selective, and it can sensitize an organism to whatever (often simple) parameters reliably specify states of affairs that matter to the specific life form" (Clark 1997: 25). If we consider this proposal in light of the modest theory of biosemantics advocated in chapter three and dispose of the notion of affordances, with a few adjustments about how to explicate the nature of such directedness, the prospects for unity should be clear.

To summarise, there is no logical difficulty in both defending commonsense psychology and agreeing with Stich that there is a, "...possibility that a sophisticated information processing theory might be constructed, none of whose various sub-components could comfortably be identified with the stores of beliefs and desires postulated by our commonsense theory" (Stich 1979: 18). In fact, like many cognitive scientists, I find this predicted outcome to be not only possible but likely. But while it spells doom for expansive classical cognitivism it should not necessarily worry defenders of commonsense psychology. We must realise these are separate enterprises. And noting this autonomy is quite healthy since it keeps cognitive science from being hide-bound by commonsense categories which do not apply to it and *vice versa.*

4. Conclusion

At this point, eliminativists would be within their rights to object that I have yet to consider propositional content which is crucial to understanding commonsense psychology. This has yet to be naturalistically accounted for. At best I have described the basic form of intentionality which is one of the pre-conditions for the more complex and socially based activity of entertaining propositions and making conceptual judgements. In effect I have only stayed the execution. In the next chapter, I consider Davidson's interpretationalist approach for the ascription of propositional content in order to show how it is complemented by the account of nonconceptual, intentionality I have defended so far. This is important because if a naturalised theory of intentionality is shown to be at least consistent with the Davidsonian approach then I will have succeeded in finally disarming eliminativism's strategy A.

CHAPTER FIVE

Interpreting Minds

"Aristotle once developed the difference between man and animal
in the following way: animals can understand each other
by indicating to each other what excites their desire
so they can seek it, and what injures them, so they can flee from it.
To men alone is *logos* given as well, so that they can make manifest
to each other what is useful and harmful, and therefore
what is right and wrong. A profound thesis."

– Gadamer, *Man and Language,* 1966

1. Introduction

The overarching purpose of this chapter is to outline, clarify and defend a Davidsonian interpretationalist understanding of the ascription of propositional content. However the chapter also has three major sub-aims which constitute its three major sections. Firstly, it seeks to establish that the relations between Davidson's interpretationalism and the kind of biosemantic theory of intentional content, defended in chapters three and four, are harmonious and complementary. Secondly, it shows that there is no tension in accepting interpretationalism and the idea that the vehicles of propositional content are logical structured, external symbols. Finally, it explains how and why interpretative practice is essentially and irreducibly normative.

2. Intensionality Introduced

In order to make the serious ascription of propositional attitudes to a subject it is necessary that they manifest finely discriminating behaviour over a considerable period of time. Davidson observes that only a person's speech acts, for the most part, constitute such finely discriminating behaviour (Davidson 1985c: 475-477, Quine 1960: 3). Consider that in daily life what makes the ascription of propositional content to linguistic utterances expedient is the defeasible assumption

that others are speaking our language. Hence if we don't understand them we can ask questions such as: 'Do you mean *this* (e.g. some proposition) when you say *that* (e.g. some other proposition)'? It is only if a speaker actively makes such choices, by accepting some propositions as adequate descriptions of their words, while rejecting others, that an interpreter is able to ascribe precise propositional contents. In this light, we can see why prompted assent, of some form or another, is the vital and minimal evidence for propositional ascription.

2.1. The Intensionality Test

Philosophically speaking, to discover what someone believes or means we put their utterances through an 'intensionality test'. The intensionality test essentially reveals in what way, and to what degree, a subject's words or thoughts is resistant to co-referential substitution.

Imagine that Farah believes that her family will arrive on the 21st of March, but that she has only limited knowledge about the ways in which that day might be correctly denoted. For example, although it is acceptable to her that we describe her thought by replacing it with propositions (2) and (3) below, she stoutly objects if we attempt to substitute it with proposition (1).

(1) "My family will arrive on the first day of spring"

(2) "My family will arrive on the 21st of March"

(3) "My family will arrive tomorrow"

By considering this simple application of the intensionality test, we learn several things. Firstly, in discovering what a person means by their utterances we are effectively locating the propositional content of such utterances in relation to other propositional contents. The content of any sentence or proposition is determined, at least in part, by its place in the 'logical geography' of contents (Davidson 1984a: 138-139, 225). Notice that the way in which Farah conceives of 'The day her family will arrive' is constrained in part by her other thoughts about that day. Farah's belief could be refined if she discovered that the 21st of March is the first day of spring. This is related to the fact that a subject who is aware of this connection, *ex hypothesi,* has a conception of that day which is different and richer. This is connected with the observations concerning Evans' generality constraint in chapter four, section 2.2.1.

Secondly, noting the pattern of inter-connections between propositional contents is the means by which we decide whether or not someone is expressing a proposition which is the same, or similar, to our own. For instance, if my expression of the proposition 'Scotland is beautiful' is to carry the same or similar

meaning to yours, then we had better agree on quite a few other important propositions; such as 'Scotland is north of England', 'Slugs are not beautiful', 'North is not east', 'Scotland is a country', and so on. As Davidson frequently reminds us, there need be no definite set of beliefs (and attendant propositional contents) upon which we must agree in order to make such similarity claims. If such comparisons are to be plausible there need only be a reasonable degree of overlap (Davidson 1985c: 475, 1984a: 257). The point is that in order to make rigorous linguistic interpretations (i.e. ascriptions of propositional contents), we must discover, in detail, the *patterns* which arise in the speech and thought of others. It is only by charting such patterns that we are able to ascribe thoughts of a truly conceptual and propositional nature.

2.2. The Essential Social Nature of Interpretation

It is vital to realise that it is not just the holistic complexity of fine-grained propositional thought that ties it to linguistic practice. Propositional thought emerges only when one learns to interpret the utterances of others. What is gained in this process is not simply a means of refining pre-existent, but rough-grained, propositional contents. If it were only a case of such refinement along a continuum then one might agree with Lockwood that "...acquiring a language or concepts may be more akin to progressively finer tuning of an instrument with a vast number of strings than it is to learning a set of rules" (Lockwood 1989: 122). This analogy is misleading: the divide is much deeper. To fully understand it we must return to questions about the nature of judgement which we left hanging at the end of chapter three. Consider Dummett's follow up remarks to the idea that judgements 'build up our picture of the world'. He writes:

> A philosophical account of the making of judgements which displayed it as no more than a skill, like playing darts, to be acquired and improved, would ....be utterly inadequate: a judgement has an enduring importance lacked by the throw of a dart. A judgement is the formation of a belief or the acquisition of a piece of knowledge: and this is manifested by our acting on what we know or believe.
>
> The same is apparent in our mastery of language. A child in the earliest stages of acquiring language is trained to say certain things – "Doggie", for instance, or, somewhat later "Doggie is asleep" – in certain recognisable circumstances: and the adults are not, at that stage, much interested in whether his remarks have point or relevance, but only in whether he makes them in the right circumstances, those that warrant them. But if his ability to do this exhausted his linguistic competence, he could not yet be properly said to *say* that anything was so, such as that a dog was there or that the dog was asleep, any more than a dog that has been trained to bark in a particular way when

> seeing his master approaching the house can be said to be saying his master is approaching. Both the child and the dog serve as extensions of others' sensory apparatus; but *neither can be conceived as making assertions until the relation becomes reciprocal, so that they also become capable of acting on the statements of others* (Dummett 1993: 157-158, first emphasis original, second emphasis mine).

Dummett is effectively endorsing the Davidsonian stipulation that we should not ascribe propositional attitudes to any creature which lacks 'a concept of belief'. Having this concept is equivalent to being able to reciprocate in the way described above. It is this capacity which is needed to generate a notion of an objective world. The subject-object split *only* emerges when we have access to an intersubjective standard; a social world.[50] Our grasp of objectivity requires a kind of triangulation with another being who simultaneously responds to things in a largely similar way to ourselves yet in a way that admits of recognition of the possibility of friction between the different perspectives. The importance of the need for an initial 'similarity of response' to 'similar objects or features of the world' is revealed if we follow Davidson's cue and consider a, "...primitive learning situation" (Davidson 1992: 262). For example, he notes:

> ...the child learning the word 'table', has already in effect noted the teacher's responses (mouthing 'table') are similar (rewarding) when its own responses (mouthing 'table') are similar. The teacher on his own part is training the child to make similar responses to what he (the teacher) perceives as similar stimuli. For this to work, it is clear that the innate responses of child and teacher – what they naturally group together – must be much alike (Davidson 1992: 264).

The dynamics of the primitive learning situation are extremely complex and several conditions must be met in order for such learning to be possible. Firstly, the child must respond to a common stimuli in ways similar to that of the teacher. If not the child would be unteachable. We would have no way of correcting it if did not respond to a common distal stimulus, such as the table, as opposed to some private proximal stimuli, such as nerve-endings or retinal images. It is only by locating features of a common world between two interpreting creatures that a key minimal condition for conceptual development and the formation of beliefs about an objective world is met. If such triangulation is to be a real possibility then we must assume that the child is capable of responding to the very same things, in a way that is similar to that of the teacher.

Another requirement is that the child must also be able to recognise that the responses of others are similarly directed and that they are similar to its own. For this reason, Davidson claims that there are, in fact, "...three similarity patterns.

The child finds tables similar, we find tables similar, and we find the child's responses in the presence of tables similar....Given these three patterns of response we can assign a location to the stimuli that elicit the child's responses" (Davidson 1992: 263). But Davidson's remark initially only concerns the teacher who, as master of the concepts, is structuring the context of learning for the child (cf. O'Hear 1991: 55, cf. also 53, Taylor 1989: 38). At some stage, if the child is to learn the concept of 'table', it must reciprocally recognise the very same three similarity patterns. This is the way in which interpreting subjects come to see themselves as playing a part in a 'referential triangle'. It is what allows them to fix upon a *common* focal point. This aspect of triangulation is graphically illustrated by figure 16.[51]

In the case of teaching and learning, mutual recognition of this kind is vital if the learner is ever to master pre-established concepts. Consider Meredith Williams apt analysis of Wittgenstein's related account of such basic learning situations. She writes:

> Only the teacher or master of the practice (the one who is training the learner) has [the normative competence displayed in a rule-governed practice]. What is distinctive, then, about the learning situation is that the learner does not have the necessary background competencies that make what she does naming (when the child calls out 'ball' in the presence of a ball) or counting or identifying a pain. That background structure can only be provided by the social environment personified in the actions of the teacher. The significance of this is that the 'judgements' of the child and the judgements of the adult cannot have the same status or mean the same thing to the two agents...In other words, the initiate speaker, makes judgements, requests, and the like only by virtue of a courtesy extended to the learner by those who have already mastered the practice (M. Williams, 1998: 193–194).

It is only at the stage when the learner can reciprocally pass judgement on the correctness or incorrectness of the teacher's response, when it masters the normative competencies, that it can be properly regarded as expressing propositional thought and having the concept of belief. Only when it can so judge will it meet the final condition for being a true believer – to borrow Dennett's apt phrase. From this point onwards the capacity for potential critique and debate develops over time and marks the onset of genuine conceptual responding. It is in

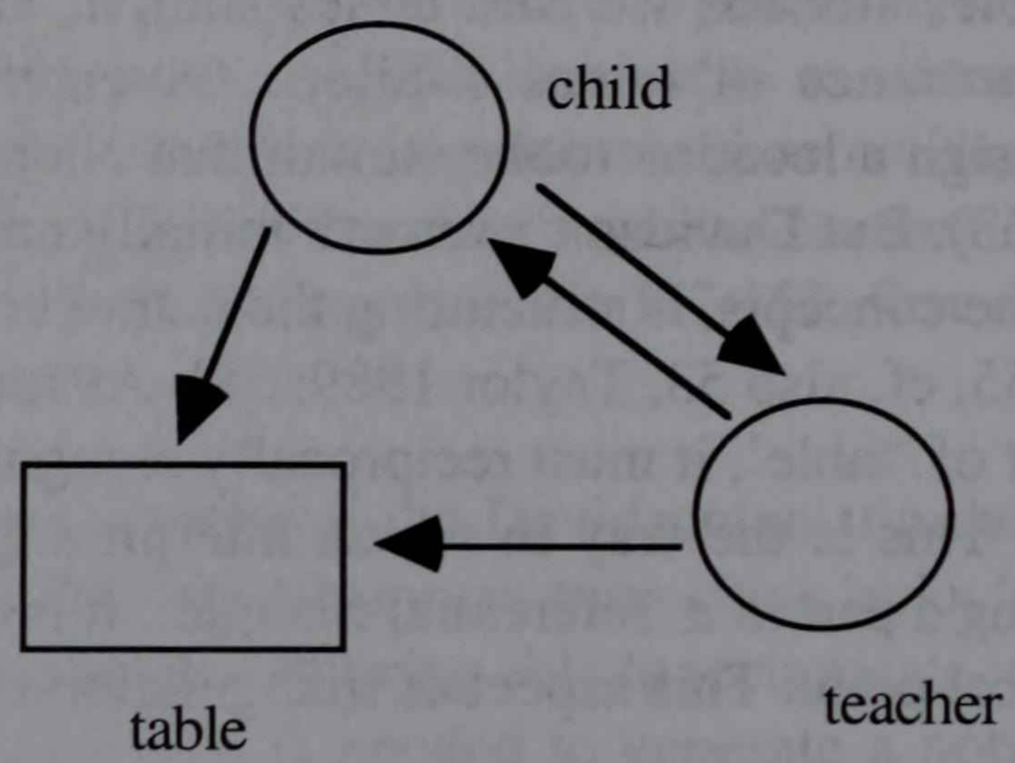

Figure 16. *Davidson's example of triangulation in primitive learning*

this way that "...*conceptual* norms govern our exercises of spontaneity in understanding" (Brandom 1998: 370, emphasis original). This is illustrated in figure 17 by an additional line indicating a capacity for judging the responses of others as being correct or incorrect.

For these reasons a subject can only be counted as a thinker-believer if it is counted as having a conceptual point of view, or perspective, on what there is. The strong claim being made here is that without the capacity for triangulation there is no way to form such a perspective because there is no way to develop the idea of a contrasting perspective on things. For only then can a subject appreciate the distinction between how things appear and how things are. Noting this is vital, for it explains why Davidson holds that only creatures aware of such a subjective-objective contrast can be plausibly counted as assenting or dissenting to a proposition or sentence and hence having a belief with propositional content.

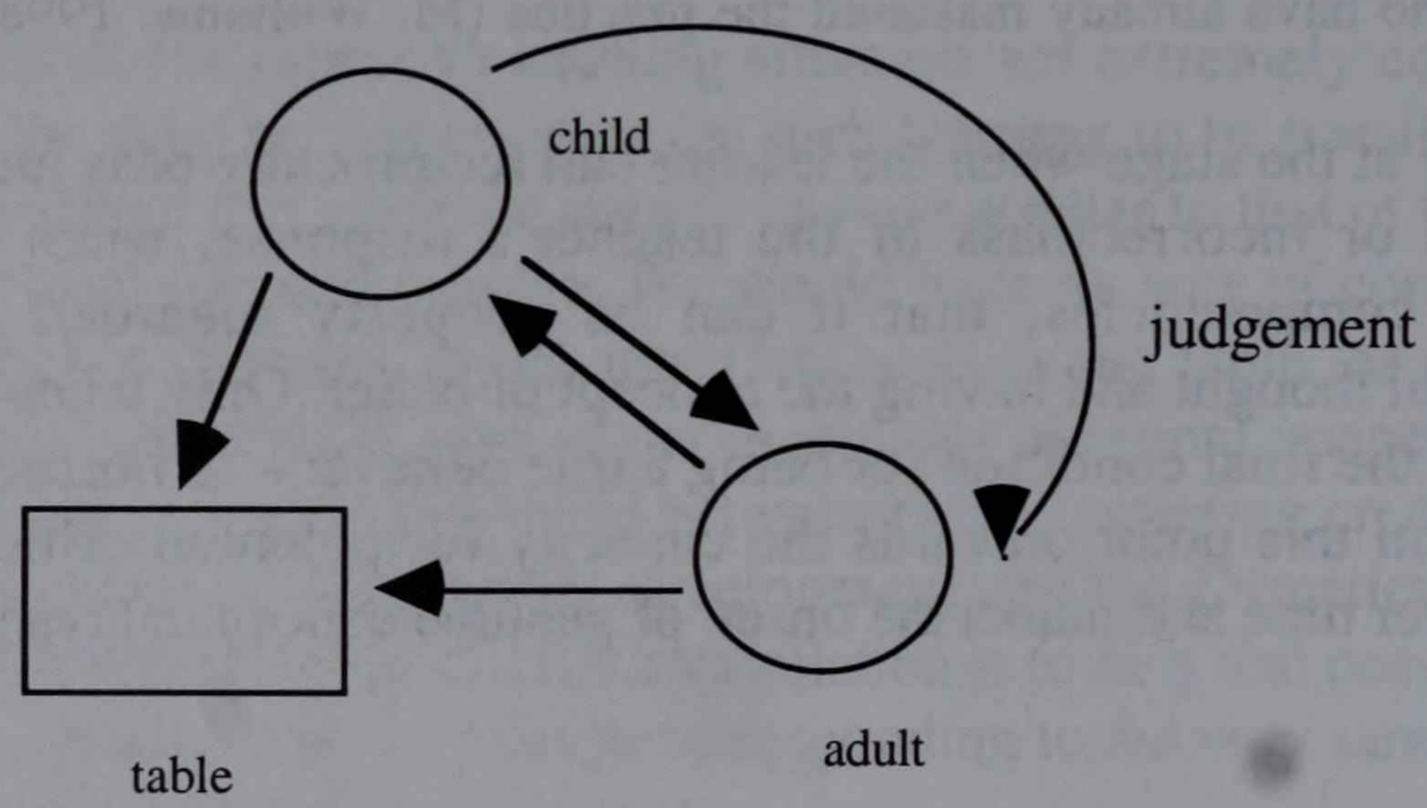

Figure 17. *The development of the concept of belief*

On this account a conceptual response to an objective world and truth conditions emerge in the process of interpreting, learning from and being able to critique others. Hence, not only do non-verbals lack the complexity of response which would enable them to pass the intensionality test, they also lack the capacity to recognise contrasting perspectives and, thereby, the capacity to develop a conceptual view of things. This is the primary reason why Davidson maintains that to be a believer of propositions requires, "...the gift of tongues" (Davidson 1985c: 473). Since the notions of objective truth and conceptual error only arise in the context of interpretation, the concepts of *subject* and *object* also emerge, as it were, simultaneously. Davidson expresses the point when he remarks that, "...the foundations of knowledge must be subjective and objective at once" (Davidson 1986b: 327). Or, as Heidegger would say, they are *equiprimordial* (Heidegger 1962: §43, §200).[52]

2.2.1. The Very Idea of Asocial Triangulation

There have been objections to the idea that social triangulation is necessary for sponsoring a conceptual perspective. For example, Child claims that, "...on the face of it, the idea that what counts as similar to what is partially defined by similarity responses could be applied using only the similarity responses of an individual" (Child 1994: 21). A first reply reminds the objector that in order to form a 'concept of belief' and its attendant notions of truth, falsity and objectivity, one needs to take on the roles of both observer and interpreter, and subsequently that of judge. To do this the subject needs first to recognise the possibility of another perspective in the world, and must be capable of simultaneously comparing this other perspective with its own with respect to some common subject matter. But Child argues that, on its own, this condition does not rule out the possibility of an individual triangulating with itself diachronically by reflecting on its earlier own responses. This kind of reflection appears to be a logical possibility for explaining the emergence of an objective world without bringing in an explicitly social standard. This proposal is illustrated by figure 18.

There are at least two different ways to respond to this criticism. For example, it could be argued, on logical grounds, that Child's proposal is unworkable because it presupposes that the subject is able to conceptually re-identify the same object over time. This, in turn, presupposes that the subject already has a grasp of objective categories (Hutto 1995b: 467-468). Unless this capacity is presupposed solitary triangulation is logically impossible. This line of response is particularly apposite if we accept the account of intentionality outlined in chapters three and four. For even if a creature is capable of responding to its environment in a sophisticated way, this alone does not justify the supposition that its responses are representative of objective features of that environment. Accepting this undercuts

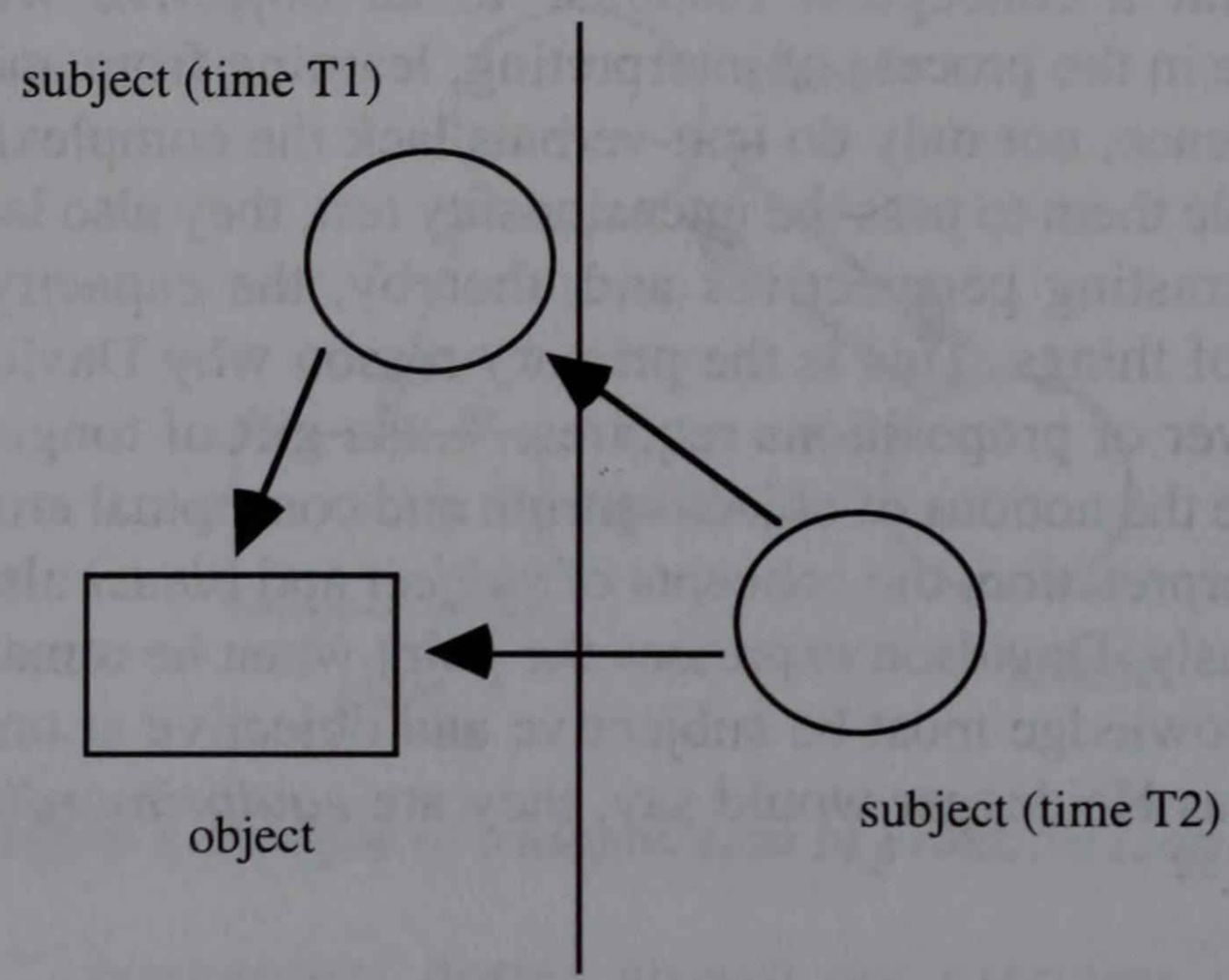

Figure 18. *An individual triangulating with itself diachronically*

what is likely the strongest motivation for wishing to sponsor the solitary triangulation thesis. That is, the idea that: unless a creature is able to form an objective representation of the world it will be unable to successfully engage with it. Yet to assume such a basic capacity clearly begs the point in question. If we consider figure 17 we can see that a capacity for nonconceptual re-identification is a necessary condition for proper triangulation, individual or social.[53] But this does not equate to a capacity for conceptual re-identification. However this debate takes us onto hallowed philosophical ground and this argument alone is unlikely to be persuasive, even if it is sound.

A second line of reply emphasises the fact that triangulation requires more than just a capacity for re-identifying objects. It also requires that the subject recognises that the responses of another can be similar or contrasting. It is this second capacity which does most of the work in making objective thought possible. In this light, we can see the need for social interaction to spur on the development of concepts by considering recent empirical work concerning the effects of enculturation on the great apes, our nearest phylogenetic relatives. Enculturation takes many forms but all involve apes that have been reared by humans. And even though the phenomena has not been fully or systematically investigated at this point, researchers in the field are confident of its general effects on certain ape-individuals. These individuals have developed relatively advanced capacities for communication and sign use due to their contact with human carers (cf. Tomasello and Call 1997: 391, Call and Tomasello, 1996). One thing that makes this possible is the natural simian capacity for joint or shared attention. By this means, "...a simple referential behaviour (e.g. gazing at X) becomes intentionally referential if it's accompanied by an ostensive behaviour (e.g. eye contact)"

(Gómez 1998: 82, 85-86, cf. also Povienelli 1996: 297-299). Moreover, despite other differences, just as in the primitive learning situation, apes set the scene for this kind of activity in complex ways, such as; touching, gesturing, looking into the eyes and so on. These behaviours enable them to initiate and check that there is joint attention. On these grounds, Gómez concludes that, "...all in all, the available evidence suggests that living apes possess an ostensive/referential system of communication..." (Gómez 1998: 84).

Despite this, it is crucially observed that apes do not display a capacity for sign use in the wild or even with conspecifics in captivity in the way they can when dealing with human trainers. Thus Call and Tomasello note that, "Apes with human contact are also the only apes to produce imperative pointing....The use of declarative gestures is much less frequent in apes, and may be confined to apes with extensive human contact" (Call and Tomasello 1996: 382). In order to explain these limitations Tomasello and Call suggest that:

> ...although apes can master the 'referential triangle' in their interactions with humans for instrumental purposes when they are raised in humanlike cultural environments, they still do not attain humanlike social motivations for sharing experience with other intentional beings (Tomasello and Call 1997: 393).

The capacity for ostensive reference and limited sign use in apes only require the ability to notice that another is seeing the same thing in the same way. But it was suggested that a proper mastery of concepts would require more than this. It eventually requires the further ability to recognise that another might see the same thing differently. This key contrastive capacity is, empirically speaking, a relatively late development in humans. For example, if we consider such experiments as the 'false belief' task, it is clear that children do not have the innate resources to fully distinguish alternative perspectives. The classical case of the false belief task is one in which children meet a puppet, called Maxi, who is shown a room in which a pile of biscuits are the main attraction. The puppet leaves the stage and the children watch as the experimenter hides the biscuits in the cupboard. The children form the belief that "The biscuits are in the cupboard". However, when it re-enters the room, it turns out that most three-year-olds have difficulty in ascribing the Maxi-puppet a belief which differs from their own current belief. On these grounds, it has been plausibly suggested that they find it difficult to understand that the puppet can have a perspective on a common scenario which differs from their own. The point is that if children had a built-in command of the concept of belief they would be able to cope with false belief task. But this ability is acquired, developed or kicks in, over time. Whichever developmental story we wish to tell, it is plausible that this capacity is only fully developed when subjects have actually attempted to compare their perspective to

those of others. There is reason to believe that we simply do not recognise the existence of alternative, as opposed to merely other, perspectives until we fully engage in the process of interpreting in a social setting (cf. Smith 1996: 353-354, Leslie and German 1995: 132).

The facts about our own development and the evidence from the study of apes gives strong grounds to suppose that the capacities to recognise other and alternative perspectives only develops in a social context as opposed to simply being triggered off by a time-delayed innate module. All of this casts empirical doubt on the idea that an asocial creature could be in a position to triangulate with itself by reflecting on its earlier responses to a common, re-identifiable stimulus.

2.3. The Origin of Conceptual Perspectives

Furthermore empirical investigations of these kinds could take us some distance in our bid to explain how such perspectives actually develop. This is important because it might supply the means of answering McDowell's charge that the Davidsonian account of triangulation "...smacks of magic" (McDowell 1998: 410). Before sketching a reply to this charge it will prove instructive to consider the standard, alternative proposal.

According to what Carruthers and Boucher label 'the communicative view', language functions simply to provide a public code for the expression of pre-existent conceptually-based thoughts. Hence advocates of this view maintain that our ability to locate a meaningful pattern in the speech acts of others presupposes that we and the subjects of interpretation must have a common *lingua franca* (Carruthers and Boucher 1998: 1). They claim, "...that you cannot learn a language whose terms express semantic properties not expressed by the terms of some language you are already able to use" (Fodor 1975: 61, cf. Chomsky 1988: 1990). But we do not share a common public language with all those we interpret. For example, we do not share a common language with our teachers when, as young children, we are learning our first language. How can this fact be accommodated? The familiar solution is to hold that human beings must come equipped with a private *lingua mentis*. If this were so, our language of thought would provide the medium through which we, our teachers and those we wish to interpret manage to understand each other (cf. Fodor 1998: 5). Consequently, as we have seen, language learning and interpreting simply become, "...problem[s] of finding what labels are used for pre-existing concepts" (Chomsky 1988: 134, cf. Fodor 1998: 9).[54]

Standardly, in line with the proposal that at least some primitive concepts must be innate, theorists like Fodor make appeal to these in order to explain more complex conceptual contents. But what is never explained is from whence such primitive concepts originate.[55] Or at least in Fodor's case, this explanation never

amounts to more than tongue in cheek remarks about what a benevolent God would have bestowed upon us (cf. Fodor 1998: 129).[56] Primitive concepts and their properties as simply assumed because of an explanatory need. This is despite the fact that postulating their existence has a paradoxical consequence – which Fodor reveals when he claims that *all* concepts must be acquired inductively (cf. Fodor 1998: 130-132). Given this it is ironic that in his eyes those who attempt to explain conceptual development by other means are guilty of sweeping, "...the problem from under the hall rug to under the rug in the parlour" (Fodor 1998: 26). But if anything is sweeping a problem under a rug it is the unexplained postulation of conceptual contents all the way down to the bottom rungs of experience. Postulating a universal *lingua mentis* with its crew of innate concepts does no proper explanatory work in this regard (cf. Putnam 1988). When we make appeal to innate, inner languages to explain the development of language we are like men, "...settling how much a toss is to be with another toss" (Wittgenstein 1967: §229). For example, putting aside the problem of naturalising representational content, we may well wonder: What explains the ontogenesis of these innate concepts (cf. Peacocke 1992a: 9)?[57] And by what miracle is it we all have the same ones?

In contrast, in advancing his strong constitutive thesis about the interdependence of conceptual thought and language, Davidson admits to being uninterested in such questions. He writes:

> The approach to the problems of meaning, belief, and desire which I have outlined is not, I am sure it is clear, meant to throw any direct light on how in real life we come to understand each other, nor how we master our first concepts and our first language (Davidson 1990b: 325).

Nonetheless by supplementing his account with the theory of nonconceptual intentionality, as advocated in chapters three and four, we can see a possible source for such answers. Recall that one condition on the possibility of triangulation is that the interpreter and interpretee are capable of being directed at the same focal point and that they can be so directed in a similar way. Although Davidson does not, and would not wish to, explain how or why such responses are similar or how or why they are similarly directed, he tells us nonetheless that, "If some...discriminative mechanisms were not built in, none could be learned" (Davidson 1992: 262, cf. M. Williams 1998: 193). Moreover, he remarks, "...it is we, because of the way we are constructed (evolution has something to do with this), that find these responses natural and easy to class together" (Davidson 1991b: 200).

A modest biosemantic theory of nonconceptual content reveals that this assumption is warranted by appeal to the fact that creatures of the same species

have been calibrated by natural selection to respond to certain objects in their environment and not others. Moreover, by and large, it is reasonable to suppose that creatures of the same species will respond to these items in similar ways (allowing for some variation). If such nonconceptual capacities are a necessary platform for the development of concepts then it is clear that the account of intentionality, outlined in chapters three and four, is not at odds with interpretationalism. And transitively, interpretationalism is therefore not at odds with recent developments in cognitive science either.

All the same, it must be emphasised that treating nonconceptual capacities as a necessary condition for the development of concepts is not an attempt to reductively explain the emergence of concepts. Furthermore if we explain the development of concepts by appeal to the training of individuals this account need not end in an explanatory regress. If triangulation is magical it is still possible to understand how the trick is pulled off. For in order to understand the ultimate origin of the concepts deployed by the teachers we would have but to look to social and linguistic development of *Homo Sapiens* (cf. M. Williams 1998: 196–198, Gómez 1998: 86-87). And here again, in providing such an account, the fact that we have similarly directed, inherited nonconceptual capacities has a vital foundational role to play, even if it cannot provide the whole story.

3. The Internal Structure of Propositions

We might wonder how interpretationalism sits with the fact that natural languages have complex syntax. After all it has long been supposed that we need to postulate an inner module concerned with syntax, which effectively provides us with a universal grammar, in order to explain the structure of natural language. In a directly connected way, this encouraged classical cognitivists to think that all thought must be intrinsically systematic and productive.

3.1. Context Invariant Symbols and External Props

However the interpretationalist line I have so far pursued is consistent with the thesis that internally structured context-invariant symbols provide the basis for the logical properties of higher-order thought. Only on this approach they are not 'stored in the mind'. Such complex structures only appear on the scene when, and if, creatures develop a sophisticated written or spoken language. The idea is that as intersubjective language developed we began to deploy context invariant symbols. Symbols of this kind do not exist as reified inner states of mind but rather as much more mundane items such as words – written and spoken.[58] Given this, *contra* Fodor, structured logical thought develops in tandem with language rather than providing a basis for it. And, echoing Wittgenstein and Ryle, the fact that we learn

to 'think in our heads' is an achievement which comes only after we have learned to think by means of employing such external symbols (cf. Dennett 1998: 284).

This proposal is particularly interesting in light of the fact that connectionist networks don't make transformations over symbolic structures with discrete elements, but instead employ context-sensitive vehicles with nonconceptual content. This was discussed in chapter four, section 2.2.1. This tallies with the fact that:

> There are classes of problems to which PDP [connectionist] approaches are apparently ill suited. These include the serial-reasoning tasks of logical inference, the temporal reasoning tasks of conscious planning, and perhaps the systematic-generative tasks of language production...as we proceed to higher, more-abstract tasks the PDP [connectionist] approach becomes harder to employ (Clark 1989: 127).

I follow Clark in speculating that although the capacity to perform such tasks does not come to us naturally we gain it when we master the use of, "...artifacts as physical representations....[or] external formalisms" (Clark 1989: 132–133). Written numbers are paradigmatic examples of such external symbols.

Furthermore the idea that syntactic structure emerges with the use of complex external symbols fills a crucial explanatory gap. For if the mind is not ready-structured by inborn concepts, then it would be difficult to explain how or why a structured universal grammar would exist prior and independently to any need (for another possible explanation see Gómez 1998: 84). In this light, it is ironic that Fodor rebukes Dennett for ignoring his explanatory burden when the latter suggests that, "...it's learning a language that makes a mind systematic" (Fodor 1998: 26).

Nevertheless it may be argued that the postulation of inner syntax is needed to explain the structural differences between sentences which are ambiguous at the surface level of syntax. Consider the following example:

(4) They are flying aeroplanes

We can parse this sentence in two different, structurally revealing ways to get two different readings.

(5) They [are flying aeroplanes].

(6) They are [flying aeroplanes].

Although I will not explore the issue in depth here, adopting an interpretationalist approach is perfectly able to cope with such structurally

ambiguous sentences without introducing a private, inner syntax. It would appear that the structural differences between (5) and (6), as driven by the differences in their semantic content, demands that there is a difference in their deep syntax. But deep syntax does not have to be thought of as the inner way in which propositions are structured or organised. If sentences are external props or tools, then like any other tool that can perform multiple functions. The ambiguity of which use is in play on a given occasion will be revealed by the context of use. Note that it is the difference in meaning that drives the two readings. And meaning is breathed into written and spoken symbols in the process of interpretation. Without wishing to invite further debate about primate thinking, this last point can be illustrated by considering an overly simplistic, but familiar, account of the quasi-interpretative practice of vervet monkeys.

> Vervet monkeys *Cercopithecus aethiops* give different alarm calls when they detect three different types of predator. Receivers dash into trees when they hear the 'leopard alarm', scan the sky after an 'eagle alarm' and scrutinize the ground after a 'snake alarm' (Seyfarth *et al.*, 1980)....Young vervets have to learn to produce each alarm call in the right circumstances and to respond appropriately to conspecific and starling calls (Seyfarth & Cheney 1986: Hauser, 1988) (Krebs and Davies 1991: 377, cf. also Gómez 1998: 78).[59]

Minimally in order to communicate effectively the monkeys must employ a set of publicly recognisable, re-identifiable symbols. In fact they use different sounds for different alarms. Each sound is associated with a particular type of predator which they can recognise. The sounds are external representations that 'go proxy' for these different creatures. Consideration of this simple case alone enables us to see how interpretationalism is in complete harmony with the view that the vehicles of full-blown propositional thought which act as basic symbolic elements are public and external. If the vervet system of communication were more sophisticated, we would expect to find 'leopard' emerging as a context-invariant symbol in complex sentences. But the fact is that, "...call repertoires *a la* vervets are fixed, not productively generated by a lexical or grammatical syntax" (Gómez 1998: 79). Nevertheless the general rule still applies to our more sophisticated forms of linguistic communication.

3.2. Syntactic Structure and Radical Interpretation

Davidson repeatedly emphasises that propositions have an internal syntactic structure which unites their parts with the parts of other propositions. He argues that this is vital because, given the unlimited number of sentences which can be produced by natural language speakers, it would be impossible to try to radically interpret speech if it had no internal structure. If our natural languages were not

so composed and could not be characterised by a finite set of axioms and compositional rules we would not have the time, given our finite life-spans, to master, employ, or generate the potentially infinite number of sentences that we can (Davidson 1984a: 8, 127-128, Schiffer 1989: 8). It is only by detecting the internal structure within the patterns of sentence use that it is possible to develop an axiomatic base necessary for learning a complex productive language or carrying out the project of radical interpretation.

For example, in a non-compositional 'language'(such that it even makes sense to speak of one), the sentences would be fused wholes. Thus "I-am-going-to-walk-the-dog" in such a language might have nothing in common with the sentence "The-dog-needs-to-go-for-a-walk". The hyphens are meant to indicate that the common structures in such sentences would be invisible (or non-existent). But if this were so then each sentence would require a wholly new and unique interpretation since there would be no way to locate words and phrases the speaker had used before. In contrast we do locate re-occurring word and phrase patterns within compositional speech. In theory a radical interpreter would be able to discern the internal patterns in linguistic use by observing which words and phrases remained constant throughout various contexts of usage and those which do not. For a crude example, imagine sentences (7) to (9) form part of our basic data for a given language:

(7) There is a cold wind today.

(8) Cold is creeping into my hands.

(9) Why is this flat so cold?

If a radical interpreter carefully observed a speaker of this 'language' in various situations, over time, they would develop an axiomatic base in which the re-occurring word 'cold' would play an important role. The point is that if a consistent and logically coherent pattern is locatable in the utterances of another then it is theoretically possible to develop an axiomatic base and series of compositional rules for those utterances. With this in hand, an interpreter could form empirical hypotheses (theorems or T-sentences) about what any given sentence in that 'language' means. The form of such theorems is captured as follows:

(T) 'Roses are red' is true iff p.

'Roses are red' is an utterance in the object (or target) language which is true if only if the conditions outlined in 'p', the meta-language sentence hold (in this case p = roses are red).[60] Thus the basic tool of the radical interpreter is a recursive

theory of truth which lists the satisfaction conditions in the meta-language for any given sentence of the object language as relativized to a language, a speaker, and a time (Davidson 1984a: xiv, 1986a: 315, 1990b: 310, Evnine 1991: 96, Ramberg 1989: 62). In this way an extensional theory of truth, of the kind Davidson advocates, provides the basis for his theory of meaning. This is a theory of truth modelled after Tarski's theory of truth for formalised languages, but which could theoretically be employed with respect to natural languages. In this regard, it is vital to note that despite earlier talk of an 'intensionality test', Davidson follows Tarski's recommendation and treats truth as an extensional relation by replacing the intensional 'means that' with the extensional 'true in L iff' (cf. Evnine 1991: 77). In effect, we can understand meaning in terms of truth, but to do so we must take truth as given; and a modified Tarskian theory does just that (Davidson 1984a: 134, 1990a: 135). If we agree that truth is generated only when the appropriate link is established between the speaker, the interpreter and the world then, to this extent, we should regard it as a basic or primitive concept (Davidson 1984a: 134, 1996).[61]

3.3. Theories of Truth and Theories of Meaning

Some philosophers have raised doubts about the very idea that the exercise in imagination which is radical interpretation could reveal the right necessary conditions for a theory of meaning. They note that even if T-sentences meet with predictive success this would not provide sufficient grounds for the ascription of content. Even if radical interpreters could in principle detect a logical, coherent and internally structured pattern in the speaker's assents and dissents which holds over a period of time this alone would not enable them to determine the meaning of a speaker's utterances. This point is often made in the following way: If we were to add a mathematical or logical truth to the axioms of the T-theory, the logical patterns of assent and dissent – the patterns of internal coherence – would be unaffected although the content of the speaker's utterance would differ. Suppose there are two T-theories such that the axioms of T are the same as T' except for the fact that all the axioms of T' have an additional logical or mathematical truth (say, 4 - 3 = 1) appended to them. T-sentences of the two theories might look like this:

(T) 'Roses are Red' is true in L iff Roses are red

(T') 'Roses are Red' is true in L iff Roses are red and 4 - 3 = 1

We can expect the interpretations of the utterances of a given speaker to display exactly the same level of formal or logical consistency on either theory. That is to say, from the point of view of a purely extensional theory of truth we have no reason to prefer T over T' as the correct T-theory for interpreting the subject.

Nevertheless it is obvious that the predictions of T and T' are not equivalent in meaning. 'Roses are red' does not *mean* the same as 'Roses are red and 4 - 3 = 1'. The problem is pervasive for as Fodor and Lepore note, "...the logical truth problem infects the axioms as well as the derivations" (Fodor and Lepore 1991: 69). Evans and McDowell make this point by insisting, "...more stringent conditions must be imposed upon a theory of truth if it is to serve as a theory of meaning" (Evans and McDowell 1976: xiv, Ramberg 1989: 61).[62] But this is not a interesting objection since Davidson is well aware of the problem. He openly tells us, "T-sentences are truths of logic, and so cannot tell us anything logic cannot tell us" (Davidson 1990b: 290). Or again, in his own words, "A theory of truth, no matter how well selected, is not a theory of meaning" (Davidson 1984a: 179).

4. Normativity and Propositional Ascription

Davidson's reply is that the logical and holistic constraints imposed by a theory of truth only lay the ground for understanding the *structure*, not the content, of propositional thought. Given this, the question for the Davidsonian is: What must we add to a theory of truth in order to make it into a full-fledged theory of meaning? Another way of phrasing the question is: How is it possible to assign meaningful content to the utterances of another when locating a coherent pattern in their responses? His answer is instructive. It reveals, ultimately, how heavily we must rely on normative principles of charity at every stage in the interpretation of the speech and propositional thought of another. In short, his answer is that if the radical interpreters are to discover a complex pattern of speech which yields meaning, they must make some important *a priori* assumptions about their subjects.

4.1. Two Principles of Charity

Two of the basic assumptions of the radical interpreter must make are:

A The subject is trying to make assertions about certain features of the world.

B The subject is normally correct or competent in making its assertions (Davidson 1980a: 256).

Together these assumptions effectively form a principle of charity which Davidson calls the principle of coherence. This principle, "...prompts the interpreter to find a degree of consistency in the thought of the speaker" (Davidson

1991a: 158). Consistency, as it is used here, refers not only to the unique internal pattern of each 'language' (e.g. its compositional rules, axioms, etc.), but also to some standard logical rules which are in place *a priori* (e.g. the law of non-contradiction). In light of this, we can see that both holistic and logical constraints conspire to compose the principle of coherence.

This principle is necessary because if a subject's use of words is very inconsistent then no attribution of meaningful speech or thought can be made. If its reactions are so far off the mark that it utters sounds indiscriminately or incompetently then no reasonable pattern will be found by which to warrant the ascription of *any* propositional content. For a simple example let us reconsider the vervets. If it turned out that a given monkey was consistently wrong about the presence or absence of, say, 'snakes' then even in principle a radical interpreter (or behavioural ecologist) would be unable to discover any meaningful pattern to its 'calls'. Lack of such a pattern might simply indicate that the monkey was unable to master the use of the call signal. If so we would expect that it would gain a reputation as an unreliable sentry. Following a sojourn with behavioural ecologists, Dennett chronicles exactly this type of case. He writes, "...a particular vervet can indeed lose credibility with the group on a particular topic, thanks to being 'framed' by experimenters" (Dennett 1987: 270, cf. Davidson 1986a: 315). Too much error on the part of a putative speaker however results, not just in a fall from grace in the eyes of one's peers, it results in not saying anything at all. Attempts at interpretation in such cases would of course break down and the claim that one was dealing with a thinking agent would have to be surrendered. Put crudely, if the speaker or thinker doesn't know what they're talking about, neither will the interpreter. This is why the very possibility of interpretation rests on making the charitable assumption, "...that we can dismiss *a priori* the chance of massive error" (Davidson 1984a: 169).

For another example, consider the case of Mrs. T. Mrs. T is an elderly patient suffering from senile dementia and severe memory loss. When asked "Was McKinley assassinated?" Mrs. T always answers in the affirmative. But, due to her memory loss, if she is asked "Who is McKinley?" or "What does 'assassinated' mean?" she cannot answer coherently. It is clear that Mrs. T is responding to the question (in some sense of 'respond'), but it would be unprincipled to say that she believed the proposition she apparently assents to. Mrs. T's senile dementia reduces her to the point that we can find no sense in her talk about McKinley. She fails to fit our norms of logical consistency and her speech acts lack an internal coherence. In other words, Mrs. T fails to meet the minimal criterion for being a rational agent. If this is widespread enough we would be forced to treat her from a different perspective. This is what Davidson means by suggesting that we, "...make others intelligible to us only to the extent that we can recognise something like our own reasoning powers at work"

(Davidson 1987b: 47). Either we find consistency in linguistic use or we must forego attribution of propositional content altogether. If we have too many doubts about the speaker's coherency (as we might if radically different conceptual schemes were possible), we must conclude that the subject's utterances have no meaning and therefore involve no conceptual beliefs. In the light of such cases, we can see why the principle of coherence is necessary if interpretation is to be possible (Davidson 1984a: 152, 159, 197, 1986b: 323, 1986a: 317, 319).

At this juncture critics often point out that there seems to be a conflict between the way in which radical interpreters are expected to employ the principle of coherence when developing the axioms of a T-theory and the way in which they are later expected to test their T-sentences. For, according to Davidson, the interpreter can only pose and test hypotheses about what a subject's sentences mean by attending to the pattern of actual sentential usage. However if the principle of coherence is employed too strictly none of the empirical hypotheses about the subject's speech acts could ever be falsified. Clearly this is problematic since no one is ever perfectly consistent or logical in their speech or thought. The odds are that there will never be a perfect or ideal correlation between past and current usage.

An example may help. Let us suppose, for the sake of argument, that we have in hand a fairly detailed T-theory for a subject. Let us also suppose that on one occasion the subject assents to P in the presence of conditions Q. However our T-theory predicts that conditions Q should not elicit the utterance of P but the utterance of R. In such a case, we have encountered a glitch in our T-theory predictions. How are we to explain this if we are prompted to locate a coherent pattern in the thought of a speaker? Given that our only evidence for T-theories is the observable behaviour of the subject such, "...evidence will never tell us which T-sentence is the source of trouble. Once an anomaly appears, it can be pushed around in the theory like a sealed air bubble" (Ramberg 1989: 86). How could an interpreter cope with such an anomaly? How could they accommodate the erroneous prediction without scrapping their entire T-theory while, at the same time, maintaining the principle of coherence? Would such an anomaly, on its own, cause them to abandon the thought that they were dealing with a believer of propositions? Not necessarily. There are several options open to the interpreter at this juncture – most of which shed light on the fact that, in the interpretation of speech and thought, we *must* rely on another crucial normative principle of charity.

First of all, the interpreter could try to accommodate the anomaly by developing a new T-theory. This move is perhaps most consistent with the principle of coherence. However if the original T-theory was developed over a long period of time and the interpreter was diligent in their original efforts there would be an immense number of logical connections holding between a speaker's past

utterances. It would be unlikely, to say the least, that a wholly new T-theory could be generated which would account for all the old assertions as well as making sense of this rogue assertion (cf. Davidson 1984a: 169). Therefore, in most cases, this strategy is unlikely to work.

A second move would be for the interpreter to assume that there was a mistake on his or her part. Perhaps all is well with the T-theory and the speaker's use but the interpreter has misread the conditions of assent. This too, by giving the subject the benefit of the doubt, favours the principle of coherence.

Finally one could invoke the further charitable assumption which Davidson calls the principle of correspondence. This principle, "...prompts the interpreter to take the speaker to be responding to the same features of the world that he (the interpreter) would be responding to under similar circumstances" (Davidson 1991a: 158, cf. Malpas 1992: 47). With this in hand, the interpreter might reasonably assume that the speaker has simply made a mistake. He or she might suspect, for some salient reason, that the speaker did not notice that conditions Q obtained; for, "...it cannot be assumed that speakers never have false beliefs. Error is what gives belief its point" (Davidson 1984a: 168). However if we should decide the subject has made a mistake, then, in accord with the principle of coherence, we must realise that, "The best we can do is cope with error holistically, given his actions, his utterances and his place in the world. About some things we will find him wrong, at the necessary cost of finding him elsewhere right" (Davidson, 1986a, p. 318). Davidson goes on to say, "As a rough approximation, finding him right means identifying the causes with the objects of his beliefs, *giving special weight to the simplest cases*, and countenancing error where it is best explained" (Davidson 1986a: 318, emphasis mine, see also Evnine 1991: 109).

Adopting this line does not contradict the principle of coherence since that principle does not ask us to regard subjects as infallible. For this reason, we must bear in mind that 'rationality' and 'normativity', as they are used here, are not to be construed as sharply defined concepts (Davidson 1990b: 325). Nonetheless if we wish to count a subject as making a mistake we must first have some evidence, in the form of largely correct utterances, by which to judge the nature of their mistake. Errors must arise against the background of a largely coherent and successful pattern of correct utterances. We can only sensibly assign irrational beliefs to a subject against such a pattern (Davidson 1984a: 153, 197). This in turn demands that we can only account for error if we assume that the speaker has largely correct views about the world (cf. Davidson 1980a: 221).

4.2. The Further Need for the Principle of Correspondence

It may still be wondered how it is possible for the radical interpreter to overcome the problem of logical truth? Here again we can employ the principle of correspondence to good effect. Quite simply, according to that principle unless there is an independent reason for thinking that the subject would append an extraneous logical or mathematical truth to *all* their thoughts, we should initially prefer theories that interpret them as making assertions about things that we would likely be speaking about if we were in their shoes. This is the default position in any case. I have already highlighted parallels between this principle and the basic assumptions of simulation theory in chapter four, section 3.2 (cf. also Hutto 1994, Davidson 1991a: 159). Moreover appeal to this principle is not an *ad hoc* means of dissolving the problem of logical truth, as it is the only means by which we can attribute some content to a subject's beliefs in the first stages of radical interpretation. Unless we make the assumption that the subject is responding as we would to the same features of the world, the project of radical interpretation would never even get off the ground. In this respect the principle of correspondence solves two crucial problems which face the interpreter in the early stages of interpretation.

The first problem is due to the holistic nature of propositional content. We cannot reasonably assign a single propositional attitude to a subject without assigning a host of others. Bearing this in mind, it seems that radical interpretation is a straightforward impossibility since the basic data upon which we erect our T-theory are *single* sentences assented to in relation to an environment. The trouble is that interpreters need to make some initial assignments prior to having a T-theory in hand. But unless some initial propositional contents can be assigned in isolation of all others a T-theory cannot be developed at all. At the same time, initial assignments cannot be made prior to having a T-theory because speech and thought are holistic. What can be done?

The problem can be resolved if we rely on a defeasible principle of correspondence in the first stages of interpretation (cf. Malpas 1992: 146, 170). In doing so, we can take reasonable first guesses about which features of the world the subject's remarks are directed at. The idea is that, in the simplest cases, interpreters come to know that the utterance of a particular word is provoked by the presence of a particular type of worldly item. They assume this by watching how the subjects interact with each other and their environment and by noting what causes them to respond as they do (Davidson 1990b: 318 1986a: 315-317). The interpreter begins by assuming that the subjects see things as the interpreter does, unless evidence suggests otherwise. As we have seen something similar goes on in the primitive learning situation. From this we can see that the first stages of interpretation depend upon our ability to recognise, "...some simple

attitude that an interpreter can recognize in an agent" (Davidson 1990b: 322). Of these simple attitudes Davidson writes:

> The assumption that such attitudes can be detected does not beg the question of how we endow the attitudes with content, since a relation, such as holding true, between a speaker and an utterance is an extensional relation which can be known to hold without knowing what the sentence means. I call such attitudes non-individuative, for *although they are psychological in nature, they do not bestow individual propositional contents on the attitudes* (Davidson 1991a: 158, emphasis mine).

Compare this with Gómez's suggestion about how to understand primitive cases of shared attention. He writes:

> The perception of X attending to O is almost the prototype or skeleton of a propositional attitude, except that there is no proposition (just a physical agent) and the attitude may, literally, be a position of the body or a direction of the eyes, accompanied or not by facial and vocal expressions. Perhaps the utterances produced by our ancestors (e.g. 'there food') started to be substituted for the actual contemplation of an object by the audience in their [shared attention episodes]...i.e. the visual perception that 'X SEES FOOD THERE' could be replaced by the visual/auditory perception 'X SAYS *FOOD THERE*'. The structure is almost the same but *FOOD THERE* is no longer a first-order, perceptual representation of food lying in a particular place, but the representation of a symbol or a string of symbols produced by another organism (Gómez 1998: 89-90).

It is worth digressing at this point to clarify the role causality plays in the process of interpreting speech acts. Some philosophers may be concerned that an invocation of causality at this juncture is at odds with the overall non-reductive nature of Davidson's interpretative account. For instance, some may think that making free with 'causal talk' revives the problem which was identified as the bane of the causal theories of content in chapter two, section 3.1 (Evnine 1991: 150). For it is surely possible to have the belief that 'There is a hippo in the fridge' because of too much to drink, without 'over drinking' in any way, figuring in the content of this belief. As this example suggests, there can be a gap between what causes my thought and what my thought is about. However the worry that a Davidsonian approach to interpretation will go the way of crude causal theories is unwarranted. It is certainly true that in his talk of causation and correspondence, Davidson wants to allow for a possible gap between the cause and the meaning of a speaker's utterances. Nevertheless he holds that it is unintelligible to view this as

a gap between a lone subject and the world. Put differently, we can make sense of this potential gap during interpretation only by triangulating with another. This inherently and non-reductively locates what causes us to use our words as we do in the social context of interpretation. Hence, although he speaks of causality and correspondence, it is vital to notice that the role these concepts play in his work is nothing like that which they asked to play in traditional correspondence theories of truth and reference which he regards as 'absurd' (Davidson 1986b: 307).[63]

I now come to the final problem which would face the interpreter in the early stages of radical interpretation. This problem also concerns holism, but holism of another sort. Propositional thought is not only holistic in that it forges logical connections with the internal structure of other thoughts, it is also holistic in that there is a hermeneutic circle holding between what a speaker means and what they believe (cf. Malpas 1992: 41, Davidson 1984a: 167, 1986a: 315, 1990b: 309). Hence if Alex says, "There is a hippo in the fridge", these words might mean that, "There is an orange in the fridge" or they might mean what they usually mean in English but for some strange reason he may honestly believe that there is a hippo in the fridge (cf. Davidson 1984a: 100-101). Other possibilities present themselves. He might be telling a bad joke, deliberately speaking in code or he might just be mouthing the words for some odd reason. Nevertheless there is no means by which to interpret Alex by simply noting the isolated speech act or even his pointing to the refrigerator. Given that this would be the sum total of the evidence available to radical interpreters in the first stages of interpretation their project seems hopeless. Put crudely, it won't help to simply ask their subject what he believes unless the interpreter is confident that they already understand his words, since we cannot ascribe beliefs without ascribing propositional contents to them. But given that only subjects capable of making linguistic utterances are sophisticated enough to harbour such contents, it follows that in order to make propositional ascriptions we must interpret the meaning of subjects' utterances. However, as Davidson rightly points out, the only way to interpret another's utterances is by having a grasp of their intentions. Thus by a series of sharp turns we are brought full circle because such intentions must be understood in terms of beliefs and desires. Consequently the ascription of propositional attitudes and the interpretation of meaningful utterances are part of a single project.

Radical interpreters must break the holistic circle obtaining between belief and meaning without recourse to either a complete theory of meaning or a detailed theory of belief for the subject (both of which they are simultaneously trying to develop). Davidson is the first to emphasise that we cannot give primacy to either belief or meaning in interpretation. As he *often* says, "We do not know what someone means unless we know what he believes; We do not know what someone believes unless we know what he means. In radical interpretation we are

able to break into this circle" (Davidson 1984a: 27, 101, 127, 134, 141, 142, 144, 146, 153, 156, 186, 1986a: 314).

Again it is only by appeal to the principle of correspondence that the radical interpreter is able to overcome this difficulty. For, "[Radical interpretation]...is intended to solve the problem of the interdependence of belief and meaning by....assigning truth conditions for alien sentences that make native speakers right when plausible possibly, according, of course, to our own view of what is right" (Davidson 1984a: 137). As before, what Davidson means by our own view of what is right is not just logical consistency. Roughly, radical interpreters, must initially assume a certain level of agreement about how they and their subjects view the world if interpretation is to be a possibility. In the case cited above, we start by interpreting Alex's words, "There is a hippo in the fridge" not as referring to hippos but to something more likely to be in the fridge. Unless, that is, in the course of further investigation we find that attributing Alex a belief about hippos makes better sense of his overall pattern of response or we find the pattern is insufficient to warrant any ascription at all.[64] Nothing in the interpretationalist account papers over the fact that when it comes to interpreting sophisticated thought, there will be many false starts and wrong turnings. Interpreting others is difficult – precisely because some thinkers believe outrageous and obscure things. But if propositional content is there to be probed we must expect that such difficulties will iron themselves out (cf. Malpas 1992: 44, 146).[65]

4.3. Irreducible Norms

As we have now seen, Davidson views propositional ascription as an essentially normative practice. Consequently he maintains that reason-explanations are irreducible to physical explanations. He writes:

> The reason mental concepts cannot be reduced to physical concepts is the *normative* character of mental concepts (Davidson 1987b: 46)

The idea is this: We cannot assign length without a physical framework. Similarly we cannot ascribe propositional attitudes without a normative, interpretative framework. Each of these domains has its own constitutive features. As long as we conceive of people 'as rational' we cannot have a closed system for propositional ascriptions such that it could be suitably reduced to a system of mere physical descriptions. This is not because it is the nature of the mental to preclude translation, it is because such ascriptions are bound up with interpretative practice. When ascribing propositional attitudes we must be prepared to cope with new evidence which needs to be coherently and normatively incorporated.

For example, imagine that a friend of yours develops the habit of continually visiting a particular cinema in town despite his general distaste for films. How can

we explain his sudden change in behaviour? We ask him and he tells us that he has been going there because it alone shows French films – which, he insists, he regards as exceptional. However in the course of time what would we make of this explanation if the cinema stopped showing French films but our friend kept going. In order to 'make sense' or 'rationalise' his behaviour we would need to revise our original assessment of his answer. Perhaps he was deceiving himself or us. Perhaps he is attracted to the girl at the ticket counter and not the French films at all. Perhaps he has changed his mind about films in general. Something like this is needed in order to make better sense of the pattern of his actions in light of the facts. This case is oversimplified but it provides insight into the fact that reason-explanations and propositional attitude ascriptions must be open to interpretative reappraisal. This is the key reason why propositional attitude psychology cannot be turned into a law-like science.

We cannot isolate, once and for all, the features which explain why our friend chose to visit the cinema in the way that we can, by appeal to some general law, explain why a particular car moving at speed X and at trajectory Y knocked over a post box. Many philosophers hold that when we speak of causes in this way we are in fact tacitly supporting the view that it is possible (at least in principle) to state a law which shows in ideal conditions how the antecedent event brings about the subsequent effect (cf. MacDonald 1989: 186-187). This view has, "...a long pedigree" (Sorabji 1980: 27). Nonetheless it is because the ascription of beliefs and desires must always take into account an agent's project, perspective and values in conjunction with events in his non-psychological surround it is a hermeneutically complicated and irreducibly revisionary affair. Making such ascriptions is, for this reason, irredeemably dis-analogous to the way in which we understand the behaviour of 'mindless' entities (Davidson 1991a: 162-163).

In this light, it is somewhat misleading when Davidson says that commonsense psychology is a, "...familiar mode of explanation of human behaviour and must be considered an organised department of common sense which may as well be called a theory" (Davidson 1984a: 158). For, in contrast to the theory-theorists, he maintains that propositional attitude psychology has special constitutive features that make it unique among 'theories'. Nevertheless his attempt to show that an irreducibly normative account of propositional attitude ascription is not odds with endorsing causalism about the attitudes is a concession to supporters of theory-theory. To square this circle he advocates a sophisticated form of non-reductive physicalism. In the next chapter, I argue that his version of anomalous monism is unfit for this work.

5. Conclusion

We have now seen how an interpretationalist account of propositional attitude psychology and a modest of biosemantic theory of intentionality are non-reductively related. It has also been demonstrated that interpretative practice is essentially and irreducibly normative. It is for this reason that an exclusively scientific treatment of propositional attitude psychology is not a live option. However, in the next chapter, I critically assess Davidson's compatibilist solution to the problem of fitting this normative account of interpretation into an orthodox physicalism. I conclude, by suggesting that the standard physicalist readings of Davidson's proposed solution are untenable. Consequently I claim that an interpretationalist understanding of propositional attitude psychology is not compatible with an exclusively science-based metaphysics.

CHAPTER SIX

Davidson's Identity Crisis

"During the heyday of neo-Wittgensteinian and Rylean philosophy of mind the era of little red books, it was said that propositional attitude explanations are not causal explanations and that beliefs, intendings, imaginings and the like are not even candidates to be causes....We have come a long way since then. The work of Davidson, Armstrong, Putnam and Fodor (amongst others) has reversed what was once orthodoxy and it is now widely agreed that propositional attitude attributions describe states and episodes which enter into causal relations."

– Lepore and Loewer, *More on Making Mind Matter*, 1989: 175

1. Introduction

If propositional attitude psychology must be understood by normative interpretative lights then it cannot be reduced or naturalised. Nonetheless some interpretationalists hope to make peace with naturalism by subscribing to a form of non-reductive physicalism. This is standardly motivated by a desire to endorse a causalist view of reasons and their mode of explanation. For reductionists and non-reductionists alike, as Lepore and Loewer suggest in the opening quotation, causalism is not only dominant in today's analytic philosophy of mind, it has the status of a new orthodoxy. This is unfortunate because, minimally, it obscures what is most important in the study of reason-explanations and, maximally, in its most common Humean forms, it is unintelligible in this domain. The concern of this chapter is to establish this and to motivate the project of finding a replacement understanding of the idea that reasons are causes. My strategy is as follows: I critically investigate Davidson's sophisticated non-reductive version of the thesis. This analysis is useful because it reminds the reader that treating reason-explanations as a kind of causal explanation does no explanatory work in answering the key methodological question: *How* do we determine for which reasons we act? It is because Davidson explicitly recognises this that his account is

both decisively distinct from and, *prima facie*, superior to reductive versions of causalism.

But if claiming that reason-explanations are a species of causal explanation was never meant to solve the 'primary reason problem' then we must ask: What really motivates Davidson to make this claim? This was the felt need to explain, what other non-reductive accounts, apparently generated: the problem of the mysterious connection between reasons and actions. But, given his own non-reductionism, he needed to develop the metaphysics of anomalous monism in order to make his solution to this latter problem intelligible. Many claim, however, that anomalous monism is unworkable because it ultimately results in epiphenomenalism. Davidson has replied to this critique with what I label 'the extension reply'. But, part of the purpose of this chapter is an attempt to reveal that if anomalous monism is to be a stable doctrine then the extension reply must be read weakly. Unfortunately, the price of this way of achieving stability is that Davidson's compatiblist metaphysics is unfit for the task of solving the mysterious connection problem. In a nutshell, endorsement of the weak reading leaves us unable to justify the principle of the nomological character of causality (hereafter the PNCC). Unless independent support for this principle can be found there is no argument for anomalous monism. More damning, without support for such a principle there is no sound basis to suppose that causal relations can, in principle, always be fully explained by appeal to the laws of an ideal physics. It appears that only a more straightforward reductionist metaphysics could provide such a basis. But, in light of the analysis of chapter two, reductive accounts are incapable of dealing with the special, normative character of reason-explanations. On these grounds I argue that we ought to reverse the orthodoxy once again and eschew standard accounts of causalism about reasons and reason-explanations. I conclude by suggesting a possible way of recasting our understanding of causation in this domain so that the mysterious connection problem never arises.

2. What Motivates Causalism?

To some philosophers it is just self-evident that our everyday discourse commits us to the claim that psychological phenomena cause, and are caused by physical events (e.g. events in our bodies and environment). That is to say, in ordinary language we commonly speak of reasons, beliefs, feelings, anxieties, pains, and a host of other psychological phenomena as if they were appropriate items to mention in the explanation of the production of 'rough grained' bodily actions. Likewise, we often explain the occurrence of particular thoughts, feelings and pains by appeal to the apparent fact that they have been caused by non-psychological events, both inside and outside our bodies (Child 1994: 109). This kind of talk is the ground for the view that, in general, a large range of

psychological phenomena, and reasons, in particular, are both causally efficacious and warrant causal forms of explanation.

2.1. The Neo-Wittgensteinian Orthodoxy

Despite the claims about the self-evident status of causalism, there was a period in recent history when its allegedly commonplace insight was far from universally accepted – at least in philosophical circles. For example, consistent with his rejection of the 'Intellectualist Legend', Ryle cautioned, "...we should not swallow the traditional stories about the occult antecedents of either deliberate or impulsive actions" (Ryle 1949: 115). Likewise, the target of Melden's *Free Action* was the, "...apparently innocuous suggestion that free action is to be distinguished from action that is not free, not by the absence of causal conditions, but by the presence of certain specific sorts of mental causes" (Melden 1961: ix).

As the opening quotation suggests, Davidson has been a central figure in reversing the non-causalist orthodoxy concerning reasons and their mode of explanation. The main thrust of his seminal article, "Action, Reasons and Causes", was to convince us that (i) a merely teleological analysis is inadequate to deal with the mysterious connection problem, and that (ii) it is at least possible that reasons could be causes. Furthermore, despite his sympathy for some aspects of the teleological approach, he endeavoured to supply sturdy support for the conviction that (iii) reason-explanations must be understood as a species of causal explanation. But Davidson's position is sophisticated. Moreover, because it highlights features that put him at odds with other, more reductive, causalists it must be handled with care.

2.2. The Primary Reason Problem

According to Davidson, when we give rationalisations of a person's behaviour we must do so in terms of beliefs and desires – the conjunction of which he calls the *primary reason* for acting. He tells us, "A primary reason consists of a belief and an attitude, but it is generally otiose to mention both" (Davidson 1980a: 6). But citing *a* reason (a primary reason) is not the same as citing *the* reason for an action. Usually a person will have more than one primary reason for acting at any given time. Yet, in many circumstances, only one reason is the reason for the action. There are a wealth of alternative belief/desire pairs which could make sense of any given action but citing one or another as the cause is allegedly necessary if we are to explain why a person acted (Davidson 1987b: 42). Consider the following scenario: I have promised my wife that I will wash the dishes before five o'clock and being a man of my word I intend to do so. According to Davidson this would be a primary reason, a rationale, for my doing the washing up before the

appointed time. However, it may be the case that, despite my intention to keep this promise, the reason I do the washing up before five o'clock is *because*, unexpectedly, I get a phone call from a friend who has decided to come round for tea at four. Since the tea cups are at the bottom of the pile, and I don't want my friend to see the kitchen in a state, I choose to do the washing up before he arrives. The latter reason (my desire to get some clean cups and to keep a tidy house for my visitor) pre-empts the former reason (my desire to be true to my word) in the production, and hence explanation, of this action. In other words, citing the latter reason tells us *why* in this instance, I actually did the washing up.

In light of such cases, Davidson argues that we must read the *because* in such common remarks as, "I did the washing up *because* I wanted to impress my visitor" as indicating a causal relation (Davidson 1980a: 254). What separates *a* reason for acting from *the* reason for acting is that when we give *the* reason we are attempting to designate the mental event(s) which, as far as we can tell, actually 'produced', 'brought about', 'were responsible for' or 'caused' the action in question (Davidson 1987b: 38). This is how the 'champion of the causal analysis' intends to show that "when we offer the fact of the desire and belief in explanation, we imply not only that the agent had the desire and belief, but that they were efficacious in producing the action" (Davidson 1980a: 232). He concludes that when we cite the reason for acting we are citing its *cause* as well.

Using Melden as a foil, Davidson sets out to show that to accept the non-causalist position is to be, "...without an analysis of the 'because' in 'He did it because...' where we go on to name a reason" (Davidson 1980a: 11). To illustrate the problem he uses the example of a man who signals by raising his arm when approaching a turn. If we treat these events, the signalling and the turning, as causally unrelated then, although we can make sense of the arm-raising in terms of social context (it being the appropriate kind of action to take by sensible road users), we have yet to explain why the man took this action in this instance. All that could be said in this case is that, "He had a reason to raise his arm, but this has not been shown to be the reason why he did it" (Davidson 1980a: 11). Thus the objection to Melden's neo-Wittgensteinian account of reason-explanations is that it fails to recognise the need to connect the reason to the action. It is mainly against this background that Davidson, contrary to the Wittgensteinian tradition of the time, insisted in his early essays that reason-explanations are a species of causal explanation and that, "A reason is a rational cause" (Davidson 1980a: 233). He claimed that:

> Much of the explanatory force of reason-explanations comes from the fact that they specify *which* pair, from among the vast number of belief-desire pairs that were suited to cause the action, actually did cause it (Davidson 1987b: 42).

2.2.1. Why Causalism doesn't Solve the Primary Reason Problem

Davidson's above claim is treacherously stated. It may appear to say that by giving attention to the causal character of reason-explanations we will be able to designate which belief/desire pair caused the action, thereby furnishing much explanatory force. But careful consideration of Davidson's wider account reveals that this cannot be right. Consider how we actually designate *the* reason for acting. We standardly ask the person and accept their answer. And unless we have good reason to think otherwise, we tend to treat others as authorities in these matters. It is with this in mind that Johnston writes:

> The individual is treated not as a site of causal processes but as an agent, and *qua* agent he is attributed a privileged role, for by citing a consideration he authoritatively locates his action in the wider network of aims and intentions which give it its sense....To return to the example Davidson considers, if we wished to find out why the man raised his arm the appropriate way to do so would be to ask him. His statement would demonstrate that not only did he have a reason to raise his arm but that was his reason for so doing. *Here Davidson would accept that the agent has a privileged role in stating his reasons for action...* (Johnston 1989: 45, emphasis mine).

We might wonder: What gives the agent the authority of saying which one out of a host of possible reasons actually causes the action? How does the agent come by this knowledge? One suggestion is that the agent has epistemically transparent access to this special class of internal causal events (i.e. reasons). To make this plausible it would be necessary to clarify the nature of this special relation between persons and their internal events in such a way that explains this privileged access.[63]

Apart from the crude metaphysics inspired by such a picture, it also suffers from the fact that it gives too much authority to the agent.[64] Occasionally self-interpreters can be badly mistaken about the reasons for their actions. The general pattern of a person's behaviour might fail to fit with their accounts of their actions. Yet this need not result in a collapse of understanding; merely a challenge to the agent's honesty (cf. Davidson 1980a: 18). Our reasons are not always obvious or transparent to ourselves or others. In developing a careful explication of my action, I need to review my reasons. Perhaps there are many and perhaps, in the end, no single or simple explanation will do. Most reason-explanations contrast with the relatively transparent situation of the 'washing up' scenario. For instance, if asked why I am typing this page right now I might answer: I am typing it to finish this chapter before next semester, or out of a sense of duty, or in the hope of getting my thoughts clear. Perhaps I am typing it for *all* of these reasons at once. I may favour one, or all, of these explanations, yet it would still be possible

that someone may offer an interpretation that will convince me that I am, in fact, mistaken or (worse) self-deceived. If so, they will furnish a more plausible or powerful account of the pattern of my behaviour. So, while recognising a certain degree of first-person authority, we must also recognise that there is always the potential for an epistemic gap between the actions of agents and the accounts they give of them.

Given this, we might ask: Would a third-personal causalist account properly characterise how we come to know about our reasons for acting? If reason-explanations are treated as straightforward causal explanations, as proposed by reductionist accounts, then saying why we act would be no different from making hypothetical judgements about causes of any other kind. But this cannot be the right story either. A wholly third personal response is not a live option because it gives too little authority to the agent.

Thus the question remains: *How* do we determine which reason, or reasons, best explains an action? When making our reply we ought to consider Wittgenstein's query:

> What is the difference between cause and motive? – How is the motive discovered, and how the cause? (Wittgenstein 1953: 224).

Davidson is mindful of this question. For him, *the* reason for acting must be designated, to borrow Hornsby's idiom, from within the 'personal point of view' (cf. also Lennon 1992: 236). We must somehow decide that one belief/desire pair has explanatory priority over others and then designate it as the cause – not the other way around. But if this is the correct explanatory order then treating the 'because' as indicating a causal relation doesn't perform the job of helping us adjudicate between competing hypotheses in the way 'looking for the cause' of say a plane crash would. In the latter case we begin by attempting to reconstruct the events surrounding the crash (as they relate to the particular case). We subsequently employ our general knowledge of what tends to happen in like cases so as to isolate the cause. But, I say again, this is not the method used when we give reasons for our actions. No appeal to inductive regularity is involved. In itself this is not damning to causalism. For as Davidson notes, we do not always have to appeal to induction in order to know that a causal relation exists or to accept that there are causal laws at work. He writes:

> ... induction is a good way to learn the truth of a law. It does not follow that it is the only way to learn the truth of a law. In any case, in order to know that a singular causal statement is true, it is not necessary to know the truth of a law; it is necessary only to know that some law covering the events at hand exists. And it is far from evident that induction, and induction alone, yields knowledge that a causal law satisfying a certain condition exists. Or to put it

differently, one case is often enough, as Hume admitted, to persuade us that a law exists, and this amounts to saying that we are persuaded without direct inductive evidence that a causal relation exists (Davidson 1980a: 18).

Hence we might think that, "...singular causal relations are....given in experience" (McDowell 1985: 398). Applied to the psychological domain, this would mean that we could, as it were, identify which reason causes an action at a glance. The main problem with this proposal is not that, "...it implies there can be 'brute' instances of causation" (C. Macdonald 1989: 186). The problem is it fails to provide an adequate criterion for identifying the cause of action. Unlike other causal investigations, in the case of reasons we begin and end by making a true singular causal statement only *after* having decided why it is that we acted. On this basis, we simply assume that there is a law available under some other description which explains the causal relation.

Given that we can be wrong about what causes us to act we need to ask: How do we identify *the* reason for acting in real cases? How do we choose between rival hypotheses about the reasons for our actions? Those attracted to the idea that we just see singular causal relations provide no answers to these questions and no framework in which to ask them. Appeal to causality in this way does nothing to eliminate or discredit any of the alternative rationalisations of our action. This is further highlighted when we consider that reason-explanations will never be reduced to pure causal explanations due to their normative character.[65] As long as we concern ourselves with the perspectives of agents our practice of citing reasons to explain actions propositional attitude psychology will not develop into a nomological science (cf. Lennon 1992, Tanney 1995: 118, Johnston 1989: 45–47). When we explain ourselves by appeal to reasons we do not have, and cannot hope for, a closed system for the assignment of psychological episodes such that it could be suitably reduced to a system of psychological laws. To get such laws we would have to completely abandon (or alter) the enterprise of commonsense psychology. If this is correct, then it should not surprise us that we must be content to talk of reasons *causing* a subject's behaviour and actions, as we often do, without being able to identify such causal relations by appeal to the standard regularity criterion, without our knowing any psychophysical laws, and even without the prospect of such laws ever being developed (Davidson 1980a: 233). Laws are the province of *some* of the natural sciences - they are not part of the understanding of one another we gain by interpreting each other as agents (cf. Hutto 1997).[66]

The point is that since Davidson steadfastly denies that *the* reason (which is also the cause) is discernible from the naturalistic perspective, he robs it of the explanatory role of eliminating competing hypotheses which is crucial in all other genuine causal investigations. As long as the reason for acting is designated with deference to the agent, as opposed to giving an impersonal analysis, then labelling

the reason as the cause does no discriminatory labour. Consequently, on this score, Davidson's account is not an advance on Melden's since both are rightly unwilling to embrace a more thorough-going naturalism. On the issue of how the reason for action is initially identified, Davidson shows himself to be in substantial agreement with the neo-Wittgensteinians.[67]

Thus simply treating the 'because' in reason-attributions as indicating a causal relation does not, in itself, provide a means of solving the primary reason problem. What is worse the causalist story tends to direct attention away from what is truly important in the study of reason-explanations because it doesn't wear this fact on its sleeve. To deal properly with the primary reason problem we might consider Julia Tanney's proposal. She proposes that, "The 'something that is missing' from the primary reason account is, if anything, a more complex interpretation that would include competing reasons, values, judgements manifesting choices, and so on, and no doubt this interpretation will be dependent on what the agent is disposed to say about his own actions" (Tanney 1995: 114). This is not the place to explore or develop this idea, but I find it congenial to the kind of narrative proposal I have begun to develop elsewhere (cf. Hutto 1997).

3. The Mysterious Connection Problem and Anomalous Monism

We have now repelled a possible first attack against the non-causalists and established that appeal to the causal character of reason-explanations gives no explanatory edge in solving the 'primary reason problem'. But, for Davidson, this was never the function of such an appeal. As Child notes, "The point of the argument for a causal view is not that we must appeal to causation in order to *tell* whether S Φ–d for one reason rather than another; it is, rather, that we must appeal to causation in order to understand *the metaphysics of the relation* between reason and action" (Child 1994: 96, second emphasis mine). The real motivation for regarding reason-explanations as a species of causal explanation is that it helps to solve the mysterious connection problem by placing reasons into a familiar metaphysical and explanatory framework. Hence having decided which belief/desire pair describes *the* reason for acting we simply claim it is connected to the subsequent action by a causal relation. Yet, on its own, appeal to an unreduced causal relation of this kind doesn't solve any mysteries. Things only improve if we regard reasons *as* causes and if we can successfully incorporate this view into a wider metaphysical account. Such an account must reveal that reason-explanations are, in at least one important sense, the same as other causal explanations. At the same time, this claim must be advanced without threatening the autonomy of reason-explanations or ignoring any of their special features.

3.1. The Token Identity Theory

Davidson offers just such an account in the form of a token identity theory which enables us to hold that one and the same event can have both mental and physical descriptions.[68] The idea is that a mental description picks out a particular event which can, in theory, be redescribed in the language of an ideal physics. To make this account viable Davidson needs to convince us of three things about events:

(1) We regularly refer to events in ordinary discourse.

(2) Our criterion of event identity is no worse than our criterion of identity for other ontologically basic entities such as objects.

(3) It is possible to re-describe events in just the way we re-describe objects.

I will digress to briefly outline Davidson's reasons for accepting claims (1) to (3). Firstly, in support of (1), his analysis of ordinary discourse suggests that we quantify over singular terms which refer to events all the time. Although this may not be immediately obvious, it is revealed by an analysis of the logical form of our action sentences (Davidson 1987b: 36). For example, according to Davidson, (a) is made more perspicuous by analysing it as (b):

(a) Smith kicked Jones

(b) $(\exists x)$ (kicked, (Smith, Jones, x)).

In this light, it becomes clear that we can, and regularly do, refer to events just as we refer to objects.

Secondly, in support of claim (2), given that we refer to events in ordinary language they ought to be regarded as ontologically basic kinds on a par with objects (Davidson 1985a: 241). Moreover Davidson claims that our method for individuating events is no more, or less, principled than that which we employ for individuating objects. This is important, for as he says, "...unless we can make sense of assertions and denials of identity we cannot claim to have made sense of the idea that events are particulars" (Davidson 1980a: 180). Originally, he proposed that we individuate events by appeal to their paradigmatically causal characters (Davidson 1980a: 179). But, as Quine remarks, this method, "...purports to individuate events by quantifying over events themselves" (Quine 1985: 166). Davidson conceded this during an interchange and in reply suggested that events could be identified as the same if they coincide spatio-temporally (Davidson 1985d: 175). This method of individuation is useful because we can still distinguish events from objects, even when they spatio-temporally coincide.

This is because events *occur* in space-time while objects *occupy* it (Davidson 1985d: 176). Still some complain that without a more elaborate underlying theory of events this criterion is shallow and apt to confuse (cf. C. Macdonald 1989: 122, 236, McLaughlin 1985: 344). In response Davidson reminds us that the same holds true of the way in which we individuate objects (cf. Davidson 1980a: 176, 180). And in his view, "Grammar allows no such confusion" (Davidson, 1985d: 176).

Finally, in support of claim (3) he has argued that not only do we refer to events in everyday parlance, but – because events are ontologically basic kinds – it is possible to re-describe them. Thus I can talk about 'what happened' at the Royal Bank of Scotland, Hemel Hempstead on Tuesday at 3:10 pm as being 'a flagrant violation of the law' or as 'a most ingenious robbery'. Both of these descriptions, with their different connotations, refer to one and the same event (cf. Davidson 1980a: 120). This is similar to the way in which I can re-describe objects. For example, I can speak of one and the same object as "The ostentatious thing that is blocking my driveway right now" or as "The neighbour's fantastic new sports coupé".

If we accept that events are ontologically basic kinds then it follows that they can have multiple descriptions. But it is not necessary that these descriptions will form part of a coherent, unified conceptual framework. With reference to Davidson's claim that particular events are individuated by their spatio-temporal occurrence, we might say an event which is designated as occurring as time (t) and place (p) can be described, for example, as 'the cause of my reasoned action' in one discourse and also as 'the cause of physical happening x' in the language of an ideal physics. How it is described matters to the conceptual connections it forges and illuminates. Noting this is especially important in the psychophysical case since these two ways of describing the same event will be forever separated due to their irreconcilably different constitutive frameworks.[69] To accept this means that, at best, non-reductive causalists can endorse a weak version of the broadly Humean view about the nature and explanation of causation. A subscriber to such a view accepts, "...that 'A caused B' entails that there exists a causal law instantiated by some *true* descriptions of A and B" (Davidson 1980a: 16, emphasis mine). Davidson's purpose in adopting this reading is to show that, "...it does not follow that we are able to dredge up a law if we know a singular causal statement to be true; all that follows is that there must be a covering law" (Davidson 1980a: 160). He believes the covering law we require can be provided under an entirely different description of the events in question. Events *simpliciter* enter into causal relations even if such relations are only truly (or fully) explained under certain descriptions.

Moreover, Davidson famously argues that acceptance of the token identity theory is the only way to consistently hold the following three orthodox principles:

(i) Principle of the Causal Interaction of the Mental: Some mental events cause, and are caused by, physical events.

(ii) Principle of the Nomological Character of Causality: Each true singular causal statement is backed by a strict law connecting events of kinds to which events mentioned as cause and effect belong.

(iii) Principle of the Anomalousness of the Mental: propositional attitude psychology is normative and cannot make appeal to homonomic laws (Davidson 1980a: 233).[70]

If we accept the token identity thesis then we can say that where there is a causal relation holding between the mental and the physical (principle i) then there must be a causal law (available under some description) which explains this relation (principle ii). But from this it does not follow that the laws which explain such relations are visible when we employ intentional idiom (principle iii).

3.1.1. Clarifying Davidson's Understanding of Supervenience

Before moving on, it is important to clarify Davidson's position on one point. He has, occasionally, tried to motivate acceptance of his token identity theory by appeal to the idea that the psychological domain is dependent on the physical. For example, he once wrote, "...although I am urging that, psychological characteristics cannot be reduced to the ....[physical, biological, physiological], nevertheless they may be (and I think they are) *strongly dependent* on them. Indeed, there is a sense in which the physical characteristics of an event determine the psychological characteristics" (Davidson 1980a: 253, emphasis mine). Furthermore he has appealed to the notion of supervenience to clarify the nature of this non-reductive relationship of dependency. Unfortunately there has been some confusion in the literature as to the exact nature of such relationships and what they entail (cf. also Kim 1978, 1979, 1982, 1984a, 1984b, 1990). This is because the concept admits of several interpretations, some stronger than others (cf. Charles 1992: 269).

Strong psychophysical supervenience has been characterised as the view that, "...no two possible events (objects, states) can agree in all physical characteristics and differ in some mental characteristic" (McLaughlin 1985: 360). This unrelativised version of supervenience is quite strong, since it implies that one could distinguish supervenient properties solely by examining the subvenient (i.e. physical) properties. And this would hold across all possible worlds and times. This implies the theoretically possibility of predicting the occurrence of a mental event knowing nothing other than what was going on at the physical level. Since

Davidson has explicitly ruled out the possibility of such predictions, strong supervenience is clearly too strong. If strong supervenience really were the proper way to understand the psychophysical relation, then given the materialist preference for purely scientific explanations, its advocates should press for what Johnston calls exhaustive monism. This is the view that, "...physical facts exhaust all the facts" (Johnston 1985: 417). Some philosophers are willing to accept and stand by this consequence (cf. Heil 1992, Kim 1987).

In contrast, some non-reductionists maintain that the psychophysical relationship should, at best, be understood in terms of weak supervenience (cf. Haugeland 1982, Seager 1988). Weak supervenience is standardly defined as the view that, "...there is no possible world in which two events (objects, states) in that world agree in all their physical characteristics and differ in some mental characteristic" (McLaughlin 1985: 360). This way of characterising weak supervenience makes the relation between the psychological and the physical looser and is therefore, comparatively, more promising. On this construal psychological properties can vary in their relations to physical properties across possible worlds.

This apparently makes better sense of the relationship which Davidson had in mind, since his paradigm of a non-reductive case of supervenience is that, "...of semantic predicates on syntactic predicates: a truth predicate for a language cannot distinguish any sentences not distinguishable in purely syntactic terms" (Davidson 1990c: 1, cf. Davidson 1985b, cf. also McLaughlin 1985: 365). Recall, from the discussion of chapter five, section 3.3, that although a theory of truth is not a theory of meaning it is a necessary condition on the possibility of meaning. If one accepts this then a clear sense is given to the idea that logical structure fixes certain facts about meaning but it doesn't determine meaning itself. Compare this to Ryle's observation that the rules of chess constrain a player's possible legal moves but they do not determine them (Ryle 1949: 76-77). As with chess, it is only because there is a set of stateable rules that moves are non-arbitrarily governed.

To continue with what is beginning to look like a 'three bears' saga, it is the absence of a clear statement of these constraints in the psychophysical domain that have led some to complain that weak supervenience is too weak. For example, it is, "...silent about across-world comparisons" (Heil 1992: 69). Unless qualified in important ways, weak supervenience permits the existence of physically (or metaphysically) possible worlds in which the physical facts are identical to ours, but which have no mental events at all. Unchecked talk of 'physically possible' worlds in the context of psychophysical supervenience claims is not useful (cf. Seager 1988: 705–707, Crane and Mellor 1990: 203–205).[71]

Nevertheless, although some worries about the plausibility of Davidson's anomalous monism have focused on how we ought to interpret his remarks about the supervenience relation, I believe this attention is misguided. As we have seen

the standard accounts of supervenience invoke the machinery of possible worlds and thus, effectively, require talk of relationships between *types* of event. Even though the talk is of logically possible worlds in which the same event occurs, it encourages us to imagine a world like this one but in which certain material facts are different. But Davidson is adamant that his account only concerns particular events which are individuated in terms of their spatio-temporal occurrence. This ties them uniquely to a series of other physical events both past and future (cf. Davidson 1993: 15). Hence, if we are speaking of events of *this* kind, it makes no sense at all to imagine the very *same* event could occur in another *physically* possible world. For to do so would be to have already imagined a different physical world and hence a different event (cf. Buekens 1997). This fits with Davidson's own attempt to make his views about supervenience clear. He writes, "The idea I had in mind is, I think, most economically expressed as follows: a predicate p is supervenient on a set of properties S if and only if p does not distinguish any entities that cannot be distinguished by S" (Davidson 1993: 4) Certainly, as he points out, this reading implies monism and is consistent with a token identity theory.

Davidson is credited with having been the first to introduce the notion of supervenience into the philosophy of mind. But in doing so his concern was merely to lend support to his token identity theory. Thus, in a reply to Harry Lewis, he wrote that:

> [when I introduced supervenience]...what I was arguing for....was *only* the identity of mental events with physical events. I wanted to emphasize that such ontological reduction does not imply that mental properties are physical properties, nor that there are causal or bridging laws relating events classed by mental properties with events classed by physical properties" (Davidson 1985b: 243–244, emphasis mine).

Given Davidson's special views on the individuation of events, to avoid confusion, his account ought only be thought of as a token identity theory.

3.2. The Epiphenomenal Objection

Nevertheless, barring these common misunderstandings, anomalous monism has been subject to the criticism that, despite advertisements to the contrary, it makes mental properties epiphenomenal.[72] For example, Honderich has argued that it is only by mentioning certain law-like properties that we can fully explain the causal relations between events. The general form of the critique goes like this: If we accept the principle of the nomological character of causality (hereafter, PNCC), as Davidson explicitly does, then it is only the law-like physical properties of an

event which explain its causal potency. The mental properties of an event are explanatorily irrelevant. Hence Davidson's critics jump from this observation about explanatory relevance to a metaphysical conclusion and argue that the mental does not, and cannot, interact with the physical in virtue of its contentful properties. Honderich's example of the French pears captures the force of their critique. He writes:

> Certain[ly] it is true that when I put some pears on the scale, something green and French did cause the pointer to move to the two-pound mark, but there in fact is no entailed law connecting greenness and Frenchness with the pointer's so moving. There is in fact no law at all connecting the event in virtue of its being something green and French with the pointer's moving to the two-pound mark (Honderich 1982: 60)

What is relevant to the explanation of the pointer's moving as it does is the weight and not the colour or nationality of the pears. For this reason Honderich advocates a revision of the PNCC which reads in terms of 'causally relevant properties'. But such an amendment means that, due to their lawless nature, the psychological properties of an event are not the sort of properties which could be causally or explanatorily relevant. Hence, Honderich's reading appears to threaten, "...the very strong requirement that the belief and desire [explain action]....in terms of their semantic contents" (Davidson 1987b: 43).

Applied to cases of alleged psychophysical interaction the position looks something like this. Imagine that Rais harbours a motive to buy his brother a toy for his birthday. He comes upon a shop which contains the perfect toy. However he meets a friend who convinces him that he should buy his brother the toy, not because it is his birthday, but just to show he cares. He resolves to do just this. Now how does Rais's original motive differ from his new resolution? Both are equally good reasons for acting, so what distinguishes the latter from the former? Recall that Davidson's response is to say that the latter is singled out as being causally efficacious in the production of the action. But Honderich reminds us that it is only the properties which are mentioned in strict causal laws which explain the efficacy of the event. And contentful properties cannot, in principle, be mentioned in such laws. Thus it makes no difference which belief/desire pair Rais harbours since what causes him to act ultimately depends entirely on the physics of the situation.

In effect Honderich claims that if we were to accept principle (iii) concerning the anomalousness of the mental, and the amended PNCC, then we must agree that, metaphysically speaking, "...the connection between the mental and the physical is accidental" (Honderich 1982: 64). This complaint is eloquently illustrated by the following story which he advanced in his debate with Peter Smith.

> As we go through the household of the Champion of the Mauve, we notice many things painted mauve. Here a light-switch, there a bottle of gin, here his bedroom slippers. Take the slippers – they have the effect of keeping his feet really warm. This is the effect, he more particularly allows, of just the Hibernian fleece with which they are lined. He also allows that their being painted mauve isn't nomically necessary to their Hibernian fleece, their causal property whose effect is the keeping of his feet really warm. But, he adds, it would be a wonderful confusion to move from that truth to any underrating of mauve in connection with his warm feet. The slippers with Hibernian fleece are identical with the mauve slippers (Honderich 1984: 86)

The standard way of representing the dilemma boils down to the proposal that Davidson must choose either to say that the mental causally interacts with the physical in nomological fashion or not at all. Honderich says, "Donald Davidson needs to choose whether a mental event as mental or that same event as physical which is causal with respect to an ensuing action, and....either choice is fatal to anomalous monism" (Honderich 1983: 147).

The choice certainly is fatal. For if Davidson maintains that propositional attitude psychology is anomalous, and he maintains the PNCC, then he must hold that the mental *qua* mental and physical *qua* physical do not interact causally at all. This would undermine his first premise of the argument for anomalous monism. On the other hand, if the mental and the physical do genuinely interact causally (in a nomological fashion) then the third principle is lost. It is for this reason, that Fodor, who favours the second reply, claims that, "It appears the cost of an intuitively adequate construal of mental responsibility is that there's no argument from mental causation to physicalism" (Fodor 1990b: 150). Clearly, if Davidson was to make either choice suggested by Honderich it would indeed undermine the argument for anomalous monism and cast doubt on the virtues of adopting the token identity theory.

3.3. Davidson's Replies

One possible reply, which Davidson has recently considered, is to advocate the view that the causal efficacy of propositional attitudes is underwritten by the possibility of developing non-strict laws that apply to that domain. He notes that, "...this defence of the causal efficacy of the mental is consistent with [anomalous monism]" (Davidson 1993: 10). But this is a dubious manoeuvre for as Kim has noted, given Davidson's views on the essentially normative nature of interpretation, his argument for the non-reducibility of propositional attitude psychology to a law-like pattern, "...should work against laws of all kinds – for example, statistical as well as deterministic ones" (Kim 1993: 25). This chimes

with Millikan's idea that the normativity involved, even with respect to intentionality, is not a norm of regularity of *any* kind or degree. For this reason it is a mistake that Davidson could, or ought to, enlist non-strict laws in defence of his metaphysical system.

3.3.1. The Weak Extension Reply

He has a second line of defence, however, which is to remind us that causal relations are extensional even though causal explanations are not. The dynamics of this claim are illustrated in figure 19. If one accepts what I dub 'the extension reply' then it would appear that the problem of the causal impotence of contentful properties no longer arises. It is the events themselves that stand in casual relations to one another, and they do not do so in virtue of their various properties (cf. McLaughlin 1993: 23). Causal relations are, as it were, primitive. This is what licenses Davidson's claim that, "...if causality is a relation between events, then it holds between them no matter how they are described" (Davidson 1993: 6).

Although, on this weak reading, we can accept that two events are causally related in extension it is vital to note that we have no clue why such a relation holds between them. If we understand events in extension in this way it implies that, in any given case, the fact that a causal relation holds between two events is an essentially unexplained brute fact. McLaughlin notes that causal relations are not accounted for by Davidson at all given that he, "...would deny that events are causally related *in virtue* of falling under strict laws" (McLaughlin 1993: 35 emphasis mine, cf. also McLaughlin 1993: 32, Kim 1993: 22, and Sosa 1993: 42). In light of his token identity solution it is clear why Davidson must take this line. Laws are intensional hence, although in certain contexts they are supposed to give us a full explanation of causal relations, they cannot make an appearance ontologically speaking. However if the PNCC is read as a physicalist doctrine about the nature of causation, as it is in reality, then Davidson's more modest monism is not truly physicalistic. It is perhaps better regarded as a variant of a Russellian neutral monism (cf. Buekens 1997).

Consequently this response is a mere papering over of the initial problem for two reasons. Firstly, it precludes what McLaughlin has called token-epiphenomenalism in that all the properties of any token event make a difference to its causal relations because if it lacked any of those properties it simply wouldn't be the same event. But this does nothing to touch the issue of type-epiphenomenalism which is the worry that there is nothing about propositional attitudes *as a class* that guarantees their causal efficacy (cf. McLaughlin 1993: 28, Sosa 1993: 42).

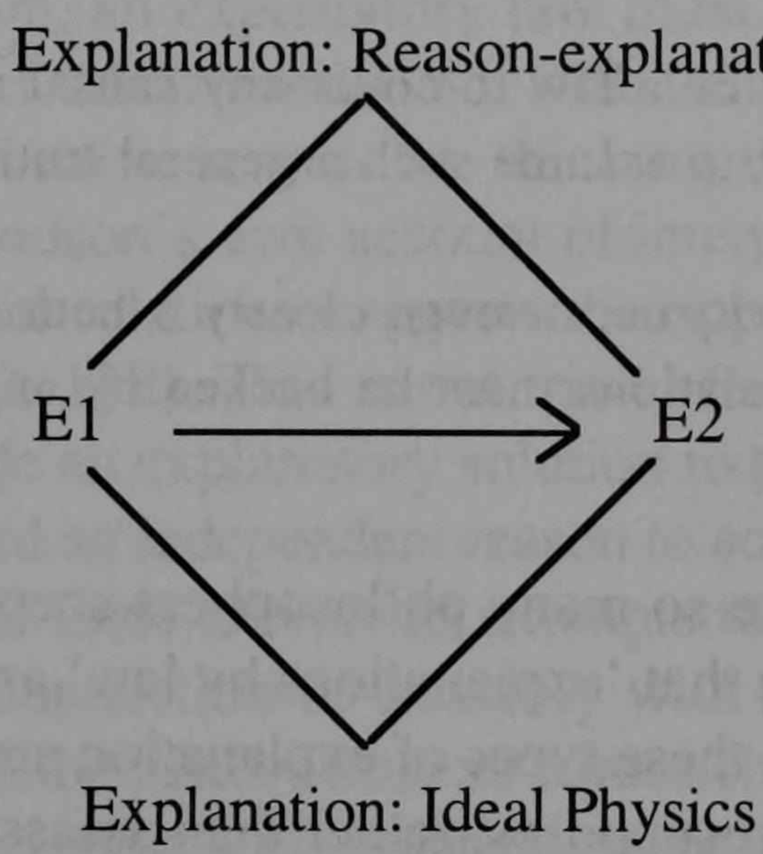

Figure 19. *The intensionality of explanation*

3.3.1.1. Scepticism about the PNCC and the Cause-Law Thesis

Secondly, and more importantly, although it is technically possible for Davidson to support the weak extension reply, his endorsing it ought to make us wonder about what secures his allegiance to the PNCC. As Tanney notes, "Davidson's move to go 'extensional' with causation, but to remain 'intensional' with laws, robs us of the explanatory connection between the laws and the particular events that the laws putatively subsume" (Tanney 1995: 117). This is crucial if we consider that despite its alleged status as an orthodox premise in contemporary philosophy of psychology, the PNCC has attracted a fair deal of scepticism over the years. Here is a small selection of quotations expressing doubts about it:

> It is over and over again assumed that any singular causal proposition implies a universal statement running 'Always when this, then that'; often assumed that true singular causal statements are derived from such 'inductively believed' universalities. Examples are indeed recalcitrant, but that does not seem to disturb. Even a philosopher acute enough to be conscious of this, such as Davidson, will say, *without offering any reason* at all for saying it, that a singular causal statement implies *that there is* such a true universal proposition – though perhaps we can never have knowledge of it. Such a thesis needs some reason for believing it! (Anscombe 1981: 147, first emphasis mine).

> I have to doubt Davidson's...premise, that all pairs of events related by 'cause' are subsumed under laws (Hornsby 1981–82: 86).

> Why must there always be a law to cover any causal relation linking events x and y? What enables us to assume such a general truth? (Sosa 1984: 278).

> I do not think it *a priori* true, or even clearly a heuristic principle of science or reason, that causal relations must be backed by any particular kind of law (Burge 1992: 35).

We might ask: Why are so many philosophers sceptical about the PNCC? In some cases they recognise that 'explanations by law' and 'explanations by cause' are quite different. In fact, these types of explanation are so different that it can be hard to see how the one presupposes, rather than assassinates the other. Strawson makes this clear. Once we have a law that relates two events, bi–conditionally, we lose many of the features we ordinarily associate with causal explanations. He remarks:

> So far as necessity and sufficiency are concerned, we have no reason for calling the A-type event a, or the, cause of the B-type event rather than vice versa; no reason, indeed, for distinguishing between cause and effect at all rather than recognising a symmetrical relation of mutual causal dependence (Strawson 1992: 129).

In a similar vein, Davidson himself tells us that once physicists develop a precise law they abandon causal language, for it is only when we are, "...lacking a law of the right kind, [that] it is essential to advert to the causal relation" (Davidson 1987b: 42). This fits with his observation that we often recognise the existence of singular causal relations, such as a stone breaking a window, without knowing how to explain the connection between events of this kind under law-like descriptions (Davidson 1980a: 16). Recognition of the differences between these forms of explanation has prompted Davidson's remark that, "I have been suggesting that appeal to causal concepts is appropriate to the explanation of action in part precisely because strict laws are not available" (Davidson 1987b: 42).

However this stance encourages a more interesting reason for scepticism about the PNCC which stems from Davidson's own argument that it is not possible, even in principle, to develop law-like explanations with respect to propositional attitude psychology (Rosenberg 1985b: 402). This is important to note, since the possible development of such laws in describing the behaviour of mindless phenomena is very likely what originally gave initial widespread confidence in the truth of the PNCC. But if there is no parallel between the kinds of causal explanations used in the mechanical sciences and those deployed in reason-explanations we must ask why should we continue to believe that the PNCC applies generally? Given that causal relations are brute facts, what could motivate

us to accept that, at bottom, an explanatory law must be at work whenever we choose to speak of them (cf. McDowell 1985: 398, Lennon 1992: 229)? Not only do we lack an obvious positive reason for thinking that there is always some covering law at work, Davidson's own account of interpretation and his views on the extensionality of causal relations suggests the opposite (Tanney 1995: 118, Strawson 1985: 133, 1992: 128). This is not trivial for it is the PNCC which is ultimately meant to provide an explanatory solution to the mysterious connection problem. Therefore we need an independent reason to accept it.

In a rare remark on this issue Davidson attempts to supply one by telling us that, "...the ground floor connection of causality with regularity is not made by experience but is *built into* the idea of objects whose changes are causally tied to other changes" (Davidson 1985b: 227, emphasis mine). In a recent paper, "Laws and Causes", he has augmented his earlier remark by discussing the views of both Hume and Ducassé on the nature of causation.[73] After some analysis, he insists that their accounts share a common *lacuna* in that they both implicitly appeal to the similarity of cases but fail to specify when 'cases are relevantly similar'. This is important to both thinkers because, "Hume needs to say when one change is relevantly similar to another ("Same Cause, same effect"); [whereas] Ducassé needs to be able to say when one state is not relevantly similar to another, i.e., when a change or event occurs" (Davidson 1995: 9). With this observation in hand he claims that the only way to designate the relevant events is if we presuppose that, at bottom, there is a law at work. Hence he concludes that, "It is not surprising then, that, singular causal statements imply the existence of covering laws: events are changes that explain and require such explanations" (Davidson 1995: 10).[74] It is on Kantian grounds such as these that he defends his cause-law thesis and insists that, "...we know in advance of evidence....that if a singular causal claim is true, there is a law that backs it, and we can know this without knowing what the law is" (Davidson 1995: 264). It is important that he claims that, "...the reason for holding the cause-law thesis....must in some sense be *a priori*, for the thesis is not a pronouncement of ordinary logic, nor can it be established empirically" (Davidson 1995: 2). But this is too strong.

The problem for Davidson's larger project is that if we accept that there is an *a priori* conceptual connection between causes and laws then we are left wondering why we ought to deny that events in extension are connected to one another in virtue of their law-like properties when causally related? The situation is exacerbated when McLaughlin points out that even if we accepted the extension reply it is still possible to specify which properties of events underwrite causal relations. He realises that even if, "...causal relations are extensional relations between events [this] is straightforwardly compatible with the claim that when events are causally related, they are so in virtue of something about each" (McLaughlin 1993: 32). Therefore we might ask, along with Kim, "What is it

about events c and e that makes it the case that c is the cause of e?" (Kim 1993: 22).

3.3.2. The Strong Extension Reply

If we accept the cause-law thesis, it would appear that a metaphysical version of PNCC is the obvious candidate for answering Kim's question. Hence we might advocate a stronger form of extension reply and hold that events are causally related in extension insofar as they so related in a law-like fashion in virtue of their ideal physical properties. This would automatically justify Davidson's view that when we say that events are causally related we are committed to the idea that there is an explanatory law available under some description. I would go so far to say that the *a priori* nature of the cause-law thesis, combined with Davidson's view that a logical analysis of language is an accurate guide to metaphysics, encourages, if not implies, this stronger reading (Davidson 1984a: 199). The difference is that on this reading the laws that explain how events connect to one another must be, in some way, generated by the law-like properties of the events themselves (or otherwise independently real and appropriately related to them).

But this simply resurrects Honderich's original problem by making the contentful aspects of an event causally impotent. Furthermore a commitment to this kind of hard physicalism opens up very serious questions about the ontological status of content. It is difficult to see how propositional attitudes could have any reality rights in a universe which, when truly described, is wholly deterministic. This would rule out the possibility of meaning because it would rule out the possibility of misrepresentation, and thereby representation itself. It is important to be clear about this point. The idea is not that norms are incompatible with a strict nomological view of the nature of causal relations. Rather norms are incompatible with the claim that the universe, when truly and *completely* described, is exclusively mechanistic in character. As we saw in chapter two, section 3.2, even Fodor has recognised that it is a formal feature of any account of representation that it must allow for an 'anomic' dimension. This is also why Davidson must employ an unreduced notion of cause in his interpretative approach, as discussed in chapter five, section 4.2 (cf. also Hutto 1996: 138–139).

The problem is that if the world, in extension, is just the world of an ideal physics, one in which all causal relations are strictly law-like, then it would seem content, representation, and intentionality in general are not genuine phenomena, since they require a universe that has room for at least some anomic characteristics. Hence, while this form of the extension reply makes space for the PNCC, it has unpleasant consequences if we wish to accept that propositional attitudes are a part, "...of reality – without which we couldn't do physics or

anything else – and [they] must occupy as fundamental a place in any credible world as matter, energy, space, time and numbers" (Nagel 1986: 7–8). To take Nagel's remarks seriously is to take propositional attitudes just as seriously as any other type of phenomena – both conceptually *and* ontologically. What becomes apparent is that the strong extension reply not only casts doubt on the causal efficacy of the propositional attitudes, it casts doubt on their reality as well. If Davidsonians were to adopt such a reading it would bring them close to the position of Dennett, who despite his preference for the title of mild realist, has occasionally avowed instrumentalism (cf. Dennett 1991a). For example, on this very issue, Dennett has claimed that, "Dennett and Davidson have pursued strikingly parallel but independent courses....[but differ only because of Davidson's steadfast] ontological realism about beliefs" (Dennett 1987: 348–349).

While strong instrumentalism, or mild realism, is certainly not a consequence Davidson seeks when advancing his non-reductive physicalism he has sometimes made remarks that reveal a serious tension in his account. For example, he has said things such as, "...the mental is not an ontological but a conceptual category" (Davidson 1987b: 46). In other instances, his metaphysical bias in favour of the ideal physical has caused him to enter into some suspicious talk about the kinds of explanations one can get from an ultimate physics. He sometimes talks of ideal physical explanations (those given in terms of strict laws) as being necessarily superior to or 'more true' than our everyday causal explanations. Consequently, despite recognising their intensional nature, he tells us that they are not 'interest-relative' and that they are to be contrasted with the causal explanations of all other 'interest-sensitive' discourses (cf. Davidson 1987b: 45, 1995: 16). Although he insists that ideal physical explanations will never replace those of commonsense psychology in the way they could replace those of folk physics, he makes remarks which suggest, nonetheless, that the absence of commonsense psychological laws reveals that, "...reason-explanations, by virtue of the features I have been depicting, are in some sense low-grade; they *explain less* than the best explanations in the hard sciences because of their heavy dependence on causal propensities" (Davidson 1987b: 42, emphasis mine, cf. also Davidson 1990c: 7). This comparative verdict seems wholly inappropriate in light of the irreconcilable differences between the explanatory concerns of commonsense psychology and ideal physics. One must ask what does 'explain less' mean in this instance? Certainly folk physical explanations explain less than ideal physical explanations. But this is because they are competitors in the same domain. But, in light of his other views, how can Davidson reasonably compare ideal physical explanations with those of commonsense psychology?

These occasional remarks indicate that he sometimes thinks that our everyday causal explanations are indeed, in some ultimate sense, false or inferior. But for this to be so it would imply that what we ordinarily designate to be the cause of an

effect does not in fact pick out what really made it happen. Furthermore this predilection for an ideal physics fits with his remark that, "...interest aside every property of every event is causally efficacious" (Davidson 1987b: 46). This remark is unsurprising given his monism, but in conjunction with the above comparisons, it can be confusing. This is because, in the context of the weak extension reply, law-like explanations, like any other, are merely intensional. It is puzzling, therefore, how they can cast interest aside, capture all the aspects of an event that make a difference while, simultaneously, staying within a single explanatory domain. The tension would dissipate if Davidson were to endorse the view that causal relations between events in extension must be law-like *in nature,* but then the aforementioned problems attendant with the strong extension reply would become explicit. To reiterate, the problem with that version of the extension reply is that if we view ultimate reality as being ideally law-like in character then we cast doubt on the very reality of the propositional attitudes.

To summarise, Davidson's dilemma is this: (i) the weak extension reply on its own isn't sufficient to support the PNCC, (ii) anything strong enough to support the PNCC would be strong enough to warrant the strong extension reply, but (iii) the strong extension reply threatens the stability of anomalous monism by threatening the causal potency and reality of the propositional attitudes. For these reasons, it would appear that either way the extension reply is bad news for anomalous monism. The only technically sustainable move, compatible with non-reductionism, is to adopt the weak reading. But such a reading renders the nature of causal relations mysterious. Hence if we are seeking a principled account of why a reason and an action are connected then Davidson is bound to disappoint. At most we know simply that they are. Consequently the problem of the mysterious connection between reasons and actions remains as much a cause for concern as ever.

4. The Mysterious Connection Problem Reconsidered

It is worth emphasising that Davidson's primary motivation for a causalist treatment of 'because' in reason-explanations derives from his admiration of Aristotle's account of action. He explicitly says, it was Aristotle who, "...introduced the concept of wanting as a causal factor" (Davidson 1980a: 11, cf. also Davidson 1987b: 40). Moreover, he has claimed, "...I would urge that, failing a satisfactory alternative, the best argument for a scheme like Aristotle's is that it alone promises to give an account of the 'mysterious connection' between reasons and actions" (Davidson 1980a: 11).

4.1. Agency and Non-Humean Causation

Given this, it is vital to recognise that Aristotle would not have formulated the mysterious connection problem in the first place. Why not? The problem only arises if one assumes a Humean-based account of causation – weak or strong (cf. Johnston 1989: 46). In this light, it is appropriate to remind ourselves of the Neo-Wittgensteinian position which Davidson was reacting against. Melden was explicit about the target of his critique. He wrote, "I am concerned with 'cause' either in the Humean sense of this term or, if this is alleged to be inadequate in certain respects, the use of the term in scientific explanations of, say, physical or physiological events, in that sense of the term in which it is in fact employed in physics or physiology" (Melden 1961: 16).

In contrast, to ally oneself with Aristotle is to re-enforce the idea that reasons gain their explanatory force by consideration of agents and their goals. It is useful to remember that Aristotelian causes and their modes of explanation are not causal in the modern sense of the term as it has developed post-Galileo and has been influenced by Humean and Kantian traditions. For this reason, Aristotle would not have worried about how such a view is to be reconciled with an impersonal, mechanical world-view.[75] Without necessarily seeking to resurrect The Philosopher's position in all its glory, noting this point is useful as a means of understanding the kind of conclusion that Gordon advocates when he writes:

> I am inclined to say that the 'because' in reason explanations indicates a 'causal' relation, but not in *the same sense* as in a naturalistic explanation of, say, a tree's being split. Rather, the relation is 'causal' in the original 'primitive' sense of 'cause', as Collingwood (1946) and Strawson (1985) have urged (Gordon 1992: 30).

What is the primitive sense of cause to which Gordon refers? To answer this we must remind ourselves of Strawson's proposal that our knowledge of causality derives, in the first instance, from our ability to imaginatively project ourselves into the world (cf. Strawson 1985: 122–123, 1992: 117–118). Strawson rejects the Humean account which links the development of the concept of cause to our intellectual habits in response to the passive observation of constant conjunctions between kinds of event. Instead, he suggests this concept has its roots in our experience of 'knocking into' things and being 'knocked into' by things. In other words, our basic notions of causation may initially stem from the phenomenological effects of our own practical engagement with the world – from a sense of felt external force and our resistance to such forces in the carrying out of our projects. This is particularly interesting in light of Horgan's criticism that non-causalist approaches do not tally with, "...the phenomenology of our own agency" (Horgan 1991: 88). He claims that agents have introspective awareness of

the causal potency of their own mental states. Dretske has replied by saying that, rather than having its basis in introspective fact, attachment to such a view about here-and-now mental causes is nothing other than a 'Davidsonian dogma' (cf. Dretske 1991: 198). But if we adopt the Strawsonian line we can accommodate Horgan's point about phenomenology precisely by rejecting a Davidsonian, or broadly Humean vision of causation.

Furthermore, according to such an account, initial attempts to extend the notion of cause to inanimate things would have, at least initially, involved a projection of human-like agency onto those things. Strawson's speculation on this possible extension of the concept is interesting because it further emphasises its anthropomorphic basis. This can be seen when he writes:

> In a great boulder rolling down the mountainside and flattening the hut in its path we see an exemplary instance of force; and perhaps, in so seeing it, we are in some barely coherent way, *identifying with* the hut (if we are one kind of person) or with the boulder (if we are another): putting ourselves imaginatively in the place of one or the other (Strawson 1985: 123, 1992: 118, emphasis mine).

This proposal fits with what we might expect if simulation theorists are correct about our initial tendency towards total projection when seeking to explain another's behaviour, as discussed in chapter four, section 3.2 (cf. Hutto 1997, 1994). Just as with non-verbals, recognising this tendency is not to endorse the practice such ascriptions in all cases. In particular, they are inappropriate when it comes to explaining the alien interactions of things impersonal, inanimate and microscopic (Strawson 1985: 124–125, 1992: 120). Learning this lesson took time for our anscestors. For example, we know that the actual historical development of our attempts to understand the natural world have this pattern. Strawson reminds us that:

> Our primitive, and not so primitive, theorists, aware of their own powers of agency and of the motives behind their exercise, aware, also, of vast effects in nature, dreaded or hoped for, quite beyond their own powers directly to avert or produce, seem to have found it utterly easy and natural to attribute these effects to the exercise of powers by superhuman agents who, capricious as their acts must have often appeared, were actuated by motives not wholly alien or inscrutable (Strawson 1985: 129, 1992: 124–125).

The point is that the way we understand causality in some domains ought not be confused with the way we understand its operation in the concept's original home: that of human agency. Nevertheless, there is some justification in thinking

that our use of the concept 'causation' is single-tracked. At a high level of abstsraction the concept of causality is *sui generis*. But we have to be careful in saying what it is that is common to all forms of causality. An examination of ordinary discourse, psychological or otherwise, reveals that, when we speak of causes, we are always pointing to, "...that which produces an effect" (Anscombe 1981: 136). Anscombe regards this as an important, albeit trite, observation. This understanding of cause is shared in both psychological and physical cases. Recognising this allows us to make sense of Melden's remark that:

> It is certainly true that we use 'cause' in speaking about the actions of agents, but we can no more infer from this verbal consideration that actions are the Humean effects of events than we can from the etymological derivation of the term 'motive' (Melden 1961: 208).

This should give us pause to ask: Are reasons causal in anything like the way physical things are? Certainly, even recognising their normative character, reason explanations are still suggestive of counterfactual patterns of behaviour. Offering a reason to explain an action sets up all sorts of defeasible, *ceteris paribus* expectations about how the agent ought to behave (cf. Hutto 1997: 70–72). For if an agent fails to behave in a consistent way we are forced to reassess our interpretation of them. But if this is the only extent in which there is a parallel with causal explanation in the impersonal domain then we have only limited grounds to call reasons causes or to think that much of their explanatory force derives from the fact that they are causal explanations. Especially since, if the analysis of this chapter is correct, the standard version of such an appeal neither solves (i) the primary reason problem nor (ii) the mysterious connection problem.

In conclusion, the main point is that if we adopt the Strawsonian line then how reasons and actions are connected is no longer a puzzle in need of a solution. If our basic or original notion of causation rests on or develops from our felt interactions with the world (as opposed to a mere Humean observation of events) then the mysterious connection problem simply does not arise. Acting is a paradigm example of causal interaction – if we think of causation in a non-reductive, non-Humean sense. Looking at things in this light aborts the mysterious connection problem before it can develop. The trick to understanding the nature of reason-explanations is not to treat them as if they were on par with other forms of causal explanation found elsewhere in the natural sciences.

5. Conclusion

In the previous chapter I argued that propositional attitude psychology must be understood by interpretationalist lights. And I have now argued that it cannot be

incorporated into a metaphysics which is inspired by physics. This brings us to the second major question which I want to address in this book: Should we eliminate propositional attitude psychology because it is out of step with a purely scientific metaphysics? I answer 'No' and give my reasons in detail in the next and final chapter.

CHAPTER SEVEN

The Poisoned Chalice

"It is still a *metaphysical faith* upon which our faith in science rests –
we seekers after knowledge, we godless anti-metaphysicians
still take our fire, too, from the flame lit by a faith
that is thousands of years old."

– Nietzsche, *The Gay Science,* §344

1. Introduction

I have endorsed a non-reductive interpretationalist line concerning propositional attitude psychology and argued against compatiblist attempts to show that it is in harmony with physicalism. Consequently with respect to the issue of commonsense psychology's prospects for naturalisation I have backed myself into a corner. Understanding agents in terms of their beliefs and desires is an irredeemably unscientific practice. This admission alone is sufficient to motivate an eliminativist verdict. Given this, there is nothing for it but to fight – *aut vincere aut mori*. In this chapter I argue that eliminativism is incoherent because the best naturalistic account of content it can offer is insufficient to explain scientific theorising.

2. Eliminativism: Plan B

In chapter one I described two eliminativist strategies. Using strategy A the eliminativist attempts to get us to believe that commonsense psychology is already in the process of being undermined by advances in cognitive science. The analyses of chapters two, three, four and five repelled this line of attack. However there is another argument for eliminativism which is *prima facie* more plausible and more insidious. Strategy B encourages us to eliminate commonsense psychology on the simple grounds that it is not, and will never be, an integral part of the rigorous and developing scientific world-view.

2.1. Churchland's Coherentist Argument

In this context Paul Churchland proposed the following three criteria against which to judge the virtues and vices of any theory or conceptual framework.

(i) Review its explanatory successes and failures.

(ii) Examine its growth, fertility, and current promise of future development.

(iii) Consider its active coherence and continuity with fertile and well-established theories in adjacent and overlapping domains.

As might be expected, commonsense psychology is not thought to fare well on any of these fronts. I discuss criteria (i) and (ii) very briefly as I do not think they provide grounds for complaint. The bulk of my attention is therefore directed at criterion (iii) since it is here that the eliminativist case must be answered.

With respect to criterion (i), it is beyond doubt that commonsense psychology serves our quotidian needs quite successfully. In doing so it makes sense of the behaviour of most of our peers, albeit the kinds of explanations it provides are not causal explanations of a naturalistic variety as I argued in chapter six, section 4. Its predictive success in this domain is also remarkably high, especially when it comes to predicting our own behaviour (cf. Gordon 1986). Even the Churchlands must admit this. In this light, Paul Churchland is wrong to complain that it fails to account for 'certain central and important mental phenomena' such as mental illness, creative imagination, the function of sleep, memory, perceptual illusions, the ground of intelligence differences, and the internal construction of 3-D visual images. In gesturing at this catalogue of alleged failure he is only stating the limits of commonsense psychology (cf. Sterelny 1993: 311). To treat these limitations as failures is to mistakenly anticipate the verdict of criteria (ii) and (iii).

To think that commonsense psychology fares badly with respect to criterion (ii) is to accuse it of stagnation and infertility. But this accusation implicitly and illegitimately assumes it to be a type of research-programme. This view was discredited in chapter one, section 2.2. Without this assumption in play, it is hardly a flaw that commonsense psychology hasn't changed much over the years. Some philosophers reject this charge by claiming that propositional attitude psychology has indeed undergone subtle change and development over the years (cf. Horgan and Woodward 1990: 152-153). Nonetheless I prefer Blackburn's approach and would rather join him in saying, "One man's stagnation is another man's certainty" (Blackburn 1991: 199, cf. Gordon 1986).

This brings us to the greatest perceived problem with commonsense psychology: its general lack of fit with the rest of science. The most convincing argument for elimination is based on the simple fact that propositional attitude

psychology will not integrate with natural science. We have seen the reasons for this verdict in chapters two, five and six. This is connected with the eliminativist adherence to a coherentist epistemology. Churchland remarks that, "Our best and most penetrating grasp of the real is still held to reside in the representations provided by our best theories. Global excellence of theory remains the fundamental measure of rational ontology" (P.M. Churchland 1989: 151). The eliminativist battle-cry is, "If our....judgements must be laden with a theory in any case, then why not have them laden with the best theory available?" (P.M. Churchland 1979: 35). The follow-up thought is that the best theories are those which are developed wholly within the bounds of natural science. In this way the charge of a lack of coherence becomes damning if we accept that *only* science can tell us what is real. This attitude is tidily expressed by Churchland in the following passage:

> If we approach *Homo Sapiens* from the perspective of natural history and the physical sciences, we can tell a coherent story of his constitution, development, and behavioural capacities which encompasses particle physics, atomic and molecular theory, organic chemistry, evolutionary theory, biology, physiology, and materialistic neuroscience (P.M. Churchland 1989: 212).

Against this background our everyday understanding of agency is nothing more than a misleading and unscientific competitor, or worse an unsightly adjunct, to a more rigorous scientific account of human nature. It is by fostering such an attitude that the eliminativist has the best hope of convincing us to abandon 'the person theory of humans'. By attacking on this front it is not necessary to show that commonsense psychology: (a) is (or will be) displaced by revolutionary theories in cognitive science, (b) is generally unsuccessful or unexplanatory in its domain or (c) is a stagnant research programme. All that needs to be shown is that it is out of step with natural science (cf. Cussins 1992).

For example, the apparent success and usefulness of the enterprise is no protection against this charge. In making this point, its enemies get us to focus on other cases in which mistaken theories have put on brave faces, and in some instances have astonished us with their apparent explanatory successes. Paul Churchland sees the 'caloric fluid theory' of heat in this category. Caloric theory conceives of heat as caused by the presence of an invisible fluid. This could account for many observable heat-related phenomena. For example, the expansion of heated substances could be explained by the increase in the volume of the fluid. Likewise the change of colour of heated substances could be explained by changes in the degree of the liquid concentration. However, despite these positive features, the 'caloric fluid theory' collapsed due to, "...its explanatory impotence in certain cases" and, "...its lack of any non-redundant place in the broader conception of

the world provided by modern physical theory" (P.M. Churchland 1979: 16-19). Since the theory was empirically adequate we might wonder: What motivated abandonment of caloric fluid theory in favour of kinetic theory? The short answer is that the latter is a better theory. But better in this instance cannot mean 'more useful' since we are assuming the theories are empirically adequate (cf. Van Fraassen 1981). Kinetic theory is only better because it is in more thorough-going harmony with other accepted scientific theories. For this reason its explanations are superior. A high premium is placed on a theory's coherence with other accepted theories. Therefore insofar as the virtues of kinetic theory over caloric fluid theory are pragmatic they are global, not local.

With this comparison in hand, as soon as we admit that propositional attitude psychology will not form part of a more scientific psychology the eliminativist declares that it should be abandoned. But this is too fast. Put crudely, the problem with this move is that it draws an arbitrary boundary around the class of scientific theories when considering the issue of coherence.

3. Assessing the 'Lack of Coherence' Charge

One immediate problem with this approach is that the eliminativist must be prepared to demarcate the boundaries of what counts as proper science. To successfully advocate exclusive scientific realism the eliminativist really ought to first answer the question "What are the credentials for being a science?" Certainly the answer to this cannot be just that such theories pass the test of unity or coherence with other scientific theories. Apart from being obviously circular, such a criterion would discredit most of the known sciences. For it is not the case that the special sciences – such as biology and quantum mechanics – integrate well or cohere internally or with other natural sciences such as physics (cf. Hutto 1998c, forthcoming-b). The problem, as Rorty notes, is that 'science' is not a natural kind term (Rorty 1988a). But it is not my purpose to rest my argument on difficulties of providing a principled answer to this question. Eliminativist plan B comes to grief for reasons other than the fact that science is not a definitionally watertight term.

A much more serious problem is that the quest for the 'best psychological theory' – and thus the best view of our self-understanding – is fraught with special difficulties; difficulties which the scientific naturalists, in their moments of zeal, fail to recognise. In the larger scheme of things when it comes to giving meaningful expression to the human condition it makes better sense, is more coherent, to accept the truth of some form of propositional attitude psychology than to deny it. Even for eliminativists, it is not possible to surrender content-ascribing propositional attitude psychology on pain of a more general incoherence. Unless an alternative picture of conceptualisation is on offer such a manoeuvre

threatens the epistemological picture which serves as a meta-condition for their attacks. Consequently eliminativism becomes incoherent when it attempts to dispense with the categories of truth, belief and rationality.

There is a delicious irony at play here which is worth highlighting. For example, Paul Churchland has remarked that a, "...'bad faith' or 'inauthenticity'....dominates current discussions of Eliminative Materialism" (P.M. Churchland 1993: 211). In his view, those who defend an irreducible propositional attitude psychology are engaged in nothing other than sentimental, "...special pleading on behalf of the old and familiar" (P.M. Churchland 1989: 125). In conciliatory fashion, he remarks that, "There need be nothing inauthentic about declining eliminative materialism....if one declines the epistemology that makes it possible" (P.M. Churchland 1993: 212). But I claim that it is precisely in accepting a broadly coherentist epistemology, described in chapter one, section 2.2.1, that the best reason for rejecting eliminativism becomes transparent. Recall that accepting a minimal coherentism provides the most stable basis for eliminativism. But, unless arbitrarily restricted, coherentism does not allow us to preclude the possibility that the metaphysics inspired by natural science, not propositional attitude psychology, wants elimination. If it turns out that there are good reasons to maintain propositional attitude psychology in the face of its resistance to naturalisation then it is materialism or physicalism which ought to be rejected. This is not a possibility that the Churchlands allow themselves to contemplate. Hence it is true that there is 'bad faith' infecting discussion, but its source is in the eliminativist camps. Insofar as they assume, on unargued *a priori* grounds, that materialism is beyond question they have not taken coherentism to heart (cf. Johnston 1993: 212-218). In contrast, I am prepared to spell out and defend my reasons for thinking we have no alternative but to accept some form of non-reductive account of human cognition and psychology.

3.1. Why Minds Matter

When attacking propositional attitude psychology the eliminativist focuses on its reliance on the irreducible notions of truth, rationality and propositional content. Nonetheless, as many have observed, they take it for granted that we are able to compare and contrast theories or conceptual schemes. But this evaluative ability rests on the fact that such theories have content to compare. It is here that eliminativism becomes self-defeating. In order to make evaluations of the virtues and vices of various theories, and their attendant referential resources, we must know what their terms mean, what kind of entities they posit, and what views they commit us to. In all we must know how one theory differs from another in its conceptual resources and semantic value. This being the case, it hardly can be coherently argued by the eliminativist that contentful, propositional attitude psychology is fit for eradication. If this counter-intuitive thesis were true then no

contentful theory, scientific or otherwise, will have ever been advanced (cf. Hornsby 1992: 157, McDonough 1991: 272).[79] Hence concepts and contents must exist for commonsense psychology to be possibly false.

From the beginning the Churchlands expected to be accused of advancing a performatively incoherent claim in the very act of stating their position. They anticipated being accused of not believing in beliefs; which is a logical impossibility. Against this charge they cite an illegitimate objection to a seventeenth century attack on vital spirits which has a similar form. In that debate the antagonist denies the existence of vital spirits and is accused, unjustly, of advancing an incoherent claim. The vital spirit theorist insists (standing by their theory) that if the antagonist's claim were true then the antagonist would be without a vital spirit. But to be without a vital spirit is to be without life and to be without life is *ipso facto* to be unable to claim anything (P.M. Churchland 1981b: 222). In the form stated, the line of argument used to discredit the opponents of vital spirit theory is rightly criticised by the Churchlands. But the eliminativists have not helped their position by drawing this analogy. This is because both the believers and non-believers in vital spirits were presumably agreed, if only implicitly, that in order to make claims of any sort a person must be alive. The source of their disagreement was about what constitutes life. That was the real focus of their debate. To deny that there was any such phenomenon would have made their debate senseless. So it is with propositional attitude psychology.

The existence of content (of some kind) is a necessary condition for theorising, conceptualising and possibly getting it wrong (cf. Baker 1988: 124). Nobody who wishes to be taken seriously denies this. Content is what makes debate possible. The eliminativists cannot coherently question the fact that thoughts have conceptual content *per se*. At best they can question that the tradition has rightly understood its essential character (cf. Cussins 1993: 235). Thus it is possible to argue that we have not understood content until we have understood it scientifically. What they must say is that we are confused about the nature of content when we elude to 'thoughts', 'truth', and 'meanings' in the way the vital spirit theorists were confused about the essential conditions necessary for life (cf. Sterelny 1993: 313). Compare with the case of Dennett's madman who claims not to believe in the existence of animals. Dennett writes:

> Suppose a madman were to claim that there were no such things as animals. We might decide to confront him with his error by taking him to the zoo and saying 'Look! What are those things, then, if not animals?'....But suppose he then said 'Oh, I know perfectly well that there are these things – lions, ostriches and boa constrictors – but what makes you think these so called animals are *animals?* In fact, they are all just fur-covered *robots* – well, actually some are covered with feathers or scales.' This may still be craziness,

> but it is a different and more defensible kind of craziness. This madman just has a revolutionary idea about the ultimate nature of animals (Dennett 1991c: 43).

So the real disagreement concerns how we are to best understand the nature of content. This is genuine arena for debate in which their can be winners and losers. Anti-reductionists deny that content can be understood in purely scientific terms. Eliminativists must assert the opposite. But in so doing they must not retain or vindicate any of the familiar, conservative theories of content. As we have seen expansive classical cognitivists are committed to explaining our semantical, referential abilities (as traditionally understood) in a non-semantic, purely scientific framework. They anticipate that one day a theory of propositional, representational content will be vindicated by a naturalistic account of how our mental machinery works in relation to stimuli all of which can be identified wholly impersonally (Cummins 1991a: 1-2). In contrast eliminativism is premised on the assumption that a retentively reductive account is not on the cards. It is partially motivated by pessimism about such a project. Indeed Paul Churchland compares the reductive strategies of identity-theorists and functionalists to the backward attempts to preserve the outdated categories of alchemy from the onslaught of a superior chemistry. Consequently in denouncing the aspirations of expansive classical cognitivism, the eliminativist is left in the awkward position of providing a purely naturalistic but non-traditional account of content and cognition (cf. Akins 1996: 338).

In a nutshell, the onus is on the exclusive scientific naturalist – classical cognitivist or eliminativist – to show that Nagel is wrong to say that, "...the mental activity of forming an objective conception of the physical world seems itself not capable of physical analysis" (Nagel 1986: 15). It is here that the eliminativist is caught on the horns of a dilemma. The first horn is: if they fail to meet Nagel's challenge their position is made incoherent due to their epistemological commitments. The second horn is: if they successfully use traditional categories in order to meet Nagel's challenge they will have vindicated the very aspects of cognition which they hoped to eliminate.

In this context, Paul Churchland confesses, "I remain committed to the idea that there exists a world, independent of our cognition, with which we interact, and of which we construct representations" (P.M. Churchland 1989: 151). Nonetheless he also explicitly recognises that acceptance of the traditional view of the nature of theorising, coupled with an attack on the classical theories of cognition, is the Achilles' heel of his early arguments (cf. P.M. Churchland 1993: 214). In response, he has vowed to provide an unorthodox account of theorising which, "...contrasts sharply with the kinds of representational and processing strategies that analytic philosophers, cognitive psychologists, and AI workers have

traditionally ascribed to us (namely, sentence-like representations...)" (P.M. Churchland 1989: 130-131). I question the prospects of this new account in its bid to deal adequately with the kind of content required for theorising.

4. The Eliminativist Burden

Eliminativists have heavy commitments. First, they must accept that content exists. Secondly, they must explain its nature in purely naturalistic terms. Finally, they must forego traditional accounts when doing so. In this regard, they seem to have just two possible routes open to them. They can either support claim (A) or claim (B) below:

(A) The nature of conception can be completely understood in neuro-computational terms.

(B) The nature of conception can be (in part) understood in neuro-computational terms and (in part) by appeal to external factors.

The structure of my argument in this section is as follows. I discredit claim (A) in order to show that eliminativists cannot avoid making claim (B). I then give reasons for thinking that claim (B) cannot be read neutrally and must be converted to claim (B'), which makes explicit mention of socio-cultural factors. I conclude by expressing scepticism about the viability of holding claim (B') in conjunction with any interesting form of exclusive scientific naturalism. I begin by assessing claim (A).

4.1. What Brains Can't Do

Paul Churchland has heroically proposed to replace the classical cognitivist account of theorising and concept formation by appeal to neuro-computationally based connectionist theories. His hope is to explain the nature of cognition by coming to, "...understand how empirical brains actually do learn" (P.M. Churchland 1989: 302). Note, however, that this hope rests on the unargued and contentious view that brains are indeed 'learners' or 'conceivers'. Thus he occasionally makes cavalier remarks such as, "...language use is something that is learned by the brain" (P.M. Churchland 1989: 16). Furthermore, because he sees no fundamental difference in kind between the information processed by people and their brains, he directs us to compare our conceptual abilities unfavourably with those of brains. He writes:

> A developing hemisphere *learns* to make use of the information the cerebral commissure deposits at its doorstep....The cable commissure consists of

roughly 200 million neurons, and even if we assume that each of these fibres is capable of only two possible states each second (a most conservative estimate), we are looking at a channel whose information capacity is greater than 2 x 10^8 binary bits per second. Compare this to the less than 500 bits/second capacity of spoken English. Now, if two distinct hemispheres can learn to communicate on such an impressive scale shouldn't two distinct human brains learn to do it also? (P. M. Churchland 1989: 20).

What Churchland suggests is that brains 'conceive of things' more efficiently than we do. Thus he encourages us to, "...look beyond such parochial virtues as Tarskian truth, since that is a feature unique to the parochial elements of human language, which is a peripheral medium of representation even for human cognition" (P.M. Churchland 1989: 301). But to talk of 'brains as conceivers' is deeply confused. Unless we are speaking metaphorically, such talk is nonsense – brains are not themselves thinkers, conceivers, planners or the like. When Churchland sketches the possible evolution of *Homo Sapiens* into non-linguistic beings who 'communicate' after the fashion of the right and left hemispheres of the brain, he is asking us to imagine an impossible future for human beings (see P.M. Churchland 1989: 17–22). In presenting his vision of the brave new world, Churchland is, in fact, advocating nothing short of cognitive suicide (Baker 1988). He is not even asking us to imagine an alien form of life. Trying to imagine how brains respond to their 'environment' belies deep confusion. Given that brains do not even interpret intentional icons they have no form of life at all. They have no more of a form of life than the life led by my kidneys or other bodily organs. Malcolm makes a similar point in an attack on mind-brain identity theories. He writes:

> Now....a brain....is not a human being; it does not have a face or limbs; it cannot see, hear, speak; it cannot exhibit surprise, desire, intention, belief; it cannot look, move, gesture, act; it does not share a language with others. It makes as little sense to ascribe experiences, wishes, thoughts, beliefs, to a brain as to a mushroom (Malcolm 1986: 186).

But not only are Churchland's remarks about the conceptual abilities of brains confused – they are misleading. They direct attention away from the important aspects of conceptualisation which cannot be understood in purely neuroscientific terms. We can see such misdirection at work by considering Bechtel's critique of Churchland's account of large scale conceptual change. For example, he writes that according to Churchland's connectionist account conceptual change occurs:

> ...when a network gets trapped in a local minimum and must be bumped out of it by an infusion of noise that significantly alters the current weights in a

> network. With luck, the network will then be able to find a deeper minimum....While this account may characterise what occurs in the scientist as he or she undergoes large-scale conceptual change, it neither explains what causes the change (the sort of noise that will bump a network out of a local minimum) nor its rationality (especially since most bumps fail to lead to deeper minimums) (Bechtel 1996: 123).

The point is that, in order to understand the nature of conceptual change properly we need to know; which concepts are being employed, what semantic value they carry, and what reasons have driven the change. Although a working brain is a necessary condition for conceptualisation, an account of the activity of our neural mechanisms is of no use in answering such questions. This alone is sufficient to show that concept development, learning and change are not simply forms of brain activity which can be entirely understood in neuro-computational terms. Moreover the conceptual content we master in cases of genuine learning does not originate in the brain. Churchland comes close to acknowledging this when discussing how science is first learned.

> In school and university *we are taught* to recognize a panoply of complex prototypical situations – falling bodies, forces of equilibrium, oxidation, nuclear fission, the greenhouse effect, bacterial infection, etc. – and *we are taught* to anticipate the prototypical elements and effects of each. This is unquestionably a process of learning....But it is just as clearly a process of socialization, a process of adopting the conceptual machinery of an antecedent society (P. M. Churchland 1989: 300, emphasis mine).

As this quotation makes clear, in the early stages learning is largely the adoption of an intellectual tradition, one preserved in writing and one which generally requires an understanding of the views and values of an antecedent society (whether this is explicit or implicit in the material learned). However the Churchlands qualify this remark by saying, "*Institutionalized* science certainly depends upon [public devices and techniques], but individual theorizing and explanatory understanding need not" (Churchland and Churchland 1996: 266). They insist that the right views to hold are:

> (a) speculative attempts to understand the world have their primary, original and still typical home within the brains of individual creatures
>
> (b) an adequate account of that original activity must precede and sustain the secondary account of its subsequent flowering within the social matrix of the occasional social animals such as *Homo Sapiens* (Churchland and Churchland 1996: 267).

4.2. The Training of Connectionist Networks

The Churchlands claim that the activity of understanding is best explained by appeal to neural representations, the basic characteristics of which are best modelled on connectionist networks. These neural representations provide the non-linguistic means by which individual brains allegedly theorise. But it is demonstrable that the activity of connectionist networks simply do not constitute theoretical activity. It is not possible to judge a neural network's conceptual responses (such that it has any) to be correct or incorrect in absence of a standard supplied by the natural world or society. For example, some connectionists have been encouraged to think that networks are capable of making cognitive responses precisely because they appear to be capable of learning. But let us also remember what such learning looks like:

> The network's information is encoded in the weights between the nodes. Typically, these weights are modifiable. The network's capacity to learn (and, in effect, its memory) depends on this modification process. Learning proceeds by slight modification of the weights between nodes. This is known as 'training' a network (Sterelny 1990: 170).

Such training can come about in a number of ways. Networks are not rigid or pre-programmed and this is what makes them so interesting. They often show the capacity to make guesses and to generalise when presented with stimuli different from the kind shown to them during their initial training sessions. Yet it is important to note that if the network is to perform intelligently at all its responses must be evaluated with reference to the norms of its teacher (which defines the training set) (cf. P.M. Churchland 1989: 133, Lockwood 1989: 119). As Sterelny notes, although networks can be trained, "...they cannot do this by themselves; they need an outside agent – something outside the network – to correlate teaching stimuli with their *correct* classifications" (Sterelny 1990: 170, emphasis mine). In other words, even in the best cases, a network learns by co-ordinating its responses to the standards set by its teacher and it is only with reference to the teacher's standards that we might be driven to think that the network has successfully mastered a concept or not.

What might serve as the outside agent or teacher of which Sterelny writes? Well, if the connectionist literature is to be believed, quite a lot of things really – another computer, evolutionary pressures, or a linguistic community (P.M. Churchland 1989: 231-253). It is partly because connectionist networks display the *prima facie* characteristics of teachablity that they are of interest to cognitive scientists. It is their ability to produce intelligent, appropriate responses – as judged by us – that makes us want to distinguish them as cognitive as opposed to

mindless. Importantly then, the first step in interpreting the responses of these networks as cognitive requires appeal to normative factors which are unavailable without teachers and their standards. But these teachers and their standards are not neuro-computational events nor are they explicable solely in terms of such events.

Furthermore, close examination of our descriptions of the behaviour of connectionist networks reveals our need to employ the categories of commonsense psychology. As Clark, notes:

> The writings of [working connectionists and connectionist-inspired philosophers] are steeped in content-ascriptions of a relatively familiar kind. Thus we find talk of networks learning to recognise typical (and atypical contents for rooms (Rumelhart, Smolensky, McClelland & Hinton: 1986), to distinguish rocks from mines (P.M. Churchland 1989: Chapter 9), to group together animate objects (Elman 1989), etc., etc. (Clark 1996: 228).

The point is that any attempt to ascribe intelligence or understanding to such networks, even in a derivative sense as licensed by the first moves in interpretative practice, reveals our reliance on our familiar categories.

4.3. The Nonconceptualist Reply

As we have seen the Churchlands are committed representationalists. But to make their heterodox account of theorising plausible they need to do more than simply appeal to connectionism in order to provide an understanding of 'neural representation'. Indeed they have concentrated too much on alternative accounts of the *vehicles* of content and not enough on developing successor proposals about the nature of *content* itself. This may not be obvious given Paul Churchland's advertisements to the contrary. For example, even in his most recent work on this topic, his article "Conceptual Similarity across Sensory and Neural Diversity", he leans on a very familiar two-factor theory of content of the kind discussed and dismissed in chapter two, section 2.2.1. According to his account representational content is a conglomerate of information, provided by the sensory modalities (the concept-to-world connection), and top down training, provided by the regimes imposed on connectionist networks (the concept-to-concept connections) (cf. P.M. Churchland 1998: 8, 12, 23, 25, 30). He is keen to suggest that it is sufficient for conceptual identity that two networks share the same geometrically discernible internal state-space such that the 'prototype trajectory families' of these networks, which link them informationally to the word, match perfectly. Alive to the likely critique that this story is undertold, he rhetorically asks, "...what makes the preceding story of activation patterns, structured spaces, and coincident trajectories a *semantic* theory? Why is it not just a new and unorthodox account of

our underlying 'syntactic' activities?" (P.M. Churchland 1998: 31). Good question. His answer is that it is not just a syntactic one on the grounds that:

> ...it gives us an account of how distinct representational contents are systematically embodied in distinct cognitive vehicles. It gives us detailed account of how those contents are causally related to propriety aspects of the external world, both in perception and action. It shows how certain aspects of the world's structure are mirrored in the relational structure of our assembled representations; that is, it gives us some idea how our background conceptual framework and specific activations within its picture both the global structure of the world and specific things and events within it....it looks like a semantic theory to me (P.M. Churchland 1998: 31).

But all this *talk* of representation simply begs the question. Churchland is not entitled to this upgrade in language, unless a theory is supplied to explain: (a) how the senses carry informational content, (b) how such content is processed, (c) how, and by what principle, it becomes representational content, (d) how, and by what principle, it interacts with representational contents that develop in top-down fashion from training, and (e) how the trained responses, themselves, become genuinely representational. It has been the burden of the majority of the chapters in this book to show that the current naturalistic theories of representational content fail on all these fronts. Since Churchland offers us nothing new in this domain the right verdict appears to be that, although connectionism offers a new means of understanding the mechanics of nonconceptual cognition, it fails to provide an alternative theory of representational content.

This would explain why many supporters of connectionism retain a traditional vision of symbolic representation. Consider Smolensky's account, according to which we are told that the connectionist paradigm retains the notion of symbol central to the classical cognitivist conception of representation.[80] As claimed in chapter one, section 3.2, the real difference between classical cognitivist and connectionist accounts concerns their theories of how content is processed. According to the traditional story the symbols themselves are the entities involved in the computational transformations, whereas the connectionist argues that computation occurs sub-symbolically (cf. Smolensky 1995: 33-34). The classical cognitivist-connectionist debate is about the nature of processing – not the nature of content. For this reason, both classical cognitivists and connectionists alike must face up to the problem of naturalising representational content. Therefore as long as eliminativists make use of the notion of representation, and rely on connectionism as their champion theory, they must explain how it is that aggregates of units manage to represent. In this context, Cussins remarks that:

> The Churchlands and Paul Smolensky (1988) amongst others have explored accounts that appeal to representational vehicles such as connectionist vectors and vector-spaces, gradient descent through weight/error space and partitions of activation-vector space. And Gareth Evans (1982), Christopher Peacocke (1989, 1992a, 1992b) and Adrian Cussins (1990, 1992b) amongst others have explored accounts of nonconceptual contents which are experiential modes of presentation whose structure is dependent upon how they are embodied in animals and embedded in the physical and social environment (Cussins 1993: 241).

He concludes from this that, "If eliminativism is ultimately to withstand the self-defeating charges....then it must combine theories of non-sentential vehicles with theories of non-conceptual contents..." (Cussins 1993: 241). Such a move would fit with a number of Churchland's remarks about his aspirations. For example, he tells us that:

> Truth, as currently conceived, might cease to be an aim of science. Not because we have lowered our sights or reduced our epistemic standards....but because we had raised our sights, in pursuit of some epistemic goal more worthy than truth (P. M. Churchland 1989: 150).

In a similar vein, when glorifying the alleged theoretical capacities of non-verbal animals Churchland, echoing an earlier remark, tells us that, "...language use appears as an extremely peripheral activity, as a biologically idiosyncratic mode of social interaction" (P.M. Churchland 1989: 16). In light of this agenda, Cussins' diagnosis appears to be the right one. Nonetheless his proposed strategy for rescuing eliminativism is flawed. He offers a poisoned chalice. In order for their account to work, the eliminativists cannot lean on a naturalised account of nonconceptual content. For whether the nonconceptualist looks to the 'realm of embodiment' or a modest biosemantics for semantic succour, it lacks the resources to explain theorising – scientific or otherwise. For example, as we have seen in chapter four, section 2.2.2, neither form is sufficient to explain reference or truth. Claims with conceptual content, of the sort appropriate to theorising, require a standard against which we can assess their truth value and determine what objects they denote. They are made and developed in the 'space of reasons' and are open to critique, judgement and debate. More conservative scientific naturalists and philosophers of science are often explicit about this commitment. For example, Boyd, Gasper and Trout write:

> We must acknowledge that the definitions of scientific concepts and the terms in scientific language are theory-dependent; any adequate account of the

> semantics of scientific terms must reflect this fact and must – whatever its author's ultimate position on semantic matters – reflect some of the insights of causal or naturalistic conceptions of definition and of reference (Boyd, Gasper and Trout 1991: xii, cf. Boyd 1991: 202).

Furthermore connectionist theories – far from providing revolutionary reasons for revising our ordinary criteria for recognising of the presence of minds – provide us with some vivid examples of why our ordinary criteria are indispensable, even if occasionally misapplied. Consequently even if the Churchlands are right that our ability to sponsor a conceptual understanding of the world is only made possible, in part, by the plasticity of our neural networks it does not follow that thinking or conceiving can be completely understood in neuro-computational terms. This fits with the fact that without appeal to a wider intersubjective standard we would have no principled grounds upon which to interpret the nonconceptual responses of networks (Clark 1996: 90-91). Thus even the eliminativists must accept that the crucial factors in the development of concepts are not purely neuro-computational. For this reason, they cannot sensibly advance claim (A).

4.4. Minding our Language

Our conception of reality is made possible, in part, by our common responses to the world and to each other. Thus it depends upon factors not discernible to neuroscience. There is no intelligible, wholly neuro-computational or mechanistic conception of the 'Real'. A conception of reality is something which emerges in the processes of intersubjective interpretation. For this reason language (or its basis in interpretative practice) cannot be regarded as a 'peripheral medium of representation' for human cognition. If the arguments of chapter five are in order then, far from being such, Gadamer is correct to say that:

> Language is the real medium of the human being....the realm of common understanding of ever-replenished common agreement – a realm as indispensable to human life as the air we breathe (Gadamer 1966: 68).

Furthermore, as Charles Taylor says:

> In talking about something you and I make it an object for us together, that is, not just an object for me which happens to be one for you, and you know, etc. The object is for us in a strong sense, which I have tried to describe elsewhere with the notion of 'public' or 'common' space. The various uses of language set up, institute, focus, or activate such common spaces....This is

the truth behind Wittgenstein's dictum that agreement in meanings involves agreement in judgements (Taylor 1989: 35).

These considerations demand something stronger than claim (B). They encourage support of claim (B').

(B') The nature of conception can be (in part) understood in neuro-computational terms and (in part) by appeal to socio-cultural factors.

Read as a claim emphasising the importance of understanding mechano-computational aspects of the workings of our minds, claim B' is not only harmless but welcome. It is harmless provided we recognise that the socially constituted aspects of content are not reducible. In essence, claim (B') leads us to believe that the contentful aspects of cognition come from without and are sustained by human society and culture. But, as it stands, claim (B') is repellent to the exclusive scientific naturalist. Consequently even though eliminativists sometimes write as if they realise that claim (B') is the only plausible claim to make about human cognition, they nevertheless want us to be open to the possibility that one day we will fully explain conceptualisation without appeal to culture and language *per se*. Thus whereas the defender of a non-reductive propositional attitude psychology ought to endorse claim (B'), the exclusive scientific naturalist must append the highly contentious (B'_{ii}) to claim (B').

(B'_{ii}) It is possible to fully explain the culturally-constituted character of content in natural scientific terms (cf. P.M. Churchland 1989: 134).

If Churchland did not support (B'_{ii}) his eliminativism would amount to nothing more than an attack on expansive classical cognitivism. In this context, he has admitted that claim (B'):

> ...may sound like a major concession to the view that the most important factors in the explanation of human behaviour are found at the cultural level rather than at the microphysical level. But to say this would be to miss the point, and to acquiesce in a false dichotomy between 'cultural' and 'physical' features (P.M. Churchland 1989: 133-134).

All the same, we are not offered any interesting analysis to show how cultural features are in fact physical features. We are merely promised that they are not, "...forever beyond a naturalistic understanding" (P.M. Churchland 1989: 134). However without a clear definition of the boundaries of such an understanding, or an account of how normative socio-cultural features reduce to the physical, it cuts little ice to say the dichotomy between the cultural and the physical is a false one.

5. Conclusion

In conclusion, we would do well to join with Putnam in saying, "I want to defend the view that there are whole domains of fact with respect to which present-day science tells us nothing at all" (Putnam 1990: 143). In this, we might even be encouraged to speak of human contexts of inquiry to help ward off reductionism and to remind ourselves that the sciences are dependent upon human culture and society (cf. McDonough 1991: 272, O'Hear 1991: 56). Having said this, we need not react against science or deny it a prominent place in our thinking. It simply cannot lead or restrict philosophical inquiry on certain matters. Relative to a particular context of inquiry, we can expect physics, the biological sciences and even cognitive science to play important and useful roles. But they do not succeed in giving us an unbiased or complete account of human nature. We must have the presence of mind to recognise what appears almost too trivial to mention: that human society, culture and language are the starting points of all debate and inquiry.[81]

Notes

Introduction

1 Lewis notes that, "These names, of course, may purport to name individuals, sets, attributes, species, states, functions, relations, magnitudes, phenomena or what have you; but they are still names" (Lewis 1972: 206).

Chapter One

2 In everyday speech we do not usually cite both a belief and a desire when we give a reason for our action. For instance, I might answer the question "Why did you eat that?" simply by saying that "I like chocolate cake" implying, but not stating, that I also believed what I ate was a piece of chocolate cake. Conversely, had I answered the question by saying "I thought it was chocolate cake" I would be implying that I desired to eat a piece of cake. Davidson notes that for reasons of accuracy it might be best to speak of pro-attitudes since any number of psychological verbs might replace 'desire' as the conative or motivational element of the folk psychological pair.

3 Other names include: belief/desire psychology and propositional attitude psychology.

4 An imaginary case of this procedure is nicely illustrated in narrative detail by Dennett when he discusses how we might relieve a group of near-English-speaking aliens of their notion of fatigues (Dennett 1985: xix - xx).

5 It is worth noting that although 'propositional attitude psychology' may occupy quite a central place in our garden variety understanding of the mind, it is not exhaustive. For instance, it leaves aside the fact that we are agents which see, hear, touch, taste and generally experience aspects of our world in a conscious fashion.

6 This Sellarsian view seems to agree – in broad outline – with Wittgensteinian views about language. It is not that we have types of raw experiences and give them names; but rather that our view on 'what there is' is much more interwoven with the fabric of our language and practices. As Schacht writes of Nietzsche's perspectives which resemble Wittgenstein's language games, "They are not to be construed as mere

vocabularies people somehow acquire and with which they articulate their experiences of pre-linguistically determinate objects and events and states of affairs of various kinds. Rather, they are 'forms of life' in the sense of spheres of human experiences and activity in which certain kinds of objects and states of affairs are fixed and differentiated in accordance with various sorts of linguistic apparatus. What counts as an object, a difference between objects, and a relation between them, is determined by the concepts and rules of particular schemes of this sort, and has no standing or meaning independently of them." (Schacht 1983: 63).

7 Using Kuhnian phraseology most cognitive scientists speak of the 'revolution' in terms of a change from the 'symbolic paradigm' (see Bechtel and Abrahamsen 1991: 1-3).

8 In his paper "Propositional Attitudes" he details a number of features which he holds act as constraints on any adequate theory of such psychology.

9 In Aristotle's terms thought involves the affirmation or negation of some proposition, while desire is concerned with the pursuit or avoidance of some object. What becomes clear, on this account, is that, "Intellect itself moves nothing...hence choice is either desiderative thought or intellectual desire" (*Nicomachean Ethics:* 1139^a 35-36, 1139^b 4-5). He described the processes that lead to action in the following way, "I want to drink, says appetite; this is drink, says sense or imagination or thought: straightaway I drink. In this way living creatures are impelled to move and to act, and desire is the last cause of movement, and desire arises through perception or through imagination and thought" (Aristotle *Movements of Animals*: 701^a 32-36, cf. also Gill 1991: 175).

10 Fodor is interested in defending commonsense psychology because he is impressed by (i) how often it works, (ii) the depth of the theory, and (iii) its indispensability (cf. Fodor 1987: ch. 1).

Chapter Two

11 Fodor denies that Stich's principle of autonomy is identical to his formality condition (cf. Fodor 1987: 42); but, ultimately, the problem for the supervenience of contentful properties on internal states emerges on either formulation.

12 This diagram is adapted from Fodor's paper on the special sciences (Fodor 1981c: 139) and his contribution to Blackwell's *A Companion to the Philosophy of Mind*. (Fodor 1994: 297)

13 Other philosophers have defended the idea that commonsense psychology is committed to some kind of narrow content (cf. Devitt 1990a, Jackson & Pettit 1993a).

14 This later view is marginally qualified. Fodor allows that contents are 'linked' by various logical connectives. Nevertheless he supports atomism about the non-logical operators.

15 Some philosophers believe this move is unnecessary since they deny that the behaviour is the same in any case. What is the same is merely their bodily movements (Adams *et. al* 1990: 224).

16 As Stich says, "Narrow content provides a third strategy for classifying brain state tokens, one which starts with broad contents but ignores context....narrow content supervenes on physiology" (Stich 1991: 248).

17 McGinn has questioned this logic with respect to theories of consciousness. On his view the fact that terms like causation, relation, and identity are topic-neutral guarantee that they can be used both 'naturally' and 'non-naturally' (cf. McGinn 1991: 57-58). If McGinn is right appealing to causality in the mental domain only links an unnatural kind with a natural kind. If this is correct we cannot take it for granted that causality is the right tool for the naturalist.

18 Quine, a die-hard behaviourist, advanced this view with an epistemological twist and took an eliminativist stance towards meanings precisely because he believed that there was no way to account for semantic content in terms of proximate stimuli. 'Stimulus meanings' are all that are acceptable to him but 'stimulus meanings' are nothing like ordinary meanings (Quine 1960: 1969). This is why Levin claims Quine makes the, "...error of supposing that terms refer to proximate stimuli" (Levin 1984: 136).

19 The misrepresentation problem is a metaphysical variant of Quine's problem of the indeterminacy of translation. Quine's arguments in favour of scepticism about meaning have had a powerful influence on naturalist philosophers. For example, Stich says, "...influenced by Quine, I have been long suspicious about the integrity and scientific utility of the commonsense notions of meaning and intentional content" (Stich 1991: 240). In a way similar to Quine's examples, the misrepresentation problem reveals that there is no non-questioning begging way to say which one, of indefinitely many possible mapping functions, is the correct one by appeal to the facts of mere causation alone (Stich 1990d: 23).

20 Ironically, a number of philosophers of science have been attracted to this kind of approach. Consider these two quotations. Newton Smith writes, "What unites the users is the intention to use the term to refer to whatever magnitude it is that is responsible for certain effects" (Newton–Smith 1981: 164). And Boyd remarks, "...a pair of terms will be coreferential just in case the social history of each of their occurrences links them, by the *relevant sort of causal relations*, to a situation of reliable belief regulation

by the actual properties of the same feature of the world" (Boyd 1991: 210, emphasis mine).

21 This result may not ultimately bother Fodor for as Dennett says, "If he has to choose between Granny [commonsense psychology] and science, he has made it clear he'll choose Granny..." (Dennett 1991b: 88). As Dennett sees it, Fodor, "...disagrees with those of us, from Sellars to the present, who in one way or another take 'functional role semantics' to be the chief beauty of functionalism" (Dennett 1991b: 90).

22 Dretske deviates from other theorists in this domain by endorsing the 'Xerox principle' about the flow of information (cf. Dretske 1981: 57). According to this principle, "...if each link in the chain carries the appropriate piece of information about its predecessor....then the *final link* will carry the *same information* about the first link (Dretske 1981: 59).

23 Dretske draws a useful analogy between this process and the game of twenty questions in which the object is to isolate one unique piece of information by asking 'yes or no' questions which narrow down the range of possibilities. For example, the player asks questions like "Is x living or dead?", "Is x male or female?", and so on.

24 In reply to his critics Dretske admirably justifies his use of a thermostat as opposed to a computer as an example. He says, "I will stick to thermostats. It is easier for me to see exactly what is causing what and why. Furthermore, as a bonus, we avoid a lot of the jargon from the information sciences (e.g. computation, data structures, storing values for variables) that really obscure the essential issues..." (Dretske 1991: 204).

Chapter Three

25 To inflate their counterfactuals the propensity theorist could appeal to what actually will benefit the organism. But what actually benefits them will only become clear at some future time – after the fact, as it were. Thus, to fix the function in this way is surely not appealing to those who looked to the propensity account for assurance that what causes action has to do with the 'here and now' capacities of an organism (cf. Coates 1997: 195, footnote 19).

26 With respect to the source of systematic reproduction she makes reference to the notion of natural laws operative *in situ* – i.e. those that need to be understood, "...by adding reference to the actual surrounding conditions" (Millikan 1984: 20).

27 She borrows the term 'icon' from Peirce because it does not carry with it the same legacy of confusion and disagreement as the term 'representation'.

28 Millikan was initially wont to speak of 'producer and interpreter devices' in *Language, Thought and Other Biological Categories*. But given that she explicitly did not require that the interpreter, "...understand *what* the sign signs", the term 'consumer' is much more appropriate (Millikan 1984: 96).

29 Millikan writer, "Roughly indeed, the idea is this. Cognitive systems are designed by evolution to make abstract pictures of the organism's environment and to be guided by these pictures in the production of appropriate actions" (Millikan 1993: 11, cf. Rowlands 1997: 283).

30 She says, "..in the frog case, [the crucial initial relation is] that the firing of the detector is coincident with the presence of an edible bug" (Millikan 1993: 127). She has expressed the same view to me in the following terms, "Connecting with something black-and-a-dot is no part of any proximate normal explanation of why it helped its ancestors survive. Neither the blackness nor the dotness helped in any way, neither need be mentioned. But the nutritious object was essential" (Millikan 1996: personal correspondence, cf. also Elder 1998: 355).

31 Interestingly Godfrey-Smith notes that, "Although it is not always appreciated, the distinction between function and malfunction can be made within Cummins' framework....If a token of a component is not able to do whatever it is that other tokens do, that plays a distinguished role in the explanation of the capacities of the broader system, then that token is malfunctional." (Godfrey-Smith 1993: 200).

32 To be fair Neander both recognises the problem and states her intention to address it.

33 Godfrey-Smith usefully outlines the full spectrum of views from the pessimistic to the optimistic (Godfrey-Smith 1994b).

34 Papineau discusses the notion of truth more fully in his *Philosophical Naturalism*. Therein he tells us that he is attracted to a redundancy theory of truth which is backed up by a 'substantial theory of content' (Papineau 1993: 85). Unsurprisingly, he tells us that, "The substantial theory of content I favour is in terms of success conditions and biological purposes" (Papineau 1993: 85).

35 It is in this regard that Millikan responds to the rhetorical question, "Is it really plausible that bacteria and paramecia, or even birds and bees, have inner representations in the same sense that we do? Am I really prepared to say that these creatures, too, have mental states, that they think? I am not prepared to say that" (Millikan 1989a: 294). Dretske makes a similar point when he writes, "To qualify as a belief a representational

content must *also* exhibit (amongst other things) the familiar opacity characteristic of the propositional attitudes..." (Dretske 1986a: 27, emphasis mine).

36 This conclusion obviously contrasts with the presumptions of those who base their teleofunctionalism on an information-theoretic account of indicator functions. Against the objection that such systems may 'carry information' nonetheless Akins concedes the point but rightly suggests that it doesn't follow that the system has any means of 'extracting' this information. Hence, she says, "The question that concerns us here is whether, given that information about the stimulus is often carried in the sensory signal, this will be of any practical use in constructing a theory of aboutness." (Akins 1996: 357).

37 Thus Rowlands is on safer ground when he suggests that, "One can, therefore, speak of the mechanism detecting flies, or enabling the frog to detect flies, but this is only in a secondary and derivative sense, and reflects neither the algorithmic nor the organismic proper function of the mechanism" (Rowlands 1997: 295).

38 Millikan tells us that it is, "...specifiable correspondence rules that give the semantics for the relevant system of representation....[and that a] certain correspondence between the representation and the world is what is normal" (Millikan 1989a: 287). She also boldly says, "I take myself to be defending the strongest possible kind of correspondence theory of truth and the most flat-footed interpretation of the truth-conditions approach to semantics" (Millikan 1993: 212).

Chapter Four

39 We can see the trap closing if we consider Armstrong's analysis. Taking this point on board he states that, "If the dog believes, then what makes his belief a belief is something which pertains to the dog, and has nothing to do with humans and their speech" (Armstrong 1973: 28). He also rightly maintains that to be counted as a believer, of any sort, one must be counted as a conceiver. Thus, he remarks, "Notice that whatever we say of beliefs must also be said of the concepts which are involved in those beliefs. So animals must have concepts as well as beliefs" (Armstrong 1973: 31). There are other extended, more recent discussions of this point (see Graham 1993: ch. 4, Heil 1992: ch. 6).

40 I explore this territory in *Beyond Physicalism* (Hutto forthcoming-b).

41 Elsewhere he says, "Conceptual and nonconceptual content are both forms of content because there is a single notion of representation applicable to both" (Bermúdez 1995: 346).

42 Someone like Dennett might object to such talk on the grounds that all intentionality must be derived. However, as Hornsby says, "Of course our intentionality might be supposed to 'derive' from our genes in some bland sense, not requiring that genes should be carrying out their own purposes" (Hornsby 1992: 169-170). She rightly notes that this 'derivation' is nothing like the derived intentionality enjoyed by human artifacts (i.e. robots, etc.). Thus when Dennett proposes arguments to the contrary he is, in effect, "...playing fast and loose with the notion of derivation" (Hornsby 1992: 170).

43 In taking this line, I am in part explicitly following up and extending Bermudéz's suggestion that that, "...subpersonal information states lend themselves to a teleological theory of content...[and that]...Correctness conditions are fixed with reference to evolutionary design and past performance" (Bermúdez 1995: 365, 366). Although this may be correct, as Cussins noted, nonconceptual content need not apply only to sub-personal states and systems ('Nonconceptual Content Conference', University of Hertfordshire, 1st November 1997). Nevertheless the biological proposal is also in the spirit of Cussins' general claim that, "Natural selection has evolved creatures, by means of gradual and continuous changes, which are capable of a mind/world distinction..." (Cussins 1990: 409).

44 This objection has been made by a number of authors in various ways. For example, many see eliminativists confusing commonsense psychology with 'syntactical computationalism' (Bogdan 1993), or as reading it in an opulent fashion which includes commitment to psychotechtonic claims as opposed to reading it in an austere fashion (Horgan 1993).

45 In this context, it is useful to recall Nelson Goodman's wonderfully satirical remark that, "...he saw no more reason to suppose that infants had the rules of grammar inscribed in their brains than the stones thrown in the air had, somewhere inscribed in them, Newton's laws of motion." (Lockwood 1989: 121).

46 The view that behaviour proper is distinct from the notion of 'mere bodily movement' may be currently unorthodox but it is not unprecedented. Malclom advanced it long ago by suggesting that, "Purposive explanations explain actions. Neurophysiological movements explain movements. Both explain behaviour: but we can only say this because we use the latter word ambiguously covers both actions and movements" (Malcolm 1968: 51-52).

47 For similar reasons Horgan claims that, "If one is a committed materialist like Dretske, then 'splitting the explanadum' is the wrong strategy for dealing with this [explanatory exclusion] problem" (Horgan 1991: 83).

48 There is a parallel here with *de re* beliefs for as Burge remarks, "In attributing a *de re* belief about a given object, one is free to substitute any correct description of the relevant object" (Burge 1977: 341).

49 Davidson has written to me on this matter saying, "I don't think the issue whether animals have beliefs is in itself of any importance – one can use words as one pleases. But if you want to talk about pre-linguistic thought, you need to explain precisely what you have in mind" (Davidson, 10th November 1991, personal communication).

Chapter Five

50 Davidson tells us, "Communication depends, then, on each communicant having, and correctly thinking the other has, the concept of a shared world. But the concept of an inter-subjective world is the concept of an objective world, a world about which the communicant can have beliefs" (Davidson 1985c: 480, Davidson 1984a: 170, Davidson 1990b: 314).

51 Davidson describes such a picture when he writes, "...one line goes from the child in the direction of the table, one line goes from us in the direction of the table, and the third line goes from us to the child. Where the lines from the child to the table and us to the table converge 'the' stimulus is located. Given our view of the child and world, we can pick out 'the' cause of the child's responses." (Davidson 1992: 263).

52 I take this to be what he means in saying, "...the world is disclosed essentially along with the Being of Dasein" (Heidegger 1962 §43, §203). Or as Gadamer puts it, "The agreement about things that takes place in language means neither a priority of things nor a priority of the human mind" (Gadamer 1960: 78).

53 The diagram is designed to show that the problem is not about possible memory failure. For 'remembering' the stimulus is not the issue – but remembering it as 'the same' as before is.

54 The Fodor-Chomsky proposal of an innate *lingua mentis* is similar to the Augustinian account of language acquisition which Wittgenstein criticises early on in the *Investigations*. He writes, "Augustine describes the learning of human language as if the child came into a strange country and did not understand the language of the country; that is, as if it had already had a language, only not this one. Or again: as if the child could already *think*, only not yet speak" (Wittgenstein 1953: §32, 1958: 41). And as he writes elsewhere, "We often think as if our thinking were founded on a thought-schema; as if we were translating from a more primitive mode of thought into ours" (Wittgenstein 1953: §597).

55 Consider that, in a similar vein, McDowell takes issue with Evans' claim that, "...conceptual capacities are first brought into operation only when one makes a judgement of experience, and at that point a different species of content comes into play" (McDowell 1994: 48). In contrast he proposes that, "A judgement of experience does not introduce a new kind of content, but simply endorses the conceptual content, or some of it, that is already possessed by the experience on which it is grounded" (McDowell 1994: 48–49).

56 All we are told is that, "...acquiring a concept is getting nomologically locked to the property that the concept expresses" (Fodor 1998: 125).

57 Chomsky would say they are, "...part of our biological endowment" (Chomsky 1988: 4). But as Putnam notes, "To have given us an innate stock of notions which includes carburettor, bureaucrat, quantum potential, etc. as required by Fodor's version of the Innateness Hypothesis, evolution would have had to be able to anticipate all the contingencies of future physical and cultural environments. Obviously it didn't and couldn't do this" (Putnam 1988: 15).

58 Clark also speculates on the possibility that structured, serial thought might be implemented as a virtual machine on a connectionist architecture. He writes, "...our basic architecture would remain a PDP system, though cunningly configured to make possible certain kinds of sequential, symbolic thought" (Clark 1989: 134).

59 Not all social communication amongst animals is geared toward merely indicating the presence of danger. Many animals signal the presence of predators, but they also signal for purposes of identification (i.e. of groups, kin, pair bonds, and individuals), to mimic more dangerous animals, to entrap other species (e.g. in the case of angler fish and their lures), to show aggression, to show submission, to indicate strengths to predators (e.g. the strutting of gazelles) and to deceive predators in order to protect offspring (grey partridges and the broken wing pretence) (cf. Krebs and Davies 1991: 374-397).

60 For example, if the metalanguage were English, p would be 'Roses are red'. This relates nicely to Wittgenstein's remark that, "The limit of language is shown by its being impossible to describe a fact which corresponds to (is the translation of) a sentence without simply repeating the sentence (This has to do with the Kantian solution the problem of philosophy)" (Wittgenstein 1980a: 10e).

61 He writes, "Truth is one of the clearest and most basic concepts we have, so it is fruitless to dream of eliminating it in favour of something simpler or more fundamental" (Davidson 1990b: 314).

62 This is one of the reasons why Quine is an eliminativist concerning meaning. Due to his commitment to behaviouristic 'stimulus meanings' Quine's approach does not have

the conceptual tools to distinguish between two extensionally equivalent T-theories (cf. Evnine 1991: 98-99, 122-124, Ramberg 1989: 122-124).

63 He tells us, "...we no longer need to explain meaning on the basis of possible confrontation" (Davidson 1986a: 307). Instead, he has adopted the slogan, "correspondence without confrontation" (Davidson 1986a: 307). He says, "...it is not the *speaker* who must perform the impossible feat of comparing his belief with reality [i.e., what caused it]; it is the *interpreter* who must take into account the causal interaction between the world and speaker to find out what the speaker means and hence what he believes" (Davidson 1986b: 332). Also, he writes, "My approach is...externalist: I suggest that interpretation depends (in the simplest and most basic situations) on the external objects and events salient to both the speaker and the interpreter, the very objects and events the speaker's words are then taken by the interpreter to have as a subject matter. It is the distal stimulus that matters to interpretation" (Davidson 1990b: 321).

64 Davidson anticipates the criticism that this method of interpretation will be culturally idiosyncratic. Against philosophers who would argue that the analysis of the meaning of our sentences is, "...distinctively Western, and inapplicable to members of other cultures" (Evnine 1991: 146) he says that, "...we have no choice....but to project our logic on to the language and beliefs of another" (Davidson 1990b: 321). It is in light of this that he says; "Given the underlying methodology of interpretation, we could not be in a position to judge that others had concepts or beliefs radically different from our own" (Davidson 1984a: 197). Thus he believes the project of radical interpretation, "...has profound consequences for the relation between thought and meaning, and for our view of truth, for it not only ensures that there is a *ground level* on which speakers share views, but also that what they share is a largely correct picture of a common world" (Davidson 1990b: 325, emphasis mine).

65 Such a principle lays the ground for the interpreter's more difficult task of, "...abstract[ing] from the evidence [those sentences held true by the speaker under specific circumstances] a workable theory of meaning and an acceptable theory of belief" (Davidson 1984a: 196). For it is from these humble beginnings that we are able to work up to, "...the fine distinctions between beliefs [which] are impossible without understood speech" (Davidson 1984a: 195).

Chapter Six

66 This is consistent with Davidson's admission that, "It comes closer to characterising first person authority to note that the self-attributer does not normally base his claims on evidence or observation..." (Davidson 1984b: 103). However, he is also aware that,

"...it is long out of fashion to explain self-knowledge on the basis of introspection" (Davidson 1984b: 103). He has done much to obviate the need to suppose the existence of special relations between agents and posited internal events by using features of radical interpretation in order to explain the source of first-person authority in the case of propositional-attitude ascriptions.

67 I discuss the problems with this kind of metaphysics in (Hutto 1995b: 465-466) and in *Beyond Physicalism* (forthcoming).

68 "If we were to drop the normative aspect from psychological explanations, they would no longer serve the purposes they do...Physics, on the other hand, has as an aim of developing laws that are as complete and precise as we can make them; a different aim" (Davidson 1991a: 163). And, this is why Davidson says, "When we try to understand the world as physicists, we necessarily employ our own norms, but we do not aim to discover rationality in the phenomena" (Davidson 1991a: 162). This is why Davidson also rightly holds that, "...we cannot turn [psychological] explanation into something more like science." (Davidson 1980a: 233, 1987b: 42-43).

69 The special features of psychological interpretation led Wittgenstein to treat reason-explanations as 'further descriptions' or 'elucidations' rather than causal hypotheses (cf. Coiffi 1991: 172-173). His view on this matter is valuable because it reminds us that we do not appeal to the methods of sciences like physiology or neurology when deciding why it is that someone acts.

70 Given this, other than noting a change in the size and colour of Routledge monographs, Lepore and Loewer are premature in their prediction of the death of red-book philosophy.

71 I use the label 'mental' for ease of exposition but it is a use of which I am generally suspicious (cf. Hutto 1995b).

72 To use J.J.C. Smart's words, the token identity theorist is, "...using 'is' in the sense of strict identity....there is one thing" (Davidson 1980a: 212).

73 Davidson defines homonomic and heteronomic as follows: a homonomic generalization is one which could be improved, with provisos, within the same general vocabulary while a heteronomic generalization is one with an ineliminable *ad hoc* character – and for which the precise law must be stated in a different vocabulary (Davidson 1980a: 219).

74 This point applies equally to global supervenience which has also emerged as a popular means of explicating the supervenience relation (cf. Petrie 1987, Paull & Sider 1992).

For a critical assessment of it on other grounds see Kim (1987), Heil (1992) and Moser (1992).

75 This criticism has been raised by several philosophers (Honderich 1982-84, Kim 1984, 1993, Sosa 1984, 1993, Johnston 1985, Fodor 1990b, Evnine 1991, McLaughlin 1993).

76 At one time Evnine noted that, "Davidson nowhere gives any argument for...[the PNCC]" (Evnine 1991: 37, 63-67). To which he added, "[since Davidson]...has disavowed any intention of analyzing causation as such...it is therefore a legitimate question to ask why he upholds [the PNCC]" (Evnine 1991: 38). Before Evnine's remarks Davidson had openly acknowledged the worry and had written, "...one would like to be given a reason for supposing there is a close connection between laws and causal explanations. I think there is such a connection, and I plan to argue for it in print one of these days" (Davidson 1985b: 247). His paper "Laws and Cause" is his first serious attempt to do so in print.

77 He also writes, "Our concept of a physical object is the concept of an object whose changes are governed by laws" (Davidson 1995: 11). We might wish to accept this in most cases of macro-'physical' causation, since the methods for investigating those types of causal relations will have persuaded us of the truth of such a view (provided we don't go too far in the quantum direction).

78 In this light it is ironic that Ricoeur offers the Aristotelian position, which gives persons a privileged place in action-theory, as a viable alternative to Davidson's approach. He writes, "Aristotle, well before the Stoics, made it apparent that action depends on the agent, in a specific sense of the relation of dependence....How, on this base is the relation of action to agent to be stated? The most concise expression of this relation is found in a formula that makes the agent the principle (*arkhé*) of his actions, but in a sense of *arché* that authorises us to say that the actions depend on (preposition *epi*) the agent himself (*autó*) (E.N. 3.1110a17)" (Ricoeur 1992: 89-90).

Chapter Seven

79 Loewer and Rey criticise Dennett along similar lines. They say, "An interpreter's taking a stance and attributing an attitude from that stance....after all, [involves the having of] intentional attitudes" (Loewer and Rey 1991:. xxi).

80 Or rather the paradigm which he feels would constitute the 'proper treatment of connectionism' (cf. Smolensky 1995: 30).

81 As Davidson writes, "A community of minds is the basis of knowledge; it provides the measure of all things. It makes no sense to question the measure of that standard, or to seek a more ultimate one." (Davidson 1991a: 164). Or as Wittgenstein stressed, "What has to be accepted, the given is – so one could say *–forms of life*" (Wittgenstein 1953: 226).

References

Adams, F. 1991. "Causal Contents." In B. McLaughlin (ed.), *Dretske and His Critics*. Oxford: Blackwell, 131–156.

Adams, F., D. Drebushenko, G. Fuller, and R. Stecker. 1990. "Narrow Content: Fodor's Folly." *Mind and Language* 5 (3): 213–229.

Akins, K. 1996. "Of Sensory Systems and the 'Aboutness' of Mental States." *Journal of Philosophy* XCIII (7)

Anscombe, G.E.M. 1981. "Causality and Determination." *Metaphysics and the Philosophy of Mind. Collected Philosophical Papers Volume II.* Oxford: Blackwell, 133–147.

Anthony, L. and J. Levin 1991. "The Nomic and the Robust" In B. Loewer and G. Rey (eds.), *Meaning in Mind: Fodor and His Critics*. London: Blackwell, 1–16.

Appel, K.O. 1991. "Is Intentionality More Basic than Linguistic Meaning?" In E. Lepore and R. Van Gulick (eds), *John Searle and His Critics.* Oxford: Blackwell, 31–58.

Armstrong, D.M. 1973. *Belief, Truth and Knowledge*. Cambridge: Cambridge University Press.

Armstrong, D.M. 1979."Perception, Sense Data and Causality." In G.F. MacDonald (ed.), *Perception and Identity: Essays Presented to A.J. Ayer with his Replies to Them*. London: Macmillian Press Ltd, 84–98.

Armstrong, D.M. 1981."The Causal Theory of Mind" In W. Lycan (ed.), 1990. *Mind and Cognition* Oxford: Blackwell, 37–47.

Armstrong, D.M. 1991."Intentionality, Perception, and Causality: Reflections on Searle's *Intentionality*" *John Searle and His Critics*. Oxford: Blackwell, 149–158.

Ashington, J. 1996."What is Theoretical about a Child's Theory of Mind?: A Vygotskian View of its Development." In P. Carruthers & P. Smith (eds.), *Theories of Theories of Mind.* Cambridge: Cambridge Universitiy Press, 184–199.

Bach, K. 1982."De Re Belief and Methodological Solipsism" In A. Woodfield (ed), *Thought and Object.* Oxford: Clarendon Press, 121–151.

Baker, L.R. 1982. "De Re Belief in Action." *The Philosophical Review* XCI (3): 363–387.

Baker, L.R. 1988. *Saving Belief: A Critique of Physicalism.* New Jersey: Princeton University Press.

Baker, L.R. 1991. "Has Content Been Naturalized?" In B. Loewer and G. Rey (eds.), *Meaning in Mind: Fodor and His Critics*. London: Blackwell, 17–32,

Baker, L.R. 1993. "Eliminativism and an Argument from Science." *Mind & Language*. 8 (2): 180–187.

Barnes, J. (ed.) 1984. *The Complete Works of Aristotle: The Revised Oxford Translation*. Vol. I & II. New Jersey: Princeton University Press.

Bechtel, W. 1988a. *Philosophy of Mind: An Overview for Cognitive Science*. New Jersey: Lawrence Erlbaum Associates.

Bechtel, W. 1988b. *Philosophy of Science: An Overview for Cognitive Science*. New Jersey: Lawrence Erlbaum Associates.

Bechtel, W. 1990. "Connectionism and the Philosophy of Mind: An Overview" In W. Lycan (ed), *Mind and Cognition*. Oxford: Blackwell, 252–273.

Bechtel, W. and Abrahamsen, A. 1991: *Connectionism and the Mind: An Introduction to Parallel Processing in Networks*. Oxford: Blackwell.

Bechtel, W. 1996. "What Should a Connectionist Philosophy of Science Look Like?" In R. McCauley (ed.), *The Churchlands and Their Critics*. Oxford: Blackwell, 121–144.

Bennett, J. 1991. "Folk Psychological Explanation" In J. Greenwood (ed.), *The Future of Folk Psychology*. Cambridge: Cambridge University Press, 176–195.

Bermudéz, J. 1994. "Peacocke's Argument Against the Autonomy of Nonconceptual Representational Content." *Mind and Language*. 9 (4): 402–417.

Bermudéz, J. 1995. "Nonconceptual Content: From Perceptual Experience to Subpersonal Computational States" *Mind and Language*. 10 (4): 333–369.

Bernier, P. 1993. "Narrow Content, Context of Thought and Asymmetric Dependency" *Mind and Language*. 8 (3). 327–341.

Bigelow, J. and R. Pargetter, 1987: "Functions" *Journal of Philosophy* 84: 181–196,

Biligrami, A. 1986. "Meaning, Holism and Use" In E. Lepore (ed.), *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*. Oxford: Basil Blackwell, 101–122,

Blackburn, S. 1991. "Losing Your Mind: Physics, Identity and Folk Burglar Protection" In J. Greenwood (ed.), *The Future of Folk Psychology*. Cambridge: Cambridge University Press, 196–225,

Block, N. (ed) 1981. *Readings in the Philosophy of Psychology: Vol. II*. London: Methuen.

Block, N. 1986a. "Psychologism and Behaviourism" *The Philosophical Review*, 95: 5–43,

Block, N. 1986b. "Advertisement for a Semantics for Psychology" Midwest Studies in Philosophy, X: 615–678.

Block, N. 1987. "Functional Role and Truth Conditions." *Proceedings of the Aristotelian Society*, Supplementary Volume. 61: 157–181.

Block, N. 1990. "An excerpt from *Troubles With Functionalism*" In W. Lycan (ed.), *Mind and Cognition* Oxford: Blackwell, 444–468.

Block, N. 1991. "What Narrow Content is Not" In B. Loewer and G. Rey (eds.), *Meaning in Mind: Fodor and His Critics*. London: Blackwell, 33–64.

Boden, M. (ed.) 1990. *The Philosophy of Artificial Intelligence*. Oxford: Oxford University Press.

Boden, M. (ed.) 1996. *The Philosophy of Artificial Life* Oxford: Oxford University Press.

Bogdan, R (ed.). 1986. *Belief.* Oxford: Clarendon Press.

Bogdan, R. 1993. "The Architectural Nonchalance of Commonsense Psychology" *Mind & Language* 8 (2): 189–205.

Boghossian, P.A. 1991. "Naturalizing Content" In B. Loewer and G. Rey (eds.), *Meaning in Mind: Fodor and His Critics*.London: Blackwell, 65–86.

Botterill, G. 1996. "Folk Psychology and Theoretical Status" In P. Carruthers & P. Smith (eds), *Theories of Theories of Mind* Cambridge: Cambridge Universitiy Press, 105–118,

Boyd, R.N. 1991. "Confirmation, Semantics and the Interpretation of Scientific Theories" In R. Boyd, P. Gasper, & J.D. Trout (eds.), *The Philosophy of Science*. Cambridge, MA: MIT Press, 3–35.

Boyd, R., P. Gasper & J.D. Trout (eds). 1991. *The Philosophy of Science.* Cambridge, MA: MIT Press.

Brandom, R. 1998. "Perception and Rational Constraint." *Philosophy and Phenomenological Research* 58 (2): 369–374.

Bratman, M. 1990. "Dretske's Desires" *Philosophy and Phenomenological Research* 50 (4): 795–800.

Brentano, F. 1874. "Psychology from an Empirical Standpoint" In G. Vesey (ed.), *Body and Mind.* 1964: London: George, Allen and Unwin, 144–154.

Brinck, I. 1997. *The Indexical 'I': The First Person in Thought and Language*. The Netherlands: Kluwer Academic Publishers.

Brown, C. 1993. "Belief States and Narrow Content" *Mind and Language* 8 (3): 343–367.

Buekens, F. 1997. "Supervenience, Identity and Causal Efficacy" (unpublished manuscript — read to the staff/postgraduate research seminar, University of Hertfordshire, 20th Feb. 1997).

Burge, T. 1977. "Belief De Re" *The Journal of Philosophy* LXXIV; 338-362.

Burge, T. 1979. "Individualism and the Mental" *Midwest Studies in Philosophy* 4: 73–121.

Burge, T. 1986a. "Individualism and Psychology" *Philosophical Review* XCV (1): 3–45.

Burge, T. 1986b. "Cartesian Error and the Objectivity of Perception" In P. Pettit and J. McDowell (eds.), *Subject, Thought and Context*. Oxford: Oxford University Press, 117–136.

Burge, T. 1989. "Individualism and Causation in Psychology" *The Pacific Philosophical Quarterly* 70: 303–322.

Burge, T. 1992. "Philosophy of Language and Mind: 1950 - 1990" *The Philosophical Review* 101 (1): 3–51.

Call, J. and M. Tomasello 1994. "Production and Comprehension of Referential Pointing by Orangutans (*Pongo pygmaeus*)" *Journal of Comparative Psychology* 108 (4): 307–317.

Call, J. and M. Tomasello 1996. "The Effect of Humans on the Cognitive Development of Apes." In A. E. Russon, K. A. Bard and S. T. Parker, (eds). *Reading into Thought*. New York: Cambridge University Press, 371–403.

Campbell, K. 1993. "What Motivates Eliminativism" *Mind & Language* 8 (2); 206–210.

Carruthers, P. 1996. "Simulation and Self-Knowledge: A Defence of Theory-Theory" In P. Carruthers & P. Smith (eds), *Theories of Theories of Mind* Cambridge: Cambridge University Press, 22–38.

Carruthers, P. 1998. *Language, Thought and Consciousness*. Cambridge: Cambridge University Press.

Carruthers, P. and P. Smith (eds.). 1996. *Theories of Theories of Mind* Cambridge: Cambridge University Press

Carruthers, P. and J. Boucher (eds.). 1998. *Language and Thought*. Cambridge: Cambridge University Press.

Charles, D. 1992. "Supervenience, Composition and Physicalism" In D. Charles and K. Lennon (eds.), *Reduction, Explanation and Realism*. Oxford: Oxford University Press, 265–296.

Chater, N. and C. Heyes 1994. "Animal Concepts: Content and Discontent" *Mind and Language* 9 (3): 209-246.

Child, W. 1994. *Causality, Interpretation and the Mind*. Oxford: Clarendon Press.

Chomsky, N. 1988. *Language and Problems of Knowledge: The Managua Lectures*. Cambridge, MA: MIT Press.

Chomsky, N. 1990. "On the Nature, Use and Acquisition of Language." In W. Lycan (ed.), *Mind and Cognition*. Oxford: Blackwell, 627–646.

Chrisley, R. 1993. "Connectionism, Cognitive Maps and the Development of Objectivity." *Artificial Intelligence Review* 7: 329–354.

Chrisley, R. 1995. "Taking Embodiment Seriously: Nonconceptual Content and Robotics" In K. Ford, Glymour, C. and Hayes, P. (eds). *Android Epistemology* Cambridge, MA: MIT Press, 141–166.

Churchland, P.S. 1980. "A Perspective on Mind-Brain Research" *The Journal of Philosophy* LXXVII (4): 185–207.

Churchland, P.S. 1986. *Neurophilosophy*. Cambridge, MA: MIT Press.

Churchland, P.S. and T. Sejnowski 1990. "Neural Representation and Neural Computation" In W. Lycan (ed.), *Mind and Cognition*. Oxford: Blackwell, 224–252.

Churchland, P.M. 1970. "The Logical Character of Action Explanations" *Philosophical Review* LXXIX (2): 214–236.

Churchland, P.M. 1979. *Scientific Realism and the Plasticity of Mind*. Cambridge: Cambridge University Press.

Churchland, P.M. 1981a. "Is Thinker a Natural Kind?" *Dialogue* 21: 223–238.

Churchland, P.M. 1981b. "Eliminative Materialism and the Propositional Attitudes" In W. Lycan (ed.), 1990. *Mind and Cognition*. Oxford: Blackwell, 206–223.

Churchland, P.M. 1985. "Reduction, Qualia, and the Direct Introspection of Brain States" *The Journal of Philosophy* 82: 8–28.

Churchland, P.M. 1988a. *Matter and Consciousness*. (2nd ed.) Cambridge, MA: MIT Press.

Churchland, P.M. 1988b. "Perceptual Plasticity and Theoretical Neutrality" in *Philosophy of Science* 55 (2): 167–187.

Churchland, P.M. 1989. *A Neurocomputational Perspective: The Nature of Mind and the Structure of Science*. Cambridge, MA MIT Press.

Churchland, P.M. 1991. "Folk Psychology and the Explanation of Human Behaviour." In J. Greenwood (ed.). *The Future of Folk Psychology*. Cambridge: Cambridge University Press, 51–69.

Churchland, P.M. 1993. "Evaluating Our Self Conception" *Mind & Language* 8 (2): 211–222.

Churchland, P.M. 1998. "Conceptual Similarity across Sensory and Neural Diversity: The Fodor/Lepore Challenge Answered" *Journal of Philosophy* XCV (1): 5–32.

Churchland, P.M. and P.S. Churchland 1990. "Stalking the Wild Epistemic Engine" In W. Lycan (ed.), *Mind and Cognition*. Oxford: Blackwell, 300–311.

Churchland, P.M. and P.S. Churchland 1996. "Replies from the Churchlands." In R. McCauley (ed.), *The Churchlands and Their Critics*. Oxford: Blackwell, 217–306.

Cioffi, F. 1991. "Wittgenstein and Freud's 'abominable mess.'" In P. Griffiths (ed.), *Wittgenstein: Centenary Essays*. Cambridge: Cambridge University Press, 169–192.

Clark, A. 1989. *Microcognition: Philosophy, Cognitive Science, and Parallel Distributing Processing*. Cambridge, MA: MIT Press.

Clark, A. 1993a. "The Varieties of Eliminativism: Sentential, Intentional and Catastrophic." *Mind & Language* 8 (2): 223–233.

Clark, A. 1993b. *Associative Engines*. Cambridge, MA: MIT Press.

Clark, A. 1996. "Dealing in Futures: Folk Psychology and the Role of Representations in Cognitive Science" In R. McCauley (ed.), *The Churchlands and Their Critics*. Oxford: Blackwell, 86–103.

Clark, A. 1997. *Being There: Putting Brain, Body and World Together Again*. Cambridge, MA: MIT Press.

Clark, A. 1998: "Magic Words: How Language Augments Human Computation" In P. Carruthers and J. Boucher. (eds.), *Language and Thought*. Cambridge: Cambridge University Press, 162–183.

Coates, P. and D. Hutto (eds.). 1996. *Current Issues in Idealism*. Bristol: Thoemmes Press.

Coates, P. 1997. "Meaning, Mistake and Miscalculation" *Mind and Machines* 7: 171–179.

Collins, A.W. 1998. "Beastly Experience" *Philosophy and Phenomenological Research* LVIII (2): 375–380.

Cram, H-R. 1992. "Fodor's Causal Theory of Representation" *Philosophical Quarterly*. 42 (166): 56–70.

Crane, T. 1992. "The Nonconceptual Content of Experience" In T. Crane (ed.), *The Contents of Experience*. Cambridge: Cambridge University Press, 136–157.

Crane, T. 1995. *The Mechanical Mind*. London: Penguin.

Crane, T. and D.H. Mellor 1990. "There is No Question of Physicalism" *Mind* 99, (394): 185–206.

Cummins, R. 1983. *The Nature of Psychological Explanation*. Cambridge, MA: MIT Press.

Cummins, R. 1991a. *Meaning and Mental Representation*. Cambridge, MA: MIT Press.

Cummins, R. 1991b. "The Role of Mental Meaning in Psychological Explanation" In B. McLaughlin (ed.), *Dretske and His Critics* Oxford: Blackwell, 102–117.

Cussins, A. 1990. "The Connectionist Construction of Concepts" In M. Boden (ed.), *The Philosophy of Artificial Intelligence* .Oxford: Oxford University Press, 368–440.

Cussins, A. 1992a. "The Limits of Pluralism" In D. Charles and K. Lennon (eds.), *Reduction, Explanation and Realism*. Oxford: Oxford University Press, 179–223.

Cussins, A. 1992b. "Content, Embodiment and Objectivity: The Theory of Cognitive Trails" *Mind* 101 (404): 651–688.

Cussins, A. 1993. "Nonconceptual Content and the Elimination of Misconceived Composites!" *Mind & Language* 8 (2): 234–251.

Davidson, D. 1980a. *Essays on Actions and Events*. Oxford: Clarendon Press.

Davidson, D. 1980b. "What Metaphors Mean" In M. Platts (ed.), *Reference, Truth and Reality*. London: Routledge and Kegan Paul, 238–254.

Davidson, D. 1984a. *Inquiries into Truth and Interpretation.* Oxford: Clarendon Press.

Davidson, D. 1984b. "First Person Authority" *Dialectica* 38 (2-3): 101–111.

Davidson, D. 1985a. "Adverbs of Action" In B. Vermazen and M. Hintikka (eds.). *Essays on Davidson: Actions and Events*. Oxford: Clarendon Press, 230–241.

Davidson, D. 1985b. "Replies to Essays I-X & X-XII" In B. Vermazen and M. Hintikka (eds.), *Essays on Davidson: Actions and Events*. Oxford: Clarendon Press, 195–253.

Davidson, D. 1985c. "Rational Animals" In E. Lepore and B. MacLaughlin (eds.), *Actions and Events: Perspectives on the Philosophy of Donald Davidson.* Oxford: Basil Blackwell, 473–480.

Davidson, D. 1985d. "Reply to Quine on Events" In E. Lepore and B. MacLaughlin (eds.). *Actions and Events: Perspectives on the Philosophy of Donald Davidson*. Oxford: Basil Blackwell, 172–176.

Davidson, D. 1985e. "Incoherence and Irrationality" *Dialectica* 39: 345–354.

Davidson, D. 1986a. "A Coherence Theory of Truth and Knowledge" In E. Lepore (ed.). *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*. Oxford: Basil Blackwell, 307–319.

Davidson, D. 1986b. "Empirical Content" In E. Lepore (ed.). *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*. Oxford: Basil Blackwell, 320–332.

Davidson, D. 1986c. "A Nice Derangement of Epitaphs" In E. Lepore (ed.), *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson.* Oxford: Basil Blackwell, 433–446.

Davidson, D. 1987a. "Knowing One's Own Mind" *Proceedings of the American Philosophical Association* 60 (3): 441–458.

Davidson, D. 1987b. "Problems in the Explanation of Action" *Metaphysics and Morality*. Oxford: Blackwell, 34–49.

Davidson, D. 1989. "What is Present to the Mind?" In J. Brandl & W. Gombocz (eds), *The Mind of Donald Davidson.* Amsterdam: Rodopi, 3–18.

Davidson, D. 1990a. "Afterthoughts, 1987" In A. Malachowski (ed.), *Reading Rorty: Critical Responses to Philosophy and the Mirror of Nature (and Beyond).* Oxford: Blackwell, 134–138.

Davidson, D. 1990b. "The Structure and Content of Truth" *Journal of Philosophy* LXXXVIII (6): 279–328.

Davidson, D. 1991a. "Three Varieties of Knowledge" In P. Griffiths (ed.). *A.J. Ayer: Memorial Essays*. Cambridge: Cambridge University Press. 153–166.

Davidson, D. 1991b. "Epistemology Externalized" *Dialectica* 45, (2-3): 191–202.

Davidson, D. 1992. "The Second Person" *Midwest Studies in Philosophy* 17: 255–267.

Davidson, D. 1993. "Thinking Causes" In J. Heil and A. Mele (eds.), *Mental Causation.* Oxford: Oxford University Press, 3–17.

Davidson, D. 1995. "Laws and Causes" *Dialectica* 49 (2-4): 263–279.

Davidson, D. 1996. "The Folly of Trying to Define Truth" *Journal of Philosophy.* XCIII (6): 263–278.

Davies, P. 1994. "Troubles For Direct Proper Functions" *Nous* 28 (3): 363–381.

Davies, M. and T. Stone (eds.). 1993. *Mental Simulation: Philosophical and Psychological Essays*. Oxford: Blackwell.

Dennett, D.C. 1969. *Content and Consciousness*. London: Routledge and Kegan Paul.

Dennett, D.C. 1984. *Elbow Room: The Varieties of Free Will Worth Wanting.* Oxford: Clarendon Press.

Dennett, D.C. 1985. *Brainstorms: Philosophical Essays on Mind and Psychology.* Cambridge, MA: MIT Press.

Dennett, D.C. 1986. "The Logical Geography of Computational Approaches: A View From the East Pole" In M. Brand and M. Harnish (eds.), *The Representation of Knowledge and Belief.* Tuscon: The University of Arizona Press, 59–79.

Dennett, D. C. 1987. *The Intentional Stance*. Cambridge, MA: MIT Press.

Dennett, D.C. 1990a. "Quining Qualia" In W. Lycan (ed.), *Mind and Cognition.* Oxford: Basil Blackwell, 519–547.

Dennett, D.C. 1990b. "The Interpretation of Texts, People, and Other Artifacts" *Philosophy and Phenomenological Research* 50: 177–194.

Dennett, D.C. 1991a. "Real Patterns" *The Journal of Philosophy* LXXXVIII (1): 27–51.

Dennett, D.C. 1991b. "Granny's Campaign for a Safe Science" In B. Loewer and G. Rey (eds). *Meaning in Mind: Fodor and His Critics*. London: Blackwell, 87–94.

Dennett, D.C. 1991c. *Consciousness Explained.* New York: Penguin Books.

Dennett, D.C. 1991d. "Two Contrasts: Folk Craft versus Folk Science, and Belief versus Opinion" In J. Greenwood (ed.), *The Future of Folk Psychology.* Cambridge: Cambridge University Press, 135–148.

Dennett, D.C. 1991e. "Ways of Establishing Harmony" In B. McLaughlin (ed.), *Dretske and His Critics.* Oxford: Blackwell, 118–129.

Dennett, D. C. 1995. *Darwin's Dangerous Idea.* New York: Simon and Schuster.

Dennett. D. C. 1997. *Kinds of Mind.* London: Phoenix.

Dennett. D.C. 1998: "Reflections on Language and Mind" In P. Carruthers and J. Boucher (eds.). *Language and Thought.* Cambridge: Cambridge University Press, 284–294.

Descartes, R. 1969. A. Anscombe and P. Geach (eds.). *Philosophical Writings.* London: Nelson.

Devitt, M. 1990a. "A Narrow Representational Theory of Mind" In W. Lycan (ed.), *Mind and Cognition*. Oxford: Blackwell, 371–398.

Devitt, M. 1990b. "On Removing Puzzles about Belief Ascription." *Pacific Philosophical Quarterly*. 71. 165–181.

Dretske, F.I. 1972. "Contrastive Statements" *Philosophical Review* 81: 411–437.

Dretske, F.I. 1977. "Referring to Events" *Midwest Studies in Philosophy*, II: 90–99.

Dretske, F.I. 1978. "Conclusive Reasons" In G. Pappas and M. Swain (eds.), *Essays on Knowledge and Justification*. New York: Cornell University Press, 41–60.

Dretske, F.I. 1981. *Knowledge and the Flow of Information*. Cambridge, MA: MIT Press.

Dretske, F.I. 1986a. "Misrepresentation" In R. Bogdan (ed.), *Belief* Oxford: Oxford University Press, 17–36.

Dretske, F.I. 1986b. "Aspects of Cognitive Representation" In M. Brand and M. Harnish (eds.), *The Representation of Knowledge and Belief*. Tucson: University of Arizona Press, 101–115.

Dretske, F.I. 1988. *Explaining Behaviour: Reasons in a World of Causes*. Cambridge, MA: MIT Press.

Dretske, F.I. 1990a. "*Précis* to Explaining Behaviour" *Philosophy and Phenomenological Research* 50 (4): 783–786.

Dretske, F.I. 1990b. "Replies to Reviewers." *Philosophy and Phenomenological Research* 50 (4): 819–840.

Dretske, F.I. 1991. "Dretske's Replies." In B. McLaughlin (ed.), *Dretske and His Critics*. Oxford: Blackwell, 180–221.

Dretske, F.I. 1995. *Naturalizing the Mind*. Cambridge, MA: MIT Press.

Dummett, M. 1973. *Frege: Philosophy of Language*. London: Duckworth.

Dummett, M. 1975. "What is a Theory of Meaning?" In S. Guttenplan (ed.), *Mind and Language*. Oxford: Clarendon Press, 97–158.

Dummett, M. 1976. "What is a Theory of Meaning II?" In G. Evans and J. McDowell (eds.), *Essays in Semantics*. Oxford: Clarendon Press, 67–137.

Dummett, M. 1978. *Truth and Other Enigmas*. London: Duckworth.

Dummett, M. 1979. "Common Sense and Physics" In G.F. MacDonald (ed.), *Perception and Identity: Essays Presented to A.J. Ayer with his Replies to Them*. London: Macmillan Press Ltd, 1–40.

Dummett, M. 1981. *The Interpretation of Frege's Philosophy*. London: Duckworth.

Dummett, M. 1993. *The Seas of Language*. Oxford: Oxford University Press.

Dupré, J. 1981. "Natural Kinds and Biological Taxa" *Philosophical Review* XC (1): 66–90.

Dupré, J. 1983. "The Disunity of Science" *Mind* XCII: 321–346.

Dupré, J. 1993. *The Disorder of Things*. Cambridge, MA: Harvard University Press.

Elder, C. 1994. "Proper Functions Defended" *Analysis* 54 (3): 167–171.

Elder, C. 1998. "What Versus How in Naturally Selected Representations" *Mind* 107, (426): 349–363.

Evans, G. 1982. *The Varieties of Reference*. Oxford: Oxford University Press.

Evans, G. and J. McDowell (eds.). 1976. *Truth and Meaning*. Oxford: Oxford University Press.

Evnine, S. 1991. *Donald Davidson*. Cambridge: Polity Press.

Feyerabend, P.K. 1963. "Mental Events and the Brain" In W. Lycan (ed.), *Mind and Cognition*. Oxford: Blackwell, 204–205.

Field, H. 1978. "Mental Representation" Reprinted in 1981, N. Block (ed.), *Readings in the Philosophy of Psychology*. Vol. II. London: Methuen, 78–114

Field, H. 1980. "Tarski's Theory of Truth" In M. Platts (ed.), *Reference, Truth and Reality*. London: Routledge and Kegan Paul, 83–110.

Fine, A. 1984. "The Natural Ontological Attitude" In J. Leplin (ed.). *Scientific Realism*. Berkeley: California University Press, 83–107.

Flanagan, O. 1984. *The Science of the Mind*. Cambridge, MA MIT Press.

Fodor, J.A. 1968. *Psychological Explanation: An Introduction to the Philosophy of Psychology*. New York: Random House.

Fodor, J.A. 1975. *The Language of Thought*. Cambridge, MA: Harvard University Press.

Fodor, J.A. 1981a. "Propositional Attitudes" In N. Block (ed.), *Readings in the Philosophy of Psychology*. Vol. II. London: Methuen, 45–63.

Fodor, J.A. 1981b. "Imagistic Representation" In N. Block (ed.), *Readings in the Philosophy of Psychology*. Vol. II. London: Methuen, 135–149.

Fodor, J.A. 1981c. *Representations: Philosophical Essays on the Foundations of Cognitive Science*. Cambridge, MA: MIT Press.

Fodor, J.A. 1982. "Cognitive Science and the Twin Earth Problem." *Notre Dame Journal of Formal Logic* 23 (2): 98–118.

Fodor, J.A. 1983. *The Modularity of Mind*. Cambridge, Mass: MIT Press.

Fodor, J.A. 1986a. "Information and Association" In M. Brand and M. Harnish (eds.), *The Representation of Knowledge and Belief*. Tucson: University of Arizona Press, 80–100.

Fodor, J.A. 1986b. "Why Paramecia Don't Have Mental Representations" *Midwest Studies in Philosophy*. X: 3–23.

Fodor, J.A. 1987. *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. Cambridge, Mass: MIT Press.

Fodor, J.A. 1990a. "Banish Discontent" In W. Lycan (ed.), *Mind and Cognition*. Oxford: Blackwell, 420–438.

Fodor, J.A. 1990b. *A Theory of Content and Other Essays*. Cambridge, MA MIT Press.

Fodor, J.A. 1990c. "You Can Fool Some of the People All of the Time, Everything Else Being Equal; Hedged Laws and Psychological Explanations" *Mind* 100, (307): 19–34.

Fodor, J.A. 1991a. "A Modal Argument for Narrow Content" *Journal of Philosophy* LXXXVIII (1): 5–26.

Fodor, J.A. 1991b. "Fodor's Guide to Mental Representation" In J. Greenwood (ed.). *The Future of Folk Psychology*. Cambridge: Cambridge University Press, 22–50.

Fodor, J.A. 1991c. "Replies" In B. Loewer and G. Rey (eds.), *Meaning and Mind* Oxford: Blackwell, 255–319.

Fodor, J.A. 1992. "The Big Idea — Can there be a Science of the Mind?" *Times Literary Supplement* July, 3: 5–7.

Fodor, J.A. 1994a. *The Elm and the Expert*. Cambridge, MA: MIT Press.

Fodor, J.A. 1994b. "Fodor, Jerry A" *A Companion to the Philosophy of Mind*. Oxford: Blackwell, 292–300.

Fodor, J.A. 1995. "A Theory of the Child's Theory of Mind" In M. Davies and T. Stone (eds.), *Mental Simulation*. Oxford: Blackwell, 109–122.

Fodor, J.A. 1998. *Concepts*. Oxford: Oxford University Press.

Fodor, J.A. and E. Lepore 1991. *Holism: A Shopper's Guide*. Oxford: Blackwell.

Fodor, J.A. and E. Lepore 1996a. "Paul Churchland and State Space Semantics" In R. McCauley (ed.), *The Churchlands and Their Critics*. Oxford: Blackwell, 145–159.

Fodor, J.A. and E. Lepore. 1996b. "Reply to Churchland" In R. McCauley (ed.), *The Churchlands and Their Critics*. Oxford: Blackwell, 159–162.

Fodor, J.A. and B. McLaughlin 1995. "Connectionism and the Problem of Systematicity" In C. MacDonald and G. MacDonald (eds.), *Connectionism: Debates on Psychological Explanation*. Oxford: Basil Blackwell, 199–222.

Fodor, J.A. and Z. Pylyshyn 1995. "Connectionism and Cognitive Architecture" In C. MacDonald and G. MacDonald (eds.), *Connectionism: Debates on Psychological Explanation*. Oxford: Basil Blackwell, 90–163.

Follesdal, D. 1985. "Causation and Explanation: A Problem in Davidson's view of Action and Mind" In E. Lepore and B. MacLaughlin (eds.). *Actions and Events: Perspectives on the Philosophy of Donald Davidson*. Oxford: Basil Blackwell, 311–323.

Freeman, N. "Theories of Mind in Collision" In M. Davies and T. Stone (eds.). *Mental Simulation*. Oxford: Blackwell, 68–86.

Frege, G. 1952. "On Sense and Reference" In P. Geach and M. Black (eds.), *Translations from the Philosophical Writings of Gottlob Frege*. Oxford: Blackwell, 56–78.

Frege, G. 1984. "Thoughts" In N. Salmon and S. Soames (eds.), *Propositions and Attitudes*. Oxford: Oxford University Press, 33–55.

Fuller, G. 1995. "Simulation and Psychological Concepts" In M. Davies and T. Stone (eds), *Mental Simulation.* Oxford: Blackwell, 19–32.

Gadamer, H-G. 1960. "The Nature of Things and the Language of Things" *Philosophical Hermeneutics*. 1976. Berekely: University of California Press, 69–81.

Gadamer, H-G. 1966. "Man and Language" *Philosophical Hermeneutics*. 1976. Berekely: University of California Press, 59–68.

Gadamer, H-G. 1976. *Philosophical Hermeneutics*. Berekely: University of California Press.

Garrett, R. and G. Graham 1991. "Why not Naturalistic Psychology?" *Philosophia* 20 (4): 377–385.

Gasper, P. 1991. "The Philosophy of Biology" In R. Boyd, P. Gasper, and J.D. Trout (eds.), *The Philosophy of Science* Cambridge, MA: MIT Press, 545–552.

Gill, C. 1991. "Is There a Concept of Person in Greek Philosophy" In S. Everson (ed.). *Psychology: Companions to Ancient Thought 2.* Cambridge: Cambridge University Press, 166–193.

Gillet, G. 1993. "Actions, Causes and Mental Ascriptions" In H. Robinson (ed.). *Objections to Physicalism* Oxford: Clarendon Press, 81–100.

Godfrey-Smith, P. 1989. "Misinformation" *Canadian Journal of Philosophy* 19, (4): 533-550.

Godfrey-Smith, P. 1992. "Indication and Adaptation" *Synthese* 92: 283–312.

Godfrey-Smith, P. 1993. "Functions: Consensus Without Unity" *Pacific Philosophical Quarterly* 74: 196–208.

Godfrey-Smith, P. 1994a. "A Modern History Theory of Functions" *Nous* 28 (3): 344–362.

Godfrey-Smith, P. 1994b. "A Continuum of Semantic Optimism." In S. Stich and T. Warfield (eds.), *Mental Representation: A Reader.* Oxford: Blackwells, 259–277,

Goldman, A. 1978. "A Causal Theory of Knowing" In G. Pappas and M. Swain (eds.), *Essays on Knowledge and Justification.* New York: Cornell University Press, 67–86.

Goldman, A. 1989. "Interpretation Psychologized" *Mind and Language* 4: 161–185.

Goldman, A. 1992a. "In Defense of Simulation Theory" *Mind and Language* 7 (1-2): 104–119.

Goldman, A. 1992b. "Empathy, Mind and Morals" *Proceedings and Addresses of the American Philosophical Association.* 66 (3): 17–41.

Goldman, A. 1993. "The Psychology of Folk Psychology" *Behavioural and Brain Sciences* 16: 15–28.

Gómez, J. 1998: "Some Thoughts about the Evolution of LADS, with Special Reference to TOM and SAM" In P. Carruthers and J. Boucher (eds.). *Language and Thought*. Cambridge: Cambridge University Press, 76–93.

Gordon, R. 1986. "Folk Psychology as Simulation" *Mind and Language* 1 (2): 158–171.

Gordon, R. 1992. "The Simulation Theory: Objections and Misconceptions" *Mind and Language* 7 (1-2): 11–34.

Gordon, R 1995. "Simulation Without Introspection or Inference from Me to You" In M. Davies & T. Stone (eds.), *Mental Simulation.* Oxford: Blackwell, 53–67.

Gordon, R. 1996. "'Radical' Simulationism" In P. Carruthers and P. Smith (eds.), *Theories of Theories of Mind* Cambridge: Cambridge University Press, 11–21.

Gopnik, A. 1996. "Theories and Modules: Creation Myths, Developmental Realities and Neurath's Boat" In P. Carruthers and P. Smith (eds.), *Theories of Theories of Mind.* Cambridge: Cambridge University Press, 169–183.

Graham, G. 1987b. "The Origins of Folk Psychology" *Inquiry* 30: 357–379.

Graham, G. 1993. *Philosophy of Mind.* Oxford: Blackwell.

Graham, G. and T. Horgan 1988. "How to be Realistic about Folk Psychology." *Philosophical Psychology* 1: 69–81.

Grandy, R. 1973. "Reference, Meaning and Belief" *Journal of Philosophy* 70: 439–452.

Greenwood, J. 1991. "Reasons to Believe" In J. Greenwood (ed.), *The Future of Folk Psychology*. (Cambridge: Cambridge University Press, 70–92.

Grice, P. 1989. *Studies in the Way of Words*. Cambridge, MA Harvard University Press.

Grimes, T. 1988. "The Myth of Supervenience" *Pacific Philosophical Quarterly.* 69: 152–160.

Haack, S. 1983. "Theories of Knowledge: An Analytic Framework" *Proceedings of the Aristotelian Society* 83: 143–157.

Haldane, J. 1988. "Understanding Folk" *Proceedings of the Aristotelian Society.* Supplementary Volume 62: 223–254.

Hannan, B. 1993. "Don't Stop Believing: The Case Against Eliminative Materialism" *Mind & Language* 8 (2): 165–179.

Hare, R. M. 1984. "Supervenience" *Proceedings of The Aristotelian Society* Supplementary Volume 58: 1–16.

Harman, G. 1973. *Thought*. New Jersey: Princeton University Press.

Harman, G. 1982. "Conceptual Role Semantics" *Notre Dame Journal of Formal Logic* 23 (2): 242–256.

Harris, P.L. 1992. "From Simulation to Folk Psychology: The Case for Development" *Mind and Language* 7 (1-2): 120–144.

Haugeland, J. (ed.) 1981. *Mind Design: Philosophy, Psychology and Artificial Intelligence*. Cambridge, MA MIT Press.

Haugeland, J. 1982. "Weak Supervenience" *American Philosophical Quarterly* 19: 93–101.

Heal, J. 1995. "How to Think about Thinking" In M. Davies and T. Stone (eds.), *Mental Simulation*. Oxford: Blackwell, 33–52.

Heal, J. 1996. "Simulation, Theory and Content" In P. Carruthers and P. Smith (eds.). *Theories of Theories of Mind.* Cambridge: Cambridge University Press, 75–89.

Heidegger, M. 1962. *Being and Time*. (trans. J. Macquarrie and E. Robinson). New York: Harper and Row.

Heil, J. 1991. "Being Indiscrete" In J. Greenwood (ed.), *The Future of Folk Psychology*. Cambridge: Cambridge University Press, 120–134.

Heil, J: 1992. *The Nature of True Minds*. Cambridge: Cambridge University Press.

Hellman, G. 1992. "Is Supervenience/Determination a Two-Way Street? Yes, But One of the Ways is the Wrong Way" *Journal of Philosophy* 89: 42–47.

Honderich, T. 1981. "Psychophysical Lawlike Connections and Their Problem" *Inquiry* 24: 277–303.

Honderich, T. 1982. "The Argument for Anomalous Monism" *Analysis* 42: 59–64.

Honderich, T. 1983. "Anomalous Monism: Reply to Smith" *Analysis* 43: 147–149.

Honderich, T. 1984. "Smith the Champion of Mauve" *Analysis* 44: 86–89.

Hookway, C. (ed.) 1984: *Minds, Machines and Evolution*. Cambridge: Cambridge University Press.

Hookway, C. 1988. *Quine: Language, Experience and Reality*. Oxford: Polity Press.

Horgan, T. 1982. "Supervenience and Microphysics" *Pacific Philosophical Quarterly* 63: 29–43.

Horgan, T. 1991. "The Explanatory Role of Content" In B. McLaughlin (ed.), *Dretske and His Critics*. Oxford: Blackwell. 73–101.

Horgan, T. 1993. "The Austere Ideology of Folk Psychology" *Mind & Language* 8 (2): 282–297.

Horgan, T. and J. Woodward 1990. "Folk Psychology is Here to Stay" In W. Lycan (ed.), *Mind and Cognition*. Oxford: Blackwell, 399–420.

Horgan, T. and G. Graham 1991. "In Defense of Southern Fundamentalism" *Philosophical Studies* 62: 107–134.

Hornsby, J. 1980-81. "Which Mental Events are Physical Events?" *Proceedings of the Aristotelian Society* 81: 73–92.

Hornsby, J. 1985. "Physicalism, Events and Part-Whole Relations" In E. Lepore and B. McLaughlin (eds.), *Actions and Events: Perspectives on the Philosophy of Donald Davidson*. Oxford: Basil Blackwell, 444–458.

Hornsby, J. 1986. "Physicalist Thinking and Conceptions of Behaviour" In P. Pettit and J. McDowell (eds), *Subject, Thought and Context*. Oxford: Oxford University Press, 95–116.

Hornsby, J. 1992. "Physics, Biology and Common-Sense Psychology" In D. Charles and K. Lennon (eds.), *Reduction, Explanation and Realism.* Oxford: Oxford University Press, 155–178.

Hornsby, J. 1993. "Agency and Causal Explanation" In J. Heil and A. Mele (eds.), *Mental Causation.* Oxford: Oxford University Press, 161–188.

Hull, D. 1984. "Historical Entities and Historical Narratives" In C. Hookway (ed.), *Minds Machines and Evolution.* Cambridge: Cambridge University Press, 17–42.

Hutto, D. 1992a. "Prins Autos Herredomme: Psykologi I Naturbidenskabens Tidsalder" (The Reign of Prince Auto: Psychology in an Age of Science). *Philosophia* 21: (1-2). 61–80.

Hutto, D. 1992b. "A Job For Philosophy" *Philosophy Now* 4: 19–23.

Hutto, D. 1992c. "Nothing Personal: Ethics Without People?" *The Philosopher: The Journal of the Philosophical Society of England:* 20–27.

Hutto, D. 1993. "A Tactical Defense of Folk Psychology" *The Proceedings of the Mind and Related Matters Conference*, Leeds *Inside/Out* 8: 12–15.

Hutto, D. 1994. "Minding Our Language: The Role of Simulation in Linguistic Interpretation" *The Proceedings of the Language and Related Matters Conference*, University of Leeds, *Inside/Out* 9: 2-5.

Hutto, D. 1995a. "The Mindlessness of Computationalism: The Neglected Aspects of Cognition" In P. Pyllkkänen (ed.). *New Directions in Cognitive Science.* Helsinki: Finnish Society for Artificial Intelligence, 201-211.

Hutto, D. 1995b. "Consciousness Demystified: A Wittgensteinian Critique of Dennett's Project." *The Monist* 78 (4): 464-478.

Hutto, D. 1996. "Was the Later Wittgenstein a Transcendental Idealist?" In P. Coates and D. Hutto (eds.), *Current Issues in Idealism.* Bristol: Thoemmes Press. 121-158.

Hutto, D. 1997. "The Story of the Self: Narrative as the Basis for Self-Development" In K. Simms (ed.), *Ethics and the Subject.* Amsterdam: Editions Rodopi, 61-75.

Hutto, D. 1998a. "Davidson's Identity Crisis" *Dialectica* 52 (1): 45-61.

Hutto, D. 1998b. "An Ideal Solution to the Problems of Consciousness" *Journal of Consciousness Studies* 5 (3): 328-343.

Hutto, D. 1998c. "Bradleian Metaphysics: A Healthy Scepticism" *Bradley Studies.* 4. (1): 82-96.

Hutto, D. 1998d. "Nonconceptual Content and Objectivity" In R. Grush (ed.), *The Electronic Journal of Analytic Philosophy: Special Issue on Gareth Evans.* http://www.phil.indiana. edu/ejap/ejap.html

Hutto, D. 1999a. "A Cause for Concern: Reasons, Causes and Explanation" *Philosophy and Phenomenological Research* 59 (2):

Hutto, D. 1999b. "Cognition Without Representation" In A. Riegler and M. Peschl. (eds.).*Understanding Representation in Cognitive Science.* New York: Plenum Press.

Hutto, D. Forthcoming-a. "Conceptual Schema and the Problem of Consciousness." In P. Pllykkänen (ed.), *Varieties and Dimensions of Conscious Experience* (Advances in Consciousness Research Series) Amsterdam, Philadelphia: John Benjamins Publishing Company.

Hutto, D. Forthcoming-b. *Beyond Physicalism. An Ideal Solution to the Problems of Consciousness.* (Advances in Consciousness Research Series) Amsterdam, Philadelphia: John Benjamins Publishing Company.

Jackson, F. and P. Pettit 1988. "Functionalism and Broad Content" *Mind* XCVII (387): 381–400.

Jackson, F. and P. Pettit 1990. "Causation in the Philosophy of Mind" *Philosophy and Phenomenological Research* 50: 195–214.

Jackson, F. and P. Pettit 1993a. "Some Content is Narrow" In J. Heil and A. Mele (eds), *Mental Causation.* Oxford: Oxford University Press, 259–282.

Jackson, F. and P. Pettit 1993b. "Folk Belief and Commonplace Belief" *Mind & Language* 8 (2): 298–305.

Jacob, P. 1997. What *Minds Can Do.* Cambridge: Cambridge University Press.

Johnston, M. 1985. "Why Having a Mind Matters" In E. Lepore and B. McLaughlin (eds.), *Actions and Events: Perspectives on the Philosophy of Donald Davidson.* Oxford: Basil Blackwell, 408–426.

Johnston, P. 1989. *Wittgenstein and Moral Philosophy.* London: Routledge.

Kim, J. 1966: "On the Psycho-Physical Identity Theory" *American Philosophical Quarterly.* 3 (3): 227–235.

Kim, J. 1978. "Supervenience and Nomological Incommensurables" *American Philosophical Quarterly*, 15 (2): 149–156

Kim, J. 1979. "Causality, Identity and Supervenience in the Mind-Body Problem" *Midwest Studies in Philosophy* IV: 31–49.

Kim, J. 1982. "Psychophysical Supervenience" *Philosophical Studies* 41: 51–70.

Kim, J. 1984a. "Epiphenomenal and Supervenient Causation" *Midwest Studies in Philosophy*, IX: 275–269.

Kim, J. 1984b. "Concepts of Supervenience" *Philosophy and Phenomenological Research* 45: 153–176.

Kim, J. 1985. "Psychophysical Laws" In E. Lepore and B. MacLaughlin (eds.), *Actions and Events: Perspectives on the Philosophy of Donald Davidson.* Oxford: Basil Blackwell, 369–386.

Kim, J. 1987. "'Strong' and 'Global' Supervenience Revisited" *Philosophy and Phenomenological Research* 48: 315–326.

Kim, J. 1990. "Supervenience as a Philosophical Concept" *Metaphilosophy* 21 (1-2): 1–27,

Kim, J. 1991. "How Reasons Explain Behaviour," In B. McLaughlin (ed.), *Dretske and His Critics*. Oxford: Blackwell, 52–72.

Kim, J. 1993. "Can Supervenience and 'Not-Strict Laws' Save Anomalous Monism?" In J. Heil and A. Mele (eds.), *Mental Causation* Oxford: Oxford University Press, 19–26.

Kraut, R. 1986. "The Third Dogma" In E. Lepore (ed.), *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*. Oxford: Basil Blackwell, 398–416.

Krebs, J.R. and N.B. Davies 1991. *Behavioural Ecology: An Evolutionary Approach*. 3rd Edition. Oxford: Blackwell Scientific Publications

Kripke, S. 1972. *Naming and Necessity*. Oxford: Blackwell.

Kripke, S. 1979. "A Puzzle About Belief" In N. Salmon and S. Soames (eds.), *Propositions and Attitudes*. Oxford: Oxford University Press, 102–148.

Kripke, S. 1982. *Wittgenstein: On Rules and Private Language*. Oxford: Blackwell.

Lear, J. 1986. "Transcendental Anthropology." In P. Pettit and J. McDowell (eds.), *Subject, Thought and Context*. Oxford: Oxford University Press, 267–298.

Leder, D. 1985. "Troubles with Token Identity" *Philosophical Studies* 47: 79–94.

Lennon, K. 1992. "Reduction, Causality and Normativity" In D. Charles and K. Lennon (eds.), *Reduction, Explanation and Realism*. Oxford: Oxford University Press, 225–238.

Lepore, E. (ed.) 1986a. *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*. Oxford: Basil Blackwell.

Lepore, E 1986b. "Truth in Meaning." In E. Lepore (ed.), *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*. Oxford: Basil Blackwell, 3–26.

Lepore, E. and B. Loewer 1989. "More on Making Mind Matter" *Philosophical Topics* XVII (1): 175–191.

Lepore E. and B. MacLaughlin (eds.). 1985: *Actions and Events: Perspectives on the Philosophy of Donald Davidson* Oxford: Basil Blackwell.

Leslie, A. and T. German 1995. "Knowledge and Ability in 'Theory of Mind': One-eyed Overview of a Debate" In M. Davies and T. Stone (eds.), *Mental Simulation*. Oxford: Blackwell, 123–150.

Lewis, D. 1970. "How to Define Theoretical Terms" *Journal of Philosophy* 67: 427–446.

Lewis, D. 1972. "Psychophysical and Theoretical Identifications" *Australian Journal of Philosophy* 50: 249–258.

Lewis, H. 1985. "Is the Mental Supervenient on the Physical?" In B. Vermazen and M. Hintikka (eds), *Essays on Davidson: Actions and Events*. Oxford: Clarendon Press, 159–172.

Loar, B. 1981. *Mind and Meaning*. Cambridge: Cambridge University Press.

Loar, B. 1982. "Conceptual Role and Truth-Conditions" *Notre Dame Journal of .Formal Logic* 23 (2): 272–283.

Loar, B. 1991. "Can We Explain Intentionality" In B. Loewer and G. Rey (eds.), *Meaning in Mind: Fodor and His Critics*. London: Blackwell, 119–135.

Loar, B. 1992. "Eliminativism versus Non-reductive Physicalism" In D. Charles and K. Lennon (eds.), *Reduction, Explanation and Realism*. Oxford: Oxford University Press, 239–264.

Lockwood, M. 1989. *Mind, Brain and the Quantum: The Compound I*. Oxford: Blackwell.

Loewer, B. and G. Rey 1991. "Editor's Introduction" In B. Loewer and G. Rey (eds.), *Meaning in Mind: Fodor and His Critics*. London: Blackwell, xi–xxxvii.

Lycan, W. 1969. "On Intentionality and the Psychological" *American Philosophical Quarterly* 6: 305–311.

Lycan, W. 1981c. "Toward a Homuncular Theory of Believing" *Cognition and Brain Science* 4: 139–159.

Lycan, W. 1986a. "Thoughts about Things" In M. Brand and M. Harnish (eds.), *The Representation of Knowledge and Belief*. Tucson: University of Arizona Press, 160–186.

Lycan, W. 1986b. "Semantics and Methodological Solipsism" In E. Lepore (ed.), *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*. Oxford: Basil Blackwell, 245–261.

Lycan, W. 1988. *Judgement and Justification*. Cambridge: Cambridge University Press.

Lycan, W. 1990a. "The Continuity of Levels of Nature" In W. Lycan (ed.), *Mind and Cognition*. Oxford: Blackwell, 77–96.

Lycan, W. 1990b. "On Respecting Puzzles about Belief Ascription" *Pacific Philosophical Quarterly* 71: 182–188.

Lycan, W. 1996. "Paul Churchland's PDP Approach to Explanation" In R. McCauley (ed.), *The Churchlands and Their Critics*. Oxford: Blackwell, 104–120.

Lyons, W. 1995. *Approaches to Intentionality*. Oxford: Oxford University Press.

MacDonald, C. 1989. *Mind-Body Identity Theories*. London: Routledge.

MacDonald, C. 1992. "Weak Externalism and Psychological Reduction" In D. Charles and K. Lennon (eds), *Reduction, Explanation and Realism*. Oxford: Oxford University Press, 133–154.

MacDonald, G. 1989. "Biology and Representation" *Mind & Language* 4 (3):186–200.

MacDonald, G. 1992. "Reduction and Evolutionary Biology" In D. Charles and K. Lennon (eds), *Reduction, Explanation and Realism*. Oxford: Oxford University Press, 69–96.

MacDonald, G. and C. MacDonald 1986. "Mental Causes and Explanation of Action" In L. Stevenson, R. Squires, and J. Haldane (eds.), *Mind, Causation, and Action*. Oxford: Basil Blackwell, 35–48.

MacDonald, C. and G. MacDonald (eds.). 1995. *Connectionism: Debates on Psychological Explanation*. Oxford: Basil Blackwell.

Maker, W. 1991. "Davidson's Transcendental Arguments" *Philosophy and Phenomenological Research* 51 (2): 345–360.

Malcolm, N. 1962. "Wittgenstein's Philosophical Investigations" In V.C. Chappell (ed.), *The Philosophy of Mind*. New Jersey: Prentice Hall, 74–100.

Malcolm, N. 1963. *Knowledge and Certainty: Essays and Lectures*. New York: Cornell University Press.

Malcolm, N. 1977. "Thoughtless Brutes" *Thought and Knowledge*. New York: Cornell University Press, 40–57.

Malcolm, N. 1986. *Nothing is Hidden*. Oxford: Blackwell.

Malpas, J.E. 1992. *Donald Davidson and the Mirror of Meaning*. Cambridge: Cambridge University Press.

Marcus, R.B. 1990. "Some Revisionary Proposals about Belief and Believing" *Philosophy and Phenomenological Research* 50: 133–153.

Margolis, J. 1991. "The Autonomy of Folk Psychology" In J. Greenwood (ed.), *The Future of Folk Psychology*. Cambridge: Cambridge University Press, 242–262.

Masson, D. 1897. "Introduction" to *The Poetical Works of John Milton*. London: Macmillan & Co., Ltd. 1–34.

Maynard-Smith, J. 1984. "The Evolution of Animal Intelligence" In C. Hookway (ed.), *Minds, Machines and Evolution*. Cambridge: Cambridge University Press. 63–72.

Maynard-Smith, John. 1990. "Explanation in Biology" In D. Knowles (ed.), *Explanation and Its Limits*. Cambridge: Cambridge University Press, 65–72.

Mayr, E. 1981. *The Growth of Biological Science*. Cambridge: Harvard University Press.

McCulloch, G. 1986. "Scientism, Mind and Meaning" In P. Pettit and J. McDowell (eds), *Subject, Thought and Context*. Oxford: Oxford University Press, 59–94.

McCulloch, G. 1990. "Dennett's Little Grains of Salt" *The Philosophical Quarterly* 40 (158): 1–12.

McDonough, R. 1989. "Towards a Non-Mechanistic Theory of Meaning" *Mind* 98: 1–22.

McDonough, R. 1991. "A Culturalist Account of Folk Psychology" In J. Greenwood (ed.). *The Future of Folk Psychology*. Cambridge: Cambridge University Press, 263–288.

McDowell, J. 1980. "On the Sense and Reference of a Proper Name." In M. Platts (ed.), *Reference, Truth and Reality*. London: Routledge and Kegan Paul, 141–166.

McDowell, J. 1984. "*De Re* Senses" In C. Wright (ed.), *Frege: Tradition and Influence*. Oxford: Basil Blackwell, 98–109.

McDowell, J. 1985. "Functionalism and Anomalous Monism" In E. Lepore and B. McLaughlin (eds.), *Actions and Events: Perspectives on the Philosophy of Donald Davidson.* Oxford: Basil Blackwell, 387–398.

McDowell, J. 1991. "Intentionality *De Re*" In E. Lepore and R. Van Gulick (eds.), *John Searle and His Critics*. Oxford: Blackwell, 215–226.

McDowell, J. 1994. *Mind and World.* Cambridge, MA: Harvard University Press.

McDowell, J. 1998. "*Précis* of Mind and World" *Philosophy and Phenomenological Research* 58 (2): 365–368.

McDowell, J. 1998. "Reply to Commentators" *Philosophy and Phenomenological Research* 58 (2): 403-431.

McGinn, C. 1980. "Truth and Use" In M. Platts (ed.), *Reference, Truth and Reality.* London: Routledge and Kegan Paul, 19–40.

McGinn, C. 1986. "Radical Interpretation and Epistemology" In E. Lepore (ed.), *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson.* Oxford: Basil Blackwell, 356–368.

McGinn, C. 1989. *Mental Content*. Oxford: Basil Blackwell.

McGinn, C. 1991. *The Problem of Consciousness*. Basil Blackwell.

McLaughlin, B. 1985. "Anomalous Monism and the Irreducibility of the Mental" In E. Lepore and B. McLaughlin (eds.), *Actions and Events: Perspectives on the Philosophy of Donald Davidson.* Oxford: Basil Blackwell, 331–368.

McLaughlin, B. 1991. "Belief Individuation and Dretske on Naturalizing Content" In B. McLaughlin (ed.), *Dretske and His Critics.* Oxford: Blackwell, 157–179.

McLaughlin, B. 1993. "On Davidson's Response to the Charge of Epiphenomenalism" In J. Heil and A. Mele (eds.), *Mental Causation.* Oxford: Oxford University Press, 27–40.

Melden, A.I. 1961. *Free Action*. London: Routledge, Kegan and Paul.

Mele, A. 1992. *The Springs of Action*. Oxford: Oxford University Press.

Miller, R. 1990. "Supervenience is a Two-Way Street" *Journal of Philosophy* 87: 695–701.

Millikan, R. 1984. *Language, Thought and Other Biological Categories*. Cambridge, MA: MIT Press.

Millikan, R. 1986. "Thoughts without Laws; Cognitive Science with Content" *Philosophical Review* 95: 47–80.

Millikan, R. 1989a. "Biosemantics" *Journal of Philosophy* 86: 281–297.

Millikan, R. 1989b. "In Defense of Proper Functions" *Philosophy of Science* 56: 288–302.

Millikan, R. 1990. "Seismograph Readings of Explaining Behaviour" *Philosophy and Phenomenological Research* 50 (4): 807–812.

Millikan, R. 1991. "Speaking Up For Darwin" In B. Loewer and G. Rey (eds.), *Meaning in Mind: Fodor & His Critics*. Oxford: Blackwell, 151–164.

Millikan, R. 1993. *White Queen Psychology and Other Essays for Alice*. Cambridge, MA: MIT Press.

Millikan, R. 1998: "A Common Structure for Concepts, Individuals, Stuffs and Real Kinds: More Mama, More Milk and More Mouse" *Behavioural and Brain Sciences* 21(1): 55–160.

Morton, A. 1980. *Frames of Mind: Constraints on the Common Sense Conception of the Mental*. Oxford: Clarendon Press.

Moser, P. 1992. "Physicalism and Global Supervenience" *The Southern Journal of Philosophy* XXX (1): 71–82.

Nagel, T. 1986. *The View From Nowhere*. Oxford: Oxford University Press.

Neander, K. 1991. "The Teleological Notion of Function" *Australasian Journal of Philosophy* 69 (4): 454–468.

Neander, K. 1995. "Misrepresenting and Malfunctioning" *Philosophical Studies* 79: 109–141.

Needham, R. 1972. *Belief, Language and Experience*. Oxford: Blackwell.

Newell, A. and H. Simon 1976. "Computer Science as Empirical Enquiry: Symbols and Search" In M. Boden (ed.). 1990. *The Philosophy of Artificial Intelligence*. Oxford. Oxford University Press, 105–132.

Nichols, S., S. Stich, A. Leslie, and D. Klein 1996. "Varieties of Off-line Simulation" In P. Carruthers and P. Smith (eds.), *Theories of Theories of Mind* Cambridge: Cambridge University Press, 39–74.

O'Hear, A. 1991. "Wittgenstein and the Transmission of Traditions" In P. Griffiths (ed.), *Wittgenstein: Centenary Essays*. Cambridge: Cambridge University Press, 41–60.

Papineau, D. 1984. "Representation and Explanation" *Philosophy of Science* 51: 550–572.

Papineau, D. 1987. *Reality and Representation*. Oxford: Blackwell.

Papineau, D. 1990. "Truth and Teleology" In D. Knowles (ed.), *Explanation and Its Limits*. Cambridge: Cambridge University Press, 21–43.

Papineau, D. 1992. "Irreducibility and Teleology" In D. Charles and K. Lennon (eds.), *Reduction, Explanation and Realism*. Oxford: Oxford University Press, 45–68.

Papineau, D. 1993. *Philosophical Naturalism*. Oxford: Blackwell.

Paull, R.C. and T.R. Sider 1992. "In Defense of Global Supervenience" *Philosophy and Phenomenological Research* 52 (4): 833–854.

Peacocke, C. 1983. *Sense and Content: Experience, Thought and Their Relations*. Oxford: Clarendon Press.

Peacocke, C. 1992a. *The Study of Concepts*. Cambridge, MA: MIT Press.

Peacocke, Christopher. 1992b. "Scenarios, Concepts and Perception" In T. Crane (ed.), *The Contents of Experience*. Cambridge: Cambridge University Press, 105–135.

Peacocke, C. 1994. "Non-conceptual Content: Kinds, Rationales and Relations" *Mind and Language* 9 (4): 419–429.

Peacocke, C. 1998. "Nonconceptual Content Defended" *Philosophy and Phenomenological Research* 58 (2): 381–388.

Perner, J. 1996. "Simulation as Explication of Prediction-Implicit Knowledge about the Mind: Arguments for a Simulation-Theory Mix" In P. Carruthers and P. Smith (eds.), *Theories of Theories of Mind* Cambridge: Cambridge University Press, 90–104.

Petrie, B. 1987. "Global Supervenience and Reduction" *Philosophy and Phenomenological Research* 48: 119–130.

Pettit, P. 1986. "Broad-Minded Explanation and Psychology" In P. Pettit and J. McDowell (eds.), *Subject, Thought and Context*. Oxford: Oxford University Press, 17–58.

Place, U.T. 1956. "Is Consciousness a Brain Process?" In W. Lycan (ed.), *Mind and Cognition*. Oxford: Blackwell, 29–36.

Place, U.T. 1988. "Thirty Five Years on – Is Consciousness Still a Brain Process?" *Australasian Journal of Philosophy* 66: 208-219.

Povinelli, D. 1996. "Chimpanzee Theory of Mind? The Long Road to Strong Inference" In P. Carruthers & P. Smith (eds.), *Theories of Theories of Mind* Cambridge: Cambridge University Press, 293–329.

Preston, J. 1989. "Folk Psychology as Theory or Practice" *Inquiry* 32: 276–303.

Putnam, H. 1975a. "The Meaning of Meaning" *Mind, Language and Reality: Philosophical Papers Vol. 2*. Cambridge: Cambridge University Press, 131–193.

Putnam, H. 1981. *Realism, Truth and History*. Cambridge: Cambridge University Press.

Putnam, H. 1986b. "Information and the Mental" In E. Lepore (ed.), *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*. Oxford: Basil Blackwell, 262–271.

Putnam, H. 1987. *The Many Faces of Realism*. Illinois: Open Court.

Putnam, H. 1988. *Representation and Reality*. Cambridge, MA: MIT Press.

Putnam, H. 1990. *Realism with a Human Face*. Cambridge, MA: Harvard University Press.

Putnam, H. 1993. *Renewing Philosophy*. Cambridge, MA: Harvard University Press.

Quine, W.V.O. 1953. *From A Logical Point of View*. Cambridge, MA: Harvard University Press.

Quine, W.V.O. 1956. "Quantifiers and Propositional Attitudes" *Journal of Philosophy* LIII (5): 177–187.

Quine, W.V.O. 1960. *Word and Object.* Cambridge, MA: MIT Press.

Quine, W.V.O. 1985. "Events and Reification" In E. Lepore and B. MacLaughlin (eds.), *Actions and Events: Perspectives on the Philosophy of Donald Davidson.* Oxford: Basil Blackwell, 162–171.

Quine, W.V.O. 1990. *The Pursuit of Truth.* Cambridge, MA: Harvard University Press.

Ramberg, B.T. 1989. *Donald Davidson's Philosophy of Language: An Introduction.* Oxford: Basil Blackwell.

Ramsey, W., S. Stich and J. Garon 1991. "Connectionism, Eliminativism and the Future of Folk Psychology." In J. Greenwood (ed.). *The Future of Folk Psychology.* Cambridge: Cambridge University Press, 93–119.

Richard, M. 1990. *Propositional Attitudes: An Essay on Thoughts and How We Ascribe Them.* Cambridge: Cambridge University Press.

Ricoeur, P. 1992. *Oneself as Another.* Chicago: Chicago University Press.

Root, M. 1986. "Davidson and Social Science" In E. Lepore (ed.), *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson.* Oxford: Basil Blackwell, 272–304.

Rorty, R. 1965. "Mind/Body Identity, Privacy, and Categories." *The Review of Metaphysics* 19: 24-121.

Rorty, R. 1980. *Philosophy and the Mirror of Nature.* Oxford: Blackwell.

Rorty, R. 1987b. "Science as Solidarity" in *Objectivity, Relativism and Truth: Philosophical Papers I.* Cambridge: Cambridge University Press, 35–45.

Rorty, R. 1988a. "Is Natural Science a Natural Kind?" in *Objectivity, Relativism and Truth: Philosophical Papers I.* Cambridge: Cambridge University Press, 46–62.

Rorty, R. 1998. "McDowell, Davidson, and Spontaneity" *Philosophy and Phenomenological Research* 58 (2): 389–394.

Rosenberg, A. 1985a. *The Structure of Biological Science.* Cambridge: Cambridge University Press.

Rosenberg, A. 1985b. "Davidson's Unintentional Attack on Psychology." In E. Lepore and B. McLaughlin (eds.), *Actions and Events: Perspectives on the Philosophy of Donald Davidson.* Oxford: Basil Blackwell, 399–407.

Rosenberg, J. 1991. "Tractarian States and Folk Psychological Explanation" In J. Greenwood (ed.), *The Future of Folk Psychology.* Cambridge: Cambridge University Press, 226–241.

Rovane, C. 1986. "The Metaphysics of Interpretation." In E. Lepore (ed.). *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson.* Oxford: Basil Blackwell.

Rowlands, M. (1997). "Teleological Semantics". *Mind* 106 (422): 279–303.

Russell, B. 1905. "On Denoting" In *Logic and Knowledge.* 1956. London: Unwin Hyman, 41–56.

Russell, B. 1919. "On Propositions: What they are and How they Mean" In *Logic and Knowledge*. 1956. London: Unwin Hyman, 285–320.

Russell, B. 1949. *Inquires into Meaning and Truth*. London: Unwin Hyman.

Ryle, G. 1949. *The Concept of Mind*. London: Hutchinson & Co.

Ryle, G. 1977. *On Thinking*. Oxford: Blackwell.

Salmon, N. and S. Soames 1988. *Propositions and Attitudes*. Oxford: Oxford University Press.

Sanford, D.H. 1985. "Causal Relata" In E. Lepore and B. McLaughlin (eds.), *Actions and Events: Perspectives on the Philosophy of Donald Davidson*. Oxford: Basil Blackwell, 282–293.

Schact, R. 1983. *Nietzsche*. London: Routledge.

Schiffer, S. 1972. *Meaning*. Oxford: Clarendon.

Schiffer, S. 1989. *Remnants of Meaning*. Cambridge, MA: MIT Press.

Seager, W. 1988. "Weak Supervenience and Materialism" *Philosophy and Phenomenological Research* 48: 697–709.

Searle, J. 1969. *Speech Acts*. Cambridge: Cambridge University Press.

Searle, J. 1979. *Expression and Meaning*. Cambridge: Cambridge University Press.

Searle, J. 1983. *Intentionality*. New York: Cambridge University Press.

Searle, J. 1984b. *Minds, Brains, and Science*. Cambridge, Mass: Harvard University Press.

Searle, J. 1990. "Is the Brain a Digital Computer" *Proceedings of the American Philosophical Association* 64 (3): 21–37.

Searle, J. 1992. *The Rediscovery of the Mind*. Cambridge, Mass: MIT Press.

Segal, G. 1996. "The Modularity Theory of Mind" In P. Carruthers and P. Smith (eds.), *Theories of Theories of Mind* Cambridge: Cambridge University Press, 141–157.

Sellars, W. 1963. "Empiricism and the Philosophy of Mind" *Science, Perception and Reality*. London: Routledge and Kegan Paul, 127–196.

Sharpe, R. 1987. "The Very Idea of Folk Psychology" *Inquiry* 30: 381–393.

Shoemaker, S. 1984. *Identity, Cause and Mind*. Cambridge: Cambridge University Press.

Shotter, J. 1991. "Wittgenstein and Psychology" In P. Griffiths (ed.), *Wittgenstein: Centenary Essays*. Cambridge: Cambridge University Press, 193–208.

Skinner, B.F. 1953. *Science and Human Behaviour*. New York: Macmillan.

Skinner, B.F. 1971. *Beyond Freedom and Dignity*. London: Penguin Books.

Skinner, B.F. 1976. *About Behaviourism*. New York: Knopf.

Smart, J.J.C. 1959. "Sensations and Brain Processes" In V. C. Chappell (ed.), *The Philosophy of Mind*. New Jersey: Prentice Hall, 160–172.

Smart, J.J.C. 1963. *Philosophy and Scientific Realism*. London: Routledge & Kegan Paul.

Smart, J.J.C. 1985. "Davidson's Minimal Materialism" In B. Vermazen and M. Hintikka (eds.), *Essays on Davidson: Actions and Events*. Oxford: Clarendon Press, 173–182.

Smart, J.J.C. 1986. "How to Turn the *Tractatus* Wittgenstein into (Almost) Donald Davidson" In E. Lepore (ed.), *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*. Oxford: Basil Blackwell, 92–100.

Smart, J.J.C. 1990. "Explanation: Opening Address" In D. Knowles (ed.), *Explanation and Its Limits*. Cambridge: Cambridge University Press, 1–19.

Smith, P. 1983. "Bad News for Anomalous Monism?" *Analysis* 42: 221–224.

Smith, P. 1984. "Anomalous Monism and Epiphenomenalism: A Reply to Honderich" *Analysis* 44: 83-86.

Smith, P. 1992. "Modest Reductions and the Unity of Science" In D. Charles and K. Lennon (eds.), *Reduction, Explanation and Realism*. Oxford: Oxford University Press, 19–43.

Smith, P. 1996. "Language and the Evolution of Mind-Reading" In P. Carruthers and P. Smith (eds.), *Theories of Theories of Mind*. Cambridge: Cambridge University Press, 344–354.

Smolensky, P. 1995a. "On the Proper Treatment of Connectionism" In C. MacDonald, and G. MacDonald (eds.), *Connectionism: Debates on Psychological Explanation*. Oxford: Basil Blackwell, 28–89.

Smolenksy, P. 1995b. "Connectionism, Constituency and the Language of Thought" In C. MacDonald, and G. MacDonald (eds.), *Connectionism: Debates on Psychological Explanation*. Oxford: Basil Blackwell, 164–198.

Sober, E. 1985. *The Nature of Selection*. Cambridge, MA: MIT Press.

Sober, E. 1990a. "Putting the Function back into Functionalism" In W. Lycan (ed.), *Mind and Cognition*. Oxford: Blackwell, 97–106.

Sober, E. 1990b. "Let's Razor Ockham's Razor" In D. Knowles (ed.), *Explanation and Its Limits*. Cambridge: Cambridge University Press, 73–93.

Sorabji, R. 1980. *Necessity, Cause and Blame: Perspectives on Aristotle's Theory*. London: Duckworth.

Sosa, E. 1984. "Mind-Body Interaction and Supervenient Causation" *Midwest Studies in Philosophy* IX: 271–281.

Sosa, E. 1993. "Davidson's Thinking Causes" In J. Heil and A. Mele (eds.), *Mental Causation*. Oxford: Oxford University Press, 41–50.

Stampe, D. 1977: "Towards a Causal Theory of Linguistic Representation" *Midwest Studies in Philosophy* II: 42–63.

Sterelny, K. 1990. *The Representational Theory of Mind*. Oxford: Blackwell.

Sterelny, K. 1993. "Refuting Eliminative Materialism on the Cheap?" *Mind & Language* 8 (2): 306–315.

Stevenson, L., R. Squires, and J. Haldane (eds.). 1986. *Mind, Causation and Action*. Oxford: Blackwell.

Stich, S. 1978. "Beliefs and Sub-doxastic States" *Philosophy of Science* 45: 499–518.

Stich, S. 1979. "Do Animals Have Beliefs?" *The Australasian Journal of Philosophy* 57: 15–28.

Stich, S. 1982. "On the Ascription Content." In A. Woodfield (ed.), *Thought and Object*. Oxford: Oxford University Press. 153–206.

Stich, S. 1983. *From Folk Psychology to Cognitive Science: The Case Against Belief*. Cambridge, MA: MIT Press.

Stich, S. 1986. "Are Belief Predicates Systematically Ambiguous?" In R. Bogdan (ed.), *Belief*. Oxford: Oxford University Press, 119–147.

Stich, S. 1990a. "Autonomous Psychology and the Belief-Desire Thesis" In W. Lycan (ed.), *Mind and Cognition*. Oxford: Blackwell, 345–361.

Stich, S. 1990b. "Dennett on Intentional Systems" In W. Lycan (ed.), *Mind and Cognition*. Oxford: Blackwell, 167–184.

Stich, S. 1990c. "Building Belief: Some Queries about Representations, Indication, and Function" *Philosophy and Phenomenological Research* 50 (4): 801–806.

Stich, S. 1990d. *The Fragmentation of Reason: Preface to a Pragmatic Theory of Cognitive Evaluation*. Cambridge, MA: MIT Press.

Stich, S. 1991. "Narrow Content Meets Fat Syntax" In B. Loewer and G. Rey (eds.), *Meaning in Mind: Fodor & His Critics*. Oxford: Blackwell, 239–254.

Stich, S. and S. Nichols 1992. "Folk Psychology: Simulation or Tacit Theory?" *Mind and Language* 7 (1-2): 35–71.

Stich, S. and S. Nichols 1995. "Second Thoughts on Simulation" In M. Davies and T. Stone (eds.), *Mental Simulation*. Oxford: Blackwell, 87–108.

Strawson, P.F. 1959. *Individuals*. London: Methuen.

Strawson, P.F. 1979. "Perception and Its Objects" In G.F. MacDonald (ed.), *Perception and Identity: Essays Presented to A.J. Ayer with his Replies to Them*. London: Macmillan Press Ltd, 41–60.

Strawson, P.F. 1985. "Causation and Explanation" In B. Vermazen and M. Hintikka (eds.), *Essays on Davidson: Actions and Events*. Oxford: Clarendon Press, 115–135.

Strawson, P.F. 1992. *Analysis and Metaphysics*. Oxford: Oxford University Press.

Tanney, J. 1995. "Why Reasons May Not Be Causes" *Mind and Language* 10 (1-2): 105–128.

Taylor, C. 1964. *Explanation of Behaviour*. London: Routledge and Kegan Paul.

Taylor, C. 1985. *Human Agency and Language: Philosophical Papers Vol. 1*. Cambridge: Cambridge University Press.

Taylor, C. 1989. *Sources of the Self: The Making of the Modern Identity*. Cambridge: Cambridge University Press.

Tenant, N. 1984. "Intentionality, Syntactic Structure and the Evolution of Language." In C. Hookway (ed.), *Minds, Machines and Evolution*. Cambridge: Cambridge University Press, 73–103.

Tomasello, M and J. Call 1997. *Primate Cognition*. New York: Oxford University Press.

Tuomela, R. 1990. "Are Reason-Explanations Explanations by means of Structuring Causes?" *Philosophy and Phenomenological Research* 50 (4): 813–818.

Van Fraassen, B.C. 1977. "The Pragmatics of Explanation" In R. Boyd, P. Gasper and J.D. Trout (eds.), *The Philosophy of Science*. Cambridge: MIT Press, 317–327.

Van Fraassen, B.C. 1981. *The Scientific Image*. Oxford: Clarendon Press.

Van Gulick, R. 1990. "Functionalism, Information, and Content" In W. Lycan (ed.), *Mind and Cognition*. Oxford: Blackwell, 107–129.

Vermazen, B. 1986. "Testing Theories of Interpretation" In E. Lepore (ed.), *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*. Oxford: Basil Blackwell, 235–244.

Vermazen, B. and J. Hintikka (eds.). 1985. *Essays on Davidson: Actions and Events*. Oxford: Clarendon Press.

Von Eckardt, B. 1993. *What is Cognitive Science?* Cambridge, MA: MIT Press.

Wagner, S. 1996. "Teleosemantics and the Troubles of Naturalism." *Philosophical Studies* 82: 81–110.

Whiten, A. 1996. "When Does Smart Behaviour-reading become Mind-Reading?" In P. Carruthers & P. Smith (eds.), *Theories of Theories of Mind* Cambridge: Cambridge University Press, 277–292.

Williams, B. 1973. *Problems of the Self*. New York: Cambridge University Press.

Williams, M. 1977. *Groundless Belief: An Essay on the Possibility of Epistemology*. Oxford: Blackwell.

Williams, M. 1998: *Wittgenstein, Mind and Meaning: Toward a Social Conception of Mind*. London: Routledge.

Wittgenstein, L. 1953. *Philosophical Investigations*. Oxford: Basil Blackwell.

Wittgenstein, L. 1958. *The Blue and Brown Books*. Oxford: Basil Blackwell.

Wittgenstein, L. 1967. *Zettel*. Oxford: Basil Blackwell.

Wittgenstein, L. 1969. *On Certainty*. New York: Harper Torchbooks.

Wittgenstein, L. 1977. *Remarks on Colour*. Oxford: Blackwell.

Wittgenstein, L. 1980a. *Culture and Value*. New York: Harper & Row.

Wittgenstein, L. 1980b. *Remarks on the Philosophy of Psychology Vol. I*. Oxford: Basil Blackwell.

Wittgenstein, L. 1980c. *Remarks on the Philosophy of Psychology Vol. II*. Oxford: Basil Blackwell.

Wittgenstein, L. 1982. *Last Writings on the Philosophy of Psychology: Preliminary Studies for Part II of the Philosophical Investigations Vol I.* Oxford: Blackwell.

Woodfield, A. 1976. *Teleology*. Cambridge: Cambridge University Press.

Woodfield, A. 1982. "On Specifying the Content of Thoughts" In A. Woodfield (ed.), *Thought and Object*. Oxford: Oxford University Press, 259–297.

Wright, C. 1980. *Wittgenstein on the Foundations of Mathematics*. London: Duckworth.

Wright, C. 1998. "McDowell's Oscillation" *Philosophy and Phenomenological Research*, 58 (2): 395–402.

Zemach, E.M. 1991. "Perceptual Realism, Naive and Otherwise" In E. Lepore and R. Van Gulick (eds.), *John Searle and His Critics*. Oxford: Blackwell, 169–179.

Index

ABC learning 98
action
  could make sense of 141
  depends on the existence of logical roles and representational contents 25
  explanation of 140, 142
  free 141
  rationalisations of 145
  relation between reason and 146
action sentences
  logical form of 147
actions
  mysterious connection between reasons and 160
affordances 84, 111
  explanatorily superfluous 85
  metaphysically extravagant 85
  representations about creature-relative 85
agency 161
agent 143
  too little authority to 144
  too much authority to 143
agents
  and their goals 161
ahistorical replica 108
ahistorical replica problem 107
AI laboratory 107
Akins 56, 81, 83, 84
Alice in Wonderland 80
analogue form
  sensory information is received in 56
analytic philosophers 87, 171
analytic-synthetic distinction 15
analytical tradition in philosophy 3
animal concepts
  deploy our concepts to describe 91
  pin down the content of 89
anomalous monism 137, 140, 151
  argument for 153
  choice is fatal to 153
  extension reply is bad news for 160
  is to be a stable doctrine 140
  metaphysics of 140
  no argument for 140
  not truly physicalistic 154
  plausibility of 150
  threatens the stability of 160
anomic characteristics 158
anomic dimension 158
Anscombe 163
anthropocentric 108
anthropomorphism 98
anti-realist 104
anti-reductionists 171
Aristotle 25, 97, 160–161
Armstrong 83, 91
artificial intelligence 20, 37
association 97, 98
associationism 98
associations 99
associative engines 97
asymmetric dependencies
  two-stage account of 47
asymmetric dependency relation
  cannot be identified with a Humean causal relation 46
  must be characterised non-lawfully 46
  of dominance 46
asymmetric dependency theory 6, 46, 47
  becomes viciously circular 49
  might provide a solution to the naturalisation problem 48
atomic structures 110
atomism 39
auditory stimuli 81
autonomous agents 111
autonomy thesis 93
  rejection of 93
axiomatic base 127
axioms
  finite set of 127

background knowledge 50, 53, 81
Baker 4, 44
Bechtel 173
bee dance 68
can malfunction 74
behaviour
as a product 106
as a process 106
as psychologically identical 40
conception of 105
explain another's 162
general pattern of a person's 143
makes sense of 166
Millikanesque 107
paradigm cases of 62
two understandings of 106
behavioural ecologists 130
behaviourism 3, 98
behaviourist 108
behaviourists 2
belief 2
content of 31, 134
detailed theory of 135
ordinary concept of 109
beliefs 1
as-intentional-attitudes 110
as-propositional-attitudes 109
realists about 33
Bermúdez 93, 101
Bernier 41
binary 50
binary decision making 50
biological artefacts 65
biological devices 67
lowest level of operation 72
naturally selected 77
reproducible character of 65
biological explanation
obstructive to 85
biological function 55, 63
biological functionality 77
biological niche 76
biological norms
must not be confused with
statistical norms 69
biological practice 75
biological purpose 63
biological sciences 181
biological strategy
prospects of 61
biologically directed responses 84
biologically purposive 103
biologically useful 55
biology 168
biopsychologist 105
behaviour of interest to the 77
biosemantic project 80
biosemantic theories
avoid the problems of distality, misrepresentation, and transitivity 69
biosemantic theory 79
has encountered the strong criticism 70
of intentional content 113
of intentionality 7, 87, 138
unlike other naturalistic theories of content 69
biosemanticist 79
biosemanticists 79
private war between 75
biosemantics 59, 62, 91
ambitious 78
attractive in light of 107
can provide a naturalistic account of intentional content 63
choice between high and low 75
contrast exists even with
consequences for the causal efficacy of content 70
consequence of adopting 107
essential feature of 77
teleofunctional theories of
indicator functions 69
high church 78
low church 76
Millikan's version of 76
misunderstandings and objections to 63
modest 7, 61, 84, 87, 91, 102, 103, 111, 138, 178
modest nonconceptual 60
modest version of 80, 86
non-referential 61
non-truth conditional 61
not a sufficient reason for rejecting 108
obvious virtues of 70
opponents of 107
philosophers who are suspicious of 107
truth-conditional versions of 81

black box 11
Blackburn 15, 18, 166
blended response 96
Block 37
bodily movement 106
bodily movements 105
Botterill 12
Boucher 122
Boyd 27, 178
brain
  is a necessary condition for conceptualisation 174
  right and left hemispheres of the 173
brains
  are indeed learners or conceivers 172
  are not themselves thinkers 173
  compare our conceptual abilities unfavourably with those of 172
  conceptual abilities of 173
  do not even interpret intentional icons 173
  have no form of life 173
  information processed by 172
Brentano 25, 67, 103
Call 121
caloric fluid 19
caloric fluid theory 167
capacity argument 76
Carruthers 91, 122
categories
  innate 15
causal account
  Kripkian-style 47
  non-naturalistic 47
causal analysis 142
causal approach
  an advantage of 42
causal conditions
  absence of 141
causal connection 45, 49
causal explanations 154, 156
  in the impersonal domain 163
  in the natural sciences 163
  of one and the same event 106
  our everyday 159
  used in the mechanical sciences 156
causal impotence 154
  raises the spectre of 39
causal impotency problem 31–32, 41
causal interaction 163
causal investigations 145
causal law 148, 149
causal laws 32, 45, 144, 152
  need not be exceptionless 45
  trade in types 42
causal powers 107
causal process 106
causal processes 24
causal relation 45, 144–146, 154
  because as indicating a 144
causal relations 42, 44, 48, 140, 154, 157
  are brute facts 156
  are extensional 154
  are strictly law-like 158
  between events 151
  between events in extension must be law-like 160
  extensionality of 157
  full explanation of 154
  lack an essential duality 44
  nature of 160
  not accounted for by Davidson 154
  strict nomological view of the nature of 158
  to identify 145
causal roles 105
causal talk 134
causal theories 42
  are riddled with problems 43
  bane of 134
  cannot account for the normative aspects of content 49
  crudity of 43
  lack at least one crucial relation 48
  leave no room for error 45
  leave no room for normative notions 45
  purely nomological 48
  really interested in distal stimuli 44
  rules out the possibility of misrepresentation 44
  unable to explain content 48
causal-explanatory 39, 107
  in a Humean sense 107
causalism 43

dominant in today's analytic philosophy of mind 139
not damning to 144
primary motivation for 160
reductive versions of 140
self-evident status of 141
causalist account
are a purely descriptive 45
third-personal 144
causalists
non-reductive 148
causality 135, 145
knowledge of 161
causally connected 45
causally related 44
causally relevant properties 152
causally unrelated 142
causation
account of 161
basic notions of 161
brute instances of 145
Humean view about the nature and explanation of 148
nature of 154, 157
original notion of 163
cause
criterion for identifying the 145
modern sense of 161
cause-law thesis 157
a priori nature of 158
causes
and their modes of explanation 161
Aristotelian 161
hypothetical judgements about 144
in the mechanistic sense 107
inside the skin of the agent 34
ceteris paribus expectations 163
Chater 89, 91, 92
Cheshire Cat 80
Child 119, 146
Chrisley 94, 96
chromosomes 66
Churchland 4, 9, 13–16, 18, 19, 104, 166–167, 169, 171–173, 176–178, 180
Churchlands 11, 13, 16, 32, 104, 166, 169, 170, 174, 175, 179
are committed representationalists 176
Clark 20, 27, 94–97, 100, 111, 125
classical architectures
deemed to be biologically implausible 28
classical cognitivism
aspirations of 171
cannot explain representational content 104
connectionist theories of information processing threaten 104
epitomised as some form of *lingua mentis* 104
Fodor's version of 6
inability to provide a naturalistic account of representational content 29
classical cognitivist 171
account of theorising 172
manoeuvres open to 29
classical cognitivist-connectionist debate 177
classical cognitivists 21, 96
commitment to representationalism 21
committed to the view that agents are organisms with internal, causally efficacious mental representations 21
classically computational 34
theory of mind 26
Cluedo 2
co-referential substitution 114
co-variance 42
connection of perfectly reliable 52
co-varied 81
co-varies 54
co-extensive 71
cognition
contentful aspects of 180
formal, or syntactic, features of 32
nature of 172
pervasive criterion of 95
cognitive dynamics 79
cognitive map 20
cognitive processes
strongly supervene 35
cognitive processing
traditional views of 110
cognitive psychologists 171

cognitive psychology 89, 92
cognitive revolution
eliminative in character 19
cognitive science 5, 111, 181
advance on a pure or minimalist version of theory-theory 22
as a multi-disciplinary investigation 27
core assumption of 27
cutting-edge work in today's 86
harmonious relation exists between intentional psychology and 110
has historically assumed the truth of the traditional account of symbolic processing 27
in the throws of an internal revolution 19
laws would not involve contents at all 40
means of understanding ourselves 3
needs to posit linguaform representations 29
only deploys causal explanations 32
recent developments in 104
requires representations 22
revolutionary advances in 19
should at best only trade in psychological laws of a narrow kind 32
threaten the status of commonsense psychology 103
cognitive scientists 111, 175
tendency of many 91
cognitive suicide 4, 173
coherence
charge of a lack of 167
high premium is placed on 168
issue of 168
patterns of internal 128
test of unity or 168
coherent pattern 129
in the utterances of another 127
coherentism 169
in its weakest form 15
minimal 169
stronger 16
coherentist epistemology 15
as a means to motivate eliminativist concerns 16
can motivate a secure eliminativism 8
minimal 8
weakest version of 15
coherentists 15
common stimuli 116
common world 85, 116
commonsense psychological terms
meaning of 14
commonsense psychology 1, 111, 137, 145, 159 (see also folk psychology, propositional attitude psychology, reason-giving psychology)
charged with being a misleading theory of mind 1
classical cognitivism fails to defend 104
critique of 110
crucial to understanding 112
debate about the future of 6
defender of 104
defenders of 111
eliminativist threat to 5
explanations it provides are not causal explanations 166
fate of 104
greatest perceived problem with 166
has a meaning-holistic character 14
has a privileged place in our self-understanding 17
has ontological import 11
hasn't changed much 166
ideal physical explanations will never replace 159
in the process of being displaced by a superior theory 20
in the process of being undermined 19
is a degenerating research programme 14
is a kind of proto-scientific theory 11
is a potentially dispensable conceptual scheme 10
is a species of theory 9
is a theory 13

is a type of research-programme 166
is an irredeemably unscientific practice 165
is at least theory-like, 14
is committed to the following claims 105
is context-sensitive 36
is the best explanation of our actions 1
key objection to 110
lack of fit with the rest of science 166
limits of 166
makes appeal to contents 34
means of naturalising 86
methodological concerns of 12
might be on the list of potentially dispensable conceptual schemes 16
misunderstanding of the very nature of 4
need not be thought of as a theory 5
need to employ the categories of 176
no special defence of 16
no special or sacred status 17
not thought to fare well 166
not threatened by the revolutionary developments of cognitive science 87
nothing more than a practical or heuristic device 18
of key philosophical importance 4
out of step with natural science 4, 167
posits inner states 108
posits inner entities 11
possibility of naturalising 61
potential revision and possible elimination of 15
prospect of naturalising 31
prospects for naturalisation 165
reputation of 110
role in underwriting a wide variety of key practices 4
scientific status of 3
scientific version of 40
scientific vindication of 26
serves our quotidian needs 166
standard scheme of 12
status and future of 3
superseded by a vastly superior theory 18
to be possibly false 170
undermined by advances in cognitive science 165
will never be, an integral part of the rigorous scientific world-view 165
will not naturalise 10
communication 120, 126
comparative psychology 89, 90
as animal cognitive science 89
experimental methodologies of 89
compatiblist 140
completed neuroscience 19
compositional rules 127, 130
compositional speech 127
compositionality 95
computational implementation 36
computational processing 36
classical account of 41
computational states
not intrinsically about anything 32
computational transformations 177
computationalism 39
mechanistically respectable 33
computationalists
draw mere analogies between our reasoning processes and computational processes 42
computer metaphor 26
computers
are formal machines 32
conceiving 100
concept 90, 103
each higher level 89
concept ascription 92
concept development 174
concept learning 21, 49
concept of belief 116, 121
having a 110, 117
to form a 119
concept of cause
anthropomorphic basis of 162
has its roots in our experience 161
is single-tracked 163
is *sui generis* 163
original home of 162

concept possession 89
concept use
not sufficient evidence of 99
concept user 93
conception
change in our 18
concepts 90, 97
a creature will necessarily need to deploy 98
constitutive 89
development of 120, 124, 179
development of full blown 102
emergence of 124
master of 117
mastery of 121
mastery of some basic 90
must be innate 122
must not assume that creatures must be using 100
no adequate translation into our 91
pre-established 117
pre-existing 122
primitive 122
standard argument that particular animals must be using 92
standard theories of 89
three main theories of 89
ultimate origin of 124
understanding the nature of 90
which we are simply unable to describe 91
conceptual analysis
of purposive behaviour 62
conceptual boundaries 90
conceptual capacity 94
conceptual categories
are essentially revisable 16
conceptual change
large scale 173
understand the nature of 174
conceptual connections 148
conceptual content 88, 89, 170, 178
developmental explanation of 93
does not originate in the brain 174
conceptual contents 109, 122
of non-verbals 92
unexplained postulation of 123
conceptual development 123
an explanation of 101
minimal condition for 116
phylogenetic theory of 93
conceptual error 119
conceptual identity 176
conceptual judgements 90
conceptual norms 118
conceptual perspective
develop a 119
conceptual point of view 118
conceptual responding
genuine 117
conceptual responses 175
conceptual role semantics 37 (see also functional role semantics, meaning)
read in a functionally acceptable fashion 37
conceptual scheme 10, 11
nothing guarantees the truth of 15
our own 80
conceptual schemes 16, 169
radically different 131
virtues and vices of 166
conceptual terms
are meaning-holistic 15
conceptual thinking
genuine hallmarks of 100
conceptualisation 97
important aspects of 173
conceptualising 170
conceptualist tradition 97
conditional probability 50, 54
conditional statements 42
connectionism 98, 176, 177
challenges our faith in our commonsense psychology 28
departs crucially from the traditional understanding of information processing 20
makings of the champion theory 19
merits of 27
threat to classical cognitivism 6
threatens the prospects of a scientific vindication of propositional attitude psychology 29
connectionist 173
network of nodes 20
nonconceptually minded 99
theories of information processing 7

threat to expansive classical cognitivism 27
connectionist architectures
attack on the possibility of using 95
connectionist challenge
to the received view of the nature of information processing 27
connectionist networks 175
are capable of learning 175
are capable of making cognitive responses 175
become sensitive to microfeatures of situations 99
behaviour of 176
blended responses are deployed in 96
capable of 'learning' from experience 28
don't make transformations over symbolic structures 125
interpretation of 100
learn by co-ordinating their responses to standards 175
make an intelligent response to a novel'blended' situation 99
meet 'real time' psychological constraints 28
microcognitive architecture of 98
need nonconceptual content 96
not rigid or pre-programmed 175
one of the great virtues of 99
regimes imposed on 176
renown for their pattern recognition abilities 97
resemble neural configurations 28
connectionist paradigm 177
connectionist processing 100
non-systematic character of 96
connectionist theories 179
neuro-computationally based 172
pose a genuine threat to commonsense psychology 6
pose a threat to classical cognitivist theories of information processing 6
real difference between classical cognitivist and 177
connectionist vehicles
need nonconceptual content 110
connectionist-classical cognitivist debate 26
connectionists
do not accept that basic information processing involves symbols 27
have to do more than provide a better story about information processing 29
must establish that their theory is incompatible with classical accounts of symbolic representation 29
no less committed to the project of cognitive science 27
conspecifics 121
constant conjunctions 161
construction-theoretic content 93
consumer device 77
consumer devices 70
consumer mechanism 68
content 158 (see also intentionality, meaning, propositions, representations)
ability to survive variations in functional role 39
account of the nature and origin of 48
appropriate to full-fledged beliefs or desires 79
ascription of 128
atomism about 39
bivalent character of 44
cannot be causally efficacious 36
cannot depend, even in part, on functional role 38
cannot strongly supervene on the internal states 36
is functionally discrete 105
causal roles in virtue of 33
causal theories of 6
conservative theories of 171
contributes nothing to the production and causal explanation of action 33
denotational theory of 39
determined by its extension 36
enters into our natural history at a very early phase 79
explanatorily redundant 32
explanatory and metaphysical role for 107

externalist view of 36
has an internal level of implementation 34
has more than one aspect 37
have not understood 170
identity conditions of 38
identity of propositional 38
if mere syntax determined 36
implemented by means of brain activity 34
internalist account of 36
is determined solely by extension 39
is epiphenomenal 33
is not causal-explanatory 107
is related to different kinds of psychological attitude 38
is unique among higher-order phenomena 35
is what makes debate possible 170
is world-directed 37
logical geography of 114
mapped in a unique way 33
naturalistic account of 5, 165
nature of 170, 171, 177
non-traditional account of 171
nonconceptualist approach to 87
normative notion of 67
of some kind 77
only causal patterns matter to 48
presuppose the naturalizability of 41
purely internalist accounts of 37
quantify over 34
questions about the ontological status of 158
relation between computational syntax and 33
representational 42
socially constituted aspects of 180
strongly supervenes upon brain states 34
strongly supervenes upon computational states 34
successor proposals about the nature of 176
theory of propositional, representational 171
traditional understanding of 101
truth-evaluable 78
understood in neuro-computational terms 172
until we stop reifying 33
contents
are identical if 38
must exist 170
should not be thought of as fused 39
context of inquiry 181
context of use 126
Copernican cosmology 18
Copernican revolution 19
Copernicus 19
copying 65
correctness conditions 101, 102
nonconceptual content has 101
not best understood as truth conditions 86
which are not truth conditions 87
which underpin intentionality 86
correspondence 135
correspondence theory of truth 79, 86, 135
counterfactual 46
counterfactuals 64
covering law 148, 157
covering laws 157 (see also laws)
Cram 47
Crane 1, 3, 23–24, 101
creature-relative property 85
crude causal theory 42 (see also causal theories)
cultural features 180
culturally relative 47
culture 180
Cummins 37, 67
Cummins'
kind of functional type-identification 67
Cummins-functions 67
current dispositions 69
Cussins 17, 88, 93–94, 96, 102, 177–178
Cussins'
notion of nonconceptual content 102
Darwin 71, 103
Davidson 109, 110, 112–113, 115–116, 118–119, 122–123, 126, 128–162
Davidsonian 81

Davidsonian dogma 162
Davidsonians 159
Davies 66, 67
Dawkins
    Marian Stampe 98
decision making 21, 26, 99
decision making processes
    a naturalised account of 26
decision theory 21, 26
decisions 97
definitional boundaries 89
definitional theory 90
demonstrative 89
Dennett 65, 73, 98, 104, 117, 125, 130, 159, 170
denotation 39
denotational accounts 41
denotational content 40
    a naturalistic account of 41
desire 2
desires 1
developmental explanation 93
developmental psychology 92
digital information 56, 58
    property-specific character of 57
digitalisation 56, 57
discriminating responses
    without relying on concepts 97
discrimination 100
    in animals 89
discriminative mechanisms
    if not built in 123
discriminatory capacities 85, 92
    nonconceptual 92
discriminatory capacity 85
disjunction problem 44
disposition
    mechanical 63
dispositional 102
dispositional response
    complex patterns of 107
dispositions 64, 108
    insufficient to underpin intentionality 108
distal stimulus 116
distality problem 43, 52, 76
domestic anthropology 108
Dretske 49–59, 69, 81, 99, 105–106, 162
dubbing ceremonies 47
Ducassé 157
Dummett 80, 115–116
Dupré 3
dysfunction
    only makes sense against the background of an understanding of proper functioning 63
effective environment 111
Elder 74, 85
eliminative materialism 169
eliminativism 10, 112, 171, 180
    another argument for 165
    assumptions which ground 10
    becomes incoherent 169
    becomes self-defeating 169
    best reason for rejecting 169
    coherence of 61
    epistemological commitments of 171
    focuses on 169
    insufficient to explain scientific theorising 165
    intellectually incoherent 8
    is incoherent 165
    left in the awkward position 171
    most convincing argument for 166
    most stable, and threatening, form of 5
    must be answered 166
    plan B comes to grief 168
    proposed strategy for rescuing 178
    two strategies of 5
eliminativist 165, 168, 169, 171
    adherence to a coherentist epistemology 167
    goal 10
    strategy A 103, 104
eliminativists 104, 168, 170, 171, 180
    accept a broadly coherentist epistemology 169
    best hope of convincing us 167
    cannot lean on a naturalised account of nonconceptual content 178
    have heavy commitments 172
    make use of the notion of representation 177
embedded cognition 7, 111
embodied ognition 7, 102
empirical foundationalists 15

empirical hypotheses 127, 131
empirical investigations 122
empirical work 120
empirically adequate 168
empiricism 3
every last residue of 16
enculturation 120
endorphin molecules 73
ends
an organism's higher 74
divide between higher and lower 73
higher level 73
various higher level 75
engineers 50
environment
internal and external 78
epiphenomenal 151
epiphenomenalism 34, 140
epistemic gap 144
epistemically transparent 143
epistemological picture 169
etiological 64
Evans 87, 94, 101–102, 114, 129
event 147
capture all the aspects of an 160
contentful aspects of an 158
no sense in imagining the very same 151
physical characteristics of 149
psychological properties of 152
refer to one and the same 148
events 147
are causally related 158
are causally related in extension 154, 158
are individuated by their spatio-temporal occurrence 148
are ontologically basic kinds 148
are particulars 147
coincide spatio-temporally 147
connect to one another 158
description of 148
designate the relevant 157
distinguished from objects 147
efficacy of 152
enter into causal relations 148
in extension 154, 157
individuate 147
individuated in terms of their spatio-temporal occurrence 151
individuation of 151
method for individuating 147
occur in space-time 148
properties of 157
re-describe 148
refer to 147, 148
regarded as ontologically basic kinds 147
special class of internal 143
stand in casual relations 154
types of 151
underlying theory of 148
everyday psychological concepts
as reducible to visible stimuli and responses 2
evolution 107
does not programme systems in advance 28
evolutionarily excessive 84
evolutionary biology 69, 103
evolutionary history 70
evolutionary pressures 55
exemplar theory 89, 90, 92
existential quantification 33
exotic subjects 108
expansive classical cognitivism 6, 25, 26, 59, 111, 180
epitomised by the language of thought 28
expansive classical cognitivist 32
expansive classical cognitivists 171
experience 102
nonconceptual account of 101
changes in our verdicts about 16
quality of 16
we conceptualise our 15
experiential judgements 17
experiential responding 102
explanation
desire and belief in 142
of a device's capacities 75
of human behaviour 137
of why a device was originally selected 71
of why a response mechanism proliferated 70
order of 72
types of 156
explanations by cause 156
explanations by law 156
explanatory exclusion 106

explanatory gap 125
explanatory law 158 (see also laws)
explanatory priority
one belief/desire pair has 144
explanatory regress 124
extension reply 140, 154, 157, 158
must be read weakly 140
strong 159, 160
stronger form of 158
weak 155, 160
extensional 128, 155
extensions 38
externalism 36, 39, 40
a necessary ingredient in a sound propositional attitude psychology 36
led to accept some form of 41
to abandon 36
false belief task 121
Field 42
final cause 77
first-person authority 144
Fodor 4, 21–26, 29, 31–34, 36–44, 46, 47–49, 70–72, 76, 91, 95–96, 103–105, 107, 122–125, 129, 153, 158
Fodor's
commitment to physicalism 34
larger project to explain central cognitive activities 49
folk physics 159
folk psychological terms 2
folk psychology 1, 9, 104, 108 (see also commonsense psychology)
folk science 11
formal properties
physically implemented 24
formal substitutions 95
formality condition 32
fortuitous benefit 62
Frankenstein monster 107
Frege 23, 25
Fregean sense 25
function 62 (see also Cummins-function, proper function)
function (mathematically conceived) 40
functional analysis 72
Cummins-style 75
lowest level of description in 72
functional indeterminacy 58
functional roles 36
functional role 23
functional role semantics
purist versions of 37
functional specification 73
functional theory 38
functionalism 22
functionalist 108
functionalists 171
Gadamer 179
Galileo 19, 161
gamete cells 66
Gasper 178
generalisations 97
involving re-identifiable substances 97
generality constraint 94, 114
genes
paradigm examples of a lower-order reproductively established family 66
genetic mechanisms 66
genetic transmission
in sexually reproducing organisms 66
genetic variation 82
ghost channel 51
Gibson 84, 111
Godfrey-Smith 43, 62, 77
Gómez 121, 134
Gordon 109, 161
graceful degradation 28
grammatical syntax 126
Grice 42
Hannan 19
hard-wired 79, 96
Harman 104
Heidegger 119
hermeneutic 137
hermeneutic circle 135
Heyes 89, 91, 92
hidden properties 97, 98
historical conditions 106
historical environment 71
historical explanations 71
are transparent 71
historical pedigree 69
historical standard 107
historically normal 81
history of ideas 18
holism 38, 135

holistic circle
obtaining between belief and meaning 135
holistic complexity 115
homestatically regulated 78
Homo Sapiens 124, 173, 174
homonomic laws 149
homuncular discharge 75
Honderich 151–153, 158
Horgan 27, 108, 161–162
horizontal axis problem 43
Hornsby 144
human condition 168
human contexts of inquiry 181
human nature
complete account of 181
scientific account of 167
human society 181
Hume 139, 157, 161–163
Humean causation
weak version of 148
hypothesis-confirmation 21, 22
icon consumption 69
dynamics and norms involved in 69
ideal physical explanations 159
folk physical explanations explain less than 159
recognising the intensional nature of159
ideal physical properties 158
ideal physics 147, 158
differences between the explanatory concerns of commonsense psychology and 159
kinds of explanations one can get from 159
language of 148
metaphysical bias in favour of 159
predilection for 160
identity theories
mind-brain 173
identity-theorists 171
illocutionary acts 38
images 99
imagistic 99
impersonal analysis 145
implementation levels 107
implementation mechanisms 34
implicit functional definitions 2
imprinting
phenomenon of 46
indexical 89
indication 54, 69
Dretske's original notion of 50
original Dretskean sense of 59
relation of 52
indicator control function
no principled way to uniquely specify 58
indicator function 59
by having a biological function 55
indicator functions (see also proper function, teleofunction)
representations have their own intrinsic 55
inference making 95
inferential 24
inferential connections 38
inferential patterns 24
inferential processes 26
inferential roles 37
information 81
can be nested in a signal 57
cannot be false 54
carried by a particular causal route 52
causal vehicle of 51
conceived as uninterpreted strings of code 32
flow of 51
is lost 56
is processed 56
Popperian creatures use stored 98
property-specific 57
provided by the sensory modalities 176
quantity of 50
share the same 51
transmission of 50
which a signal carries 52
information loss
vivid illustration of 56
information processing 111
algorithmic 20
by means of logical rules 20
classical account of 26, 32
connectionist account of 96
connectionist approaches to 87

machinery 32
models of 20
serial symbol manipulation
account of 26
uses meaningless symbol-strings 21
information scientists 50
information sensitivity 59
information states
truth-conditional 101
information-theoretic 69
account of indication 58
accounts of content 49
approaches are distinct from causal theories 51
notion of indication 59
informational content
assigned a function 55
distinction between representational content and 49
Dretskean account of 54
equates sensory content to 54
senses carry 177
technical definition 50
informational links
distinguished from causal links 51
informational sensitivity 81
distinction between intentional direction and 82
should not be understood in Dretskean form 81
informational state
nonconceptual 88
informational states
postulated base of nonconceptual 101
informationally sensitive 81, 82, 97
to certain features of the world 82
innate
categories 15
inner languages 123
no concepts need be 93
innate concepts
ontogenesis of 123
innate module 122
inner functional states
are related to representations 23
serve as representations 23
inner mechanisms 111
inner state tokens 46
inner states
reified 124
semantically interpretable 108
syntactically structured 32
instrumentalism 159
intellectual tradition 174
intellectualist legend 141
intensional 128, 155
descriptions are 80
intensional indeterminacy
of our selectionist explanations 80
intensionality test 114, 119
intentional content 5, 72, 105
ascription of 81
biological norms which underwrite 86
bivalent character of 86
can be best understood in relation to full blown propositional content 61
naturalistic account of 86
nonconceptual character of 86
which proper function, matters to 75
intentional contents
are not things or events 107
intentional icon
content appropriate to an 78
content of 74, 79
has both indicative and imperative aspects 68
ought to point to 72
paradigm example of 68
intentional icons 67, 75, 77
assigns too much content to 76
content of 85
direct proper function of 69, 70
do not have Fregean sense 85
do not represent objects 84
minimal conditions for 77
proper function of 73
to state the content of 80
intentional idiom 149
intentional object
we mis-describe the 90
intentional psychology 105, 110
in harmony with cognitive science 110
intentional responses 110
intentionality 25, 95, 102, 103, 107, 108, 112, 154, 158
account of 119, 124

built-in 49
focus of 81
naturalised theory of 112
nonconceptual account of 7, 92, 109
of an individual's thoughts 47
of complex responses 74
original, medieval image of 81
plays a role in underpinning propositional judgements 86
pre-existent 47
pre-existing 54
the most rudimentary forms of 61
theory of nonconceptual 123
thought of in terms of directedness. 103
unsafe to infer 77
intentionally directed 81
at the same type of objects 82
interest-sensitive discourses 159
internal properties
of the agent 36
internal representations
some animals are capable of using 98
internal-external dichotomy 77
internalism
more than a merely syntactical account of 36
not compatible with commonsense psychology 37
pure forms of 37
internalist 36
interpret
meaning of subjects' utterances 135
interpretation 110, 126, 127, 130, 134, 135, 144, 146
context of 119
essentially normative nature of 153
first move in 103
first stages of 133, 135
intersubjective 179
is essentially and irreducibly normative 113
is to be a possibility 136
normative account of 138
of meaningful utterances 135
of speech and thought 131
of the speech and propositional thought 129
possibility of 130
potential gap during 135
social context of 135
subjects of 122
to reassess 163
interpretationalism 113, 124, 126
non-reductive nature of 134
not at odds with recent developments in cognitive science 124
interpretationalist 5, 125, 165
understanding of the ascription of propositional content 113
interpretationalist account 136
interpretationalists
hope to make peace with naturalism 139
interpretations
rigorous linguistic 115
interpretative framework 136
interpretative investigation 109
interpretative practice 138, 179
first moves in 176
interpreter 114, 119, 123, 128, 131, 132
interpreters
make some initial assignments 133
interpreting
only by triangulating 135
process of 122
role causality plays in 134
sophisticated thought 136
interpreting subjects 117
intersubjective 124
intersubjective standard 116, 179
introspection
cannot be a neutral judge 17
introspective awareness 161
intuition 108
irrational 33
Jacob 43
Johnston 143, 150
judgement
can reciprocally pass 117
capacity for propositional 80
must be characterised in terms of concepts 88
nature of 115

not equivalent to 103
with a propositional content 88
judgement making 101
judgements
according to prototype theory 90
conceptually based 92
making conceptual 112
necessary conditions for the making of 100
Kant 157, 161
Kim 106, 107, 153, 157, 158
kinetic theory 168
know-how 97
Kripke 47
Lakatos 12
language
a peripheral medium of representation 173, 179
interdependence of conceptual thought and 123
logical analysis of 158
written or spoken 124
language learning
infinite regress of 22
language of thought 26, 122
is innate 22
not itself a learned language 22
offers no real projection for commonsense psychology 29
sentences in the 33
version of computationalism 25
language of thought hypothesis 25–26, 32–34, 91
language users 90, 92
law
once physicists develop a precise 156
law abiding naturalists 45
law of non-contradiction 79, 130
law-like descriptions 156
law-like explanations 156
are merely intensional 160
law-like fashion 158
law-like pattern 153
law-like physical properties 151
law-like properties 151, 157, 158
law-like science 137
lawless nature 152
laws 145, 152 (see also causal law, causal laws)
absence of commonsense
psychological 159
are intensional 154
computational 39
connection between causes and 157
content-based 41
content-based psychological 34
does not essentially concern the strictness of 45
general 3, 137
hedged predictive 33
input-output 105, 107
natural 3
non-strict 153, 154
of a scientific psychology 39
of an ideal physics 140
of cognitive science 32
of our best scientific psychology 33
of the universe 3
psychological 33
strict causal 44
that relate two events, bi-conditionally 156
to be contentfully-construed 33
learning
concepts 21
first concepts, 21
first natural language 22
language 21
our first language 122
learning situation
primitive 116, 121, 133
learning situations 117
Lepore 129, 139
Lewis 2, 14, 151
lingua mentis 91, 122, 123
linguistic policies 47
linguistic policy 48
linguistic practices
depend on the pre-existence of a system of mental representation 49
linguistic use
internal patterns in 127
Loar 12, 13, 25, 37
Lockwood 115
Loewer 46, 139
logical connections 131, 135
logical impossibility 170

logical inferences 79
logical rules 110
logical structure 150
logical thought
develops in tandem with language 124
logical truth 133
logical truth problem 129
logico-linguistic thinking 95
logico-syntactical properties 23
Malcolm 90, 173
malfunction 72
mapping rule 77
march of science 16
assertion of faith in 19
Marcus 91
marine bacteria 58, 85
materialism 169
is beyond question 169
materialist
preference for purely scientific explanations 150
McDowell 101, 122, 129
McGinn 32, 78
McLaughlin 154, 157
meaning 115
complete theory of 135
empiricist view of 14
given in certain types of experience 14
holistic accounts of 38
is generated holistically 14
is manifest in 'use' 38
necessary condition on the possibility of 150
of ordinary psychological terms 14
of our words 38
rule out the possibility of 158
something by our words 38
meaning-is-use thesis
extreme forms of 38
meaningful pattern 130
meaningful speech
no attribution of 130
meanings
are autonomous with respect to use 38
means-end fashion 73
means-end reasoning 25
mechanical mind 24
mechanisms
operation of 73
mechanistic arms race 76
mechano-computational 180
mechano-functions 69
meiosis 66
Melden 141–142, 146, 161, 163
memory loss 130
mental
and the physical do genuinely interact causally 153
anomalousness of 152
cannot be reduced 149
cannot, interact with the physical 152
is dependent on the physical 149
mental causation 6, 153
mental causes 141
mental concepts 2
mental event
occurrence of a 149
mental events 142, 150
mental formulae 23
mental properties
are explanatorily irrelevant 152
mental states 1
mentalistic vocabulary 13
meta-language 127
metaphysical entities
bizarre 85
metaphysical glue 49
metaphysics 1, 140, 143, 146
accurate guide to 158
exclusively science-based 138
inspired by natural science 1, 5, 169
purely scientific 164
methodological solipsism 32, 34, 40
mild realism 159
mild realist 159
Millikan 62–79, 86, 97–98, 105–108, 154
Millikan's
definition of proper functions 62
notion of reproduction 65
Milton 18, 19
misinformation
equivalent to no information 54
misrepresentation
normative assessment of 70
possibility of 158

misrepresentation problem 44, 45, 49
molecular biology 75
monism 160
exhaustive 150
Russellian neutral 154
morphine molecules 73
Mrs. T 130
Müller-Lyer illusion 85, 89
mutually associated 98
mysterious connection problem 140–141, 146, 157, 161, 163
does not arise 163
no longer a puzzle 163
Myth of the Given 101
Nagel 159, 171
naming ceremonies 47
narrative proposal 146
narrow content 39, 40, 41
always difficulties in giving a positive clarification of 40
defined negatively as 40
depends on having an account of denotational content 41
does not concern reference or truth conditions 40
Fodor's technical way of defining 40
is radically inexpressible 40
isn't reducible to a mere syntactic or functional role 40
only way to characterise 40
supposed to be 'another kind of intentionality' 40
narrow contents
enter into the same partial function 40
narrow psychology
was itself a revisionist manoeuvre 40
would miss valuable psychological generalisations 40
natural indicators 58
natural meaning 50
natural sciences 145
natural selection 55, 64, 86
calibrated by 103, 124
does not care about truth 84
functions are products of 65
norms requires appeal to 64
recent episodes of 64
natural sign 54
natural sign indicators 54
natural world
our attempts to understand the 162
naturalisation 5
current thinking about 3
naturalised theory of content
best we are likely to get from cognitive science 8
inadequate to explain scientific theorising 8
naturalism 146
exclusive scientific 8
requires the reduction of intentionality to the non-semantic 47
resistance to 169
naturalistic perspective 145
Neander 66, 71–77, 81
Neander's
definition of proper functions 66
recommended low road 76
neo-Wittgensteinian 142, 161
neo-Wittgensteinians 146
nerve-endings 116
nested information
analytical 57
nomic 57
nesting
is not necessarily apparent 57
neural machinery 104
neural mechanisms 82
account of the activity of 174
neural network 175
neural networks
plasticity of our 179
neural representation 176
neural representations 175
neural response 81
neural responses 81
neural systems 81
neuro-computational 174, 176, 179, 180
neurology 107
neurophysiological challenge 83
neuroscience 104, 179
Newell 20
niche-dependent sensing 111
non-causalist 142
orthodoxy 141
non-causalist approaches 161

non-causalists 146
non-compositional 127
non-objective 96
non-observable mental phenomena 2
non-reductionism 140, 160
non-reductionists 139, 150
non-reductive 34, 102, 140, 163, 165, 169
relationship of dependency 149
non-reductive physicalism 137, 139, 159
non-reductively 135
non-teleological 75
non-theoretical entities and events 2
non-verbal animals
alleged theoretical capacities of 178
nonconceptual 105
discriminatory capacities 92
Evans' distinction between the conceptual and the 101
intentional content 100
intentional responses 102
perceptual responses 100
reading of animal behaviour 100
nonconceptual capacities 124
foundational role of 124
nonconceptual cognition
mechanics of 177
nonconceptual content 7, 87–89, 93–94, 96, 103, 125
a primary motivation for 88
appropriate vehicles of 87
becomes interesting 92
biosemantic theory of 111
cannot be unpacked by means of pregiven notions of truth and reference 102
construction-theoretic account of 93
even in the most basic cases of 101
helps us to understand conceptual development 101
in desperate need of clarification 102
is already thought of as truth-conditional 102
is not symbolically structured 110
is regarded as representational 101
lacks correctness conditions 94
modest biosemantic theory of 123
most appropriate for non-systematic, context-sensitive PDP-processing 96
not equivalent to propositional content 7
object of much serious attention 87
potentially more than one form of 93
pre-objective 94
the link between connectionist vehicles and 96
whether or not there is such a thing as 87
nonconceptual contents 101, 102
do not map onto what we would recognise as objective features of the world 96
have correctness conditions 102
nonconceptual knowledge
would be context-bound 99
nonconceptual responses 92
can be behaviourally influential 88
can be misdirected 103
creatures which are only capable of 95
experiential aspect of 100
nonconceptualist 178
norm of regularity 154
normal conditions 64, 69, 71
normal environment 75, 81
normal explanations 64
are seen as specifying structuring causes 106
normal habitat 71
normative 136, 138, 145, 163, 180
normative competencies 117
normative dimension 102
needed to explain semantics 45
normative factors 176
normative notions 45
normative understanding 72
normativity 132, 154
norms 102, 158
novel scenario 99
novel situation 98
object language 128
object permanence 94

objective
perspective on the world 102
representation of the world 120
objective categories 119
objective features
representative of 119
objective thought 93, 120
objective truth
notions of 119
objective world 119
formation of beliefs about an 116
notion of 116
the emergence of 119
objectivity 95, 119
objects
individuating 147
re-describe 148
way in which we individuate 148
observable causes and effects 2
observable facts 2
observation 161
observation statements 15
Ockham's razor 59, 92
ontogenetic 93
ontogenetic ladder 79
ontological economy 85
opacity 57
is a hallmark of semantic content 57
operant conditioning 56
operational definitions 2
opioid receptors 73
organic chemistry 75
original intentionality 103 (see also intentionality)
ostensive behaviour 120
ostensive reference 121
over-determination 32
Papineau 46, 78
Paradise Lost 18
particulars 42
pattern
locate a coherent 131
of each 'language' 130
patterns
of actual sentential usage 131
of assent and dissent 128
Peacocke 93–94, 101
perceiving 100
perception 102
as a neutral basis for assessment 15
traditional view of 83
perceptual capacities 75, 76
perceptual contents 42
perceptual experiences 88
fine-grained character of 89
perceptual illusions
can influence actions 88
perceptual mechanisms 81
perceptual relativism 89
perceptual response
concepts are inappropriate in a principled description of 88
perceptual systems 82
perfect correlations 81
person theory of humans 167
personal point of view 144
perspective 121
develop the idea of a contrasting 118
on a common scenario 121
recognise the possibility of another 119
perspectives 122
alternative 122
alternative, as opposed to merely other 122
friction between different 116
to recognise contrasting 119
to recognise other 122
phenomenological effects 161
phenomenology 161, 162
phenotypes 66
philosophers of language 38
philosophers of science 15, 178
philosophical assumption 83
philosophy of mind 151
philosophy of psychology 155
phlogiston 19
phylogenetic 93
phylogenetic relatives 120
phylogenetic tree 79
physical 104, 106
relation between the psychological and 150
physical events 140
physical features 180
physical properties 150
physical reduction 107
physical replica 107

physical twins 40
physicalism 3, 138, 153, 165, 169
hard 158
physically identical 35, 105
physically possible world 151
physically possible worlds
unchecked talk of 150
physics 168, 181
physio-chemical 75
physiological 161
physiology 107
place-holding terms 2
plasticity 79
PNCC 140, 151–153, 155–157, 160
confidence in the truth of 156
has attracted a fair deal of scepticism 155
makes space for 158
metaphysical version of 158
read as a physicalist doctrine 154
scepticism about 156
Popper 98, 103
Popperian creatures
are necessarily concept users 98
Positivism 3
positivist
philosophy of science 2, 3
positivistic view 15
positivists
epistemological concerns of 3
possibilities 50
need not be narrowed by the character of the signal itself 53
possible state of affairs 50
possible worlds 149, 150, 151
logically 151 (see also physically possible worlds)
practical knowledge 98
need to invoke concepts when it comes to understanding 97
practical simulation 109
practical syllogism 38
pre-objective 94
pre-verbals
the mentality of 89
primary reason 141
will have more than one 141
primary reason problem 140, 146, 163
no explanatory edge in solving 146
principle of autonomy 34, 39, 107
principle of charity
normative 131
principle of coherence 129, 131–132
principle of correspondence 109, 132–133, 136
principle of the anomalousness of the mental 149
principle of the causal interaction of the mental 149
principle of the nomological character of causality 140, 149, 151 (see also PNCC)
principles of charity 129
privileged access 143
producer device 77
productivity 95
projection rule 68
lower-level, algorithmic 78
projection rules
proximal and distal 78
prompted assent 114
propensity theory 64
appears inadequate 64
future prospects of 64
proper function 72, 81, 85
adapted, derived 74
algorithmic 74
assigns predominance to a device's direct 74
biological device has a 77
current 65
derived adapted 74, 77–78
direct 66, 73, 74, 75, 78, 84
due to its history 69
general 74
historically-based 64
organismic 73, 78
selectionist explanation of 80
proper functional description
level of 74
proper functions 62, 66–67, 69, 105
are essentially normative 63
benefit-based account of 73, 75
correct descriptions of 72
idea that current devices have 65
in statistical terms 76
integrity of the notion of 86
kind of functionality that underwrites 64
logically stacked 73

lowest level of a mechanism's 75
multiple 73
natural selectionist accounts of 66
reach out to the external world 77
standardly complement one another 74
terminology of 72
to understand 64
understood in terms of natural selection 86
property-specificity
rests on a digital informational base 58
proposition
any meaningful sentential assertion can be regarded as a 12
each has its own unique geometry 24
propositional ascription 109
is an essentially normative practice 136
minimal evidence for 114
propositional ascriptions 12, 135, 137
cannot have a closed system for 136
propositional attitude 11, 133
propositional attitude ascription
normative account of 137
propositional attitude psychology 5–6, 11, 25, 105, 110, 137–138, 156, 165, 169–170 (see also commonsense psychology, reason-giving psychology)
a scientifically respectable linguaform version of 31
abandon the prospect of a naturalistic vindication of 39
an interpretationalist approach to 7
attempt to absorb it into a scientific psychology 8
cannot be incorporated into a metaphysics which is inspired by physics 163
cannot be reduced 139
content-ascribing 168
depends on a shared theory about human psychology 12
elimination of 5
has undergone subtle change 166
has special constitutive features 137
interpretationalist account of 138
interpretationalist understanding of 138
is a kind of low-grade theory 1
is anomalous 153
is best understood by interpretationalist lights 8
is fit for eradication 169
limited scope of 110
logical aspects of 26
misrepresented when regarded as 'theoretical' in a narrow sense 5
mode of explanation 25
must be understood by interpretationalist lights 163
naturalistic defenders of 42
naturalistic vindication of 37
no safe haven for 34
non-reducibility of 153
non-reductive 180
potential for defending 25
rationalistic 39
realistic 39
requires that contents be logically distinct 38
scientifically respectable version of 41
should we eliminate 164
threatened by the developments in today's cognitive science 5
to vindicate 33
truth of some form of 168
understood by normative interpretative lights 139
will not develop into a nomological science 145
will not form part of a more scientific psychology 168
will not integrate with natural science 167
propositional attitudes 116, 136, 154, 158, 159
ascription of 135
causal efficacy of 153, 159
causal potency and reality of 160
causalist understanding of 8
doubts concerning the very reality of 160

doubts concerning their reality 159
structural affinities with theoretical predicates 13
propositional content 5, 11, 79, 86, 91, 105, 108, 109, 112, 114, 136
able to ascribe precise 114
ascription of 112, 113, 130
distinction between intentional content and 110
emergence of 79
forego attribution of 131
having a belief with 118
holistic nature of 133
is required for an proper account of full-blown propositional attitude psychology 7
non-verbals don't have 108
only humans really have beliefs with 79
tendency to ascribe possible 108
vehicles of 113
propositional contents 90, 115, 135
ascriptions of 109, 115
inter-connections between 114
refining pre-existent, but rough-grained, 115
propositional substitutions 108 (see also formal substitutions, subsitutions)
propositional thought 117
internal structure of 135
structure, not the content, of 129
vehicles of 126
propositions 90
are structured 126
believer of 119, 131
proto-beliefs 79
proto-concepts 86
prototype generalization 97
prototype theories 89, 90, 92
prototype theory 90
prototype trajectory families 176
prototypically similar 98
proximal causal descriptions 77
proximal content 77
proximal stimuli 78, 116
proximate causes 104
psychological attitudes
can be directed at the same propositional content 38
examples of 11
functionally defined 38
psychological behaviourism 2
psychological laws 145
psychological similarity 40
psychological theorising 2
psychology
generalisations of 33
psychophysical 148, 150
psychophysical interaction 152
psychophysical law 45
psychophysical laws 145
psychophysical relation 150
psychophysical relationship 150
psychotechtonic 104
psychotechtonics 27
Ptolemaic cosmology 18–19
Ptolemy 19
public code 122
public language 122
purposive behaviour 62
understood in dispositional terms 62
Putnam 35, 37, 49, 181
Pylyshyn 95, 96
quantify
over singular terms 147
quantum mechanics 168
Quine 15, 16, 147
radical interpretation 127–128, 135
(see also interpretation)
early stages of 135
first stages of 133
project of 133
radical interpreter 127, 130, 133, 136
basic assumptions of 129
basic tool of 127
radical interpreters 128, 129, 135, 136
sum total of the evidence available to 135
radically interpret 126
Ramsey 104
random mutant 107
rational
people as 136
rational agent 130
rational decision making 25
rational thought processes
require the re-occurrence of logical structures 26
rationalistic explanation 26

rationalistic psychology
would be impossible 38
rationality 132
re-identification 97
conceptual 120
nonconceptual 120
re-identified individuals 97
re-identify 97
real
only science can tell us what is 167
realism
about beliefs 159
realist
about intentional content 39
realists 104
reality
conception of 179
realm of embodiment 102, 178
realm of reference 102
reason
citing one or another as the cause 141
for acting 142, 145
is a rational cause 142
need to connect the 142
reason for acting
how we actually designate 143
reason-explanations 56, 142, 143, 156
are a species of causal explanation 140, 142
are irreducible to physical explanations 136
as a kind of causal explanation 139
as a species of causal explanation 146
autonomy of 146
causal character of 143, 146
eschew standard accounts of causalism about 140
explain less than the best explanations in the hard sciences 159
in some sense low-grade 159
must be understood as a species of causal explanation 141
never be reduced to pure causal explanations 145
normative character of 140
open to interpretative reappraisal 137
same as other causal explanations 146
suggestive of counterfactual patterns of behaviour 163
study of 139
treated as straightforward causal explanations 144
trick to understanding the nature of 163
truly important in the study of 146
reason-giving explanations 9
reason-giving psychology 11
reasoning processes 24
reasons 163
actions performed for 105
and their mode of explanation 141
are causes 139
are not always obvious or transparent 143
causal role of 106
causalist view of 139
could be causes 141
for our actions 144
in ordinary language we commonly speak of 140
mode of explanation of 139
our practice of citing 145
play an explanatory role in behaviour 56
warrant causal forms of explanation 141
recent history
in evolutionary terms 65
recent history thesis 65
distinguishing between a mere beneficial effect and proper function 65
reduction
of theories 3
reductionism 181
goal of 3
reductionist accounts 144
reductionist metaphysics 140
reductionists 139
reductive causalists 141
reference 45, 102, 135, 178
our own standard scheme of 80

our usual scheme of 96
referential
veridically 103
referential relations 37
referential triangle 117
referentially opaque 25
regularity
connection of causality with 157
regularity criterion 145
reified content 37
represent
capacity to 72
representation 78, 102, 158, 177
as a sentence in a natural language 23
as a symbol of an internal code 23
as an abstract object 23
can be naturalistically explained 43
classical cognitivist conception of 177
context-sensitive 20
distinct types and levels of 79
established convention of 54
explained by an unmysterious naturalistic relation 42
formal feature of any account of 158
inner 84
naturalistic account of 42, 49
primitive kinds of 43
structured 99
systems of 79
traditional vision of symbolic 177
truth-conditional account of 102
two sorts of naturalistic theories of 42
veridical 84
representational capacities 72
representational capacity 77
representational content 23, 41, 45, 91, 101, 176–177
alternative theory of 177
condition on the very possibility of 54
divine the nature of 55
most fundamental type of 93
naturalistic theories of 177
of icons 72
problem of naturalising 123, 177
representational contents 22
odd 85
representational medium 21
representational relation
doubled-edged nature of 44
representational systems
three main types of 54
type I 54
type II 54
type III 55
representational theory 21
representational relation 70
representationalism
is our best theory of mind 22
representations 79, 85
a tendency to reify 84
can still be true or false 79
differences between human and animal 79
gain their content 49
less sophisticated 79
linguaform 6, 26
must indicate just one thing 57
natural selection of 55
natural systems of 55
naturalistic theory of 6
need to manipulate discrete 99
no structured 100
pre-linguistic 42
provided by our best theories 167
that 'go proxy' for 126
we misleadingly reify 85
which have no intrinsic powers of representation 54
reproduction 65
operates at the genetic level 66
reproductive success
future chances for 64
reproductively established families
biologically-based 66
higher-order 66
reproductively established family 65, 66
resemblance theories 42
retinal images 116
retinal pattern 77
retinal stimuli 43
revolutionary conceptual shift 18
Rey 46
Rorty 168
Rosenberg 75
Rowlands 61, 73–74, 78, 84

Ryle 124, 141, 150
satisfaction conditions 128
scenario content 93
science
is first learned 174
is not a definitionally watertight term 168
is not a natural kind term 168
scientific credibility 2
scientific explanation
must be of a causal character 34
scientific generalisations 13
scientific metaphors 27
scientific naturalism
exclusive 172
scientific naturalist
exclusive 171, 180
scientific naturalists 15, 168
conservative 178
scientific psychology
content-ascribing 41
could be construed narrowly 39
requires narrow contents 41
scientific realism
exclusive 8, 168
Searle 21
selectionist explanation 75
self-deceived 144
self-interpreters 143
semantic content 126
semantic predicates 150
semantic relations 48
semantic substitutions 24
semantic value 174
senile dementia 130
sense 89
determines reference 80
sense-data 3
senses
traditional assumption about the function of 84
sensory mechanisms 76
sensory perception 84
sensory systems
animal 55
less than perfect 75
sentence-use
patterns of 127
sentences
are external props 126
unlimited number of 126
serious scientific psychology 34, 110
shared attention 120, 134
sign use 120–121
similarity patterns 116–117
similarly directed 116
Simon 20
simulation theorists 109, 162
simulation theory
basic assumptions of 133
fundamental maxim of 109
singular causal claim 157
singular causal relations 145, 156
singular causal statement 145, 148
situated robotics 111
Skinner 98, 103
Skinnerian creatures 98
Smith 152
Smolensky 95, 177
social 116
social context 122
social interaction 120
social setting 122
social standard 119
socio-cultural factors 172, 180
socio-cultural features 180
space of reasons 178
spandrels 65
special science 21, 34
special sciences 168
speech acts 113, 131, 134
lack an internal coherence 130
meaningful pattern in 122
spontaneity 118
state of affairs 57
can unequivocally specify one 53
state-space 176
states of affairs 90
statistical causal account 45
does not solve the disjunction problem 45
statistical norms 45
statistically frequent 90
Sterelny 175
Stich 22, 32–37, 84, 104, 108, 110–111
stimulus discrimination 92
stimulus-driven responses 56
stimulus-response behaviour 2
stimulus-response devices 106
Strawson 94, 156, 161–163
strict laws 159 (see also laws)

structured pattern 128
structured symbols 20
structuring cause 106
structuring causes 106, 107
sub-doxastic states 79, 86
sub-symbolic 177
subject 135
subject-object split 116
subject-predicate dichotomy 96
subjective-objective contrast 118
subjective-relative properties 85
substances 97
substitutions (see also formal substitutions, intensionality test)
  Fregean restraint on 109
  in belief contexts 109
supervenience 34, 39
  admits of several interpretations 149
  context of psychophysical 150
  non-reductive case of 150
  notion of 149, 151
  of content upon putative internal events 39
  of semantic content 36
  standard accounts of 151
  strong 150
  strong psychophysical 149
  unrelativised version of 149
  weak 150
supervenience relation 150
supervenient properties 149
supervening 35
symbol
  notion of 177
  tokening of 45
  uniquely shaped 24
symbol-manipulation
  is the essence of intelligence 20
  necessary and sufficient means for general intelligent action 20
symbol-strings
  form whole propositions 24
  form whole representations 25
symbols
  are physical tokens of an internal code 23
  attaching to reality 49
  basic entities of the science of the mind 21
  can be given alternative semantic interpretations 21
  context-invariant 95, 100, 124
  do not intrinsically represent anything 21
  external 125
  have both representational capacities and formal structures 25
  logical structured 113
  logico-syntactic properties of 23
  manipulated in cognition 21
  mechanically realisable 23
  must have unique shapes 25
  ontological status of 21
  publicly recognisable 126
  represent worldly features and objects 21
  semantically evaluable 23
  understood after the fashion of other theoretical entities 20
syntactic structures
  have different inferential connections 36
syntactic theory of mind 37
syntactic tokens 32
syntactically structured 32
syntax
  as a kind of 'shape' 23
  deep 126
  does not reduce to physical shape 23
  inner 125, 126
  private 126
  unorthodox account of 176
system
  overall response of 67, 75
systematicity 95, 100
T-sentences 127, 128, 131
  are truths of logic 129
T-theories 128
T-theory 128, 131, 132, 133
  axioms of 128, 131
Tanney 146, 155
Tarski 128, 173
Taylor 179
teleofunction 107 (see also proper function)
  determines representational content 59
teleofunctional 56

levels are complementary 75
notion of an indicator 59
teleofunctional level
lowest possible 73
teleofunctional theory 6
teleofunctionalism 55, 105
allows for misrepresentation 59
need-based 59
structural instability that lies at the base of 58
teleofunctionalist account 54, 58
teleofunctions 69 (see also proper functions)
low-level 81
teleological account
without a top-down 71
teleological analysis 141
teleological approach 141
teleological functions 62
silent about the origins of 62
teleology
outright rejection of 72
The Philosopher 161
theoretic shifts 18
theoretical revolution
of great magnitude 17
theoretical statements
testing ground for the truth of 15
theoretical terms 2
a sub-set of conceptual terms 14
are meaning-holistic 14
gain their meaning by 2
theoretically embedded vocabularies 13
theories
accepted scientific 168
carry ontological implications 3
class of scientific 168
essential properties of 14
have content to compare 169
mistaken 167
scientific 168
to compare and contrast 169
virtues and vices of 169
within the bounds of natural science 167
theorising 178
heterodox account of 176
kind of content required for 172
necessary condition for 170
traditional view of 171
unorthodox account of 171
theory
can affect our observations 16
global excellence of 17, 167
theory change
ontologically radical 19
theory choice 17
theory of meaning 128, 129
necessary conditions for 128
theory of truth 129
extensional 128
for formalised languages 128
is not a theory of meaning 129, 150
logical and holistic constraints imposed by 129
recursive 127
theory-theorists
in contrast to 137
theory-theory 1, 3, 11
a modest form of 14
concession to supporters of 137
core 12
core principles 12
defines mental states 22
general background assumptions 12
knowledge of anthropocentric facts 12
narrow 12
thermoreception 83, 84
thermoreceptor 83
thermoreceptors 83
thinking
general forms of structural 100
thought
a capacity for non-objective 95
fine-grained propositional 115
logical properties of 124
more general modes of 95
non-objective modes of 95
thoughts
conceptually-based 122
having contentful 85
token identity solution 154
token identity theory 147, 148, 149, 151
cast doubt on the virtues of adopting the 153
token identity thesis 149
token-epiphenomenalism 154

Tomasello 121
total projection 162
transitivity 59
transitivity problem 58
arises at the teleofunctional level 59
only applies to cases of analytically or nomically nested information 58
triangulate
asocial creature could be in a position to 122
triangulating
with itself diachronically 119
triangulation 117, 124
condition on the possibility of 123
if it is to be a real possibility 116
necessary condition for 120
objectivity requires a kind of 116
requires more than just a capacity for re-identifying objects 120
smacks of magic 122
social 119
solitary 119, 120
without the capacity for 118
triggering cause 106
triggering causes 106, 107
Trout 178
truth 102, 119, 173, 178
as a basic or primitive concept 128
assessment of 80
generated only when the appropriate link is established 128
understand meaning in terms of 128
truth condition 79
truth conditions 40, 79, 102, 119
must play a quasi-independent role in the determination of meaning 38
truth predicate 150
truth relation 80
truth value 178
truth values 40
truth-conditional
mistake to think of such content as 80
truth-conditional content 79 (see also propositional content)
truth-conditions 45
truth-evaluable perspective 110
Twin Earth 35, 105
two-factor theories 37
must be rejected 39
result in each psychological state having its own, unique conditions of individuation 38
threaten autonomy with respect to content 38
unable to support the required predictive generalisations 38
two-factor theorists 41
two-factor theory 39, 176
type II representations
are only partially naturalistic 55
type III representations
of immediate interest to the naturalist 55
type-epiphenomenalism 154
typical causes and effects 2
Umwelt 111
unconscious 20
universal grammar 125
universals 42
universe
ideally law-like in character 160
is exclusively mechanistic 158
wholly deterministic 158
utterances 113–115, 127–128, 131, 134–135
pattern of 132
vehicles of content 96, 176
a connectionist understanding of 110
context-sensitive 125
not like those that the tradition has postulated 96
veridicality 100
vertical axis problem 44
vervet monkeys 126
visual illusions 89
visual stimulation 43
vital spirits 170
voluntary behaviour 56
is essentially learned 56
Von Eckardt 27
Von Uexkull 111
wide content 40
Williams 91
Meredith 117

Wittgenstein 117, 124, 144
Wittgensteinian tradition 142
Wright 62
Wright's
    definition of functions 62

In the series ADVANCES IN CONSCIOUSNESS RESEARCH (AiCR) the following titles have been published thus far or are scheduled for publication:

1. GLOBUS, Gordon G.: *The Postmodern Brain.* 1995.
2. ELLIS, Ralph D.: *Questioning Consciousness. The interplay of imagery, cognition, and emotion in the human brain.* 1995.
3. JIBU, Mari and Kunio YASUE: *Quantum Brain Dynamics and Consciousness. An introduction.* 1995.
4. HARDCASTLE, Valerie Gray: *Locating Consciousness.* 1995.
5. STUBENBERG, Leopold: *Consciousness and Qualia.* 1998.
6. GENNARO, Rocco J.: *Consciousness and Self-Consciousness. A defense of the higher-order thought theory of consciousness.* 1996.
7. MAC CORMAC, Earl and Maxim I. STAMENOV (eds): *Fractals of Brain, Fractals of Mind. In search of a symmetry bond.* 1996.
8. GROSSENBACHER, Peter G. (ed.): *Finding Consciousness in the Brain. A neurocognitive approach.* n.y.p.
9. Ó NUALLÁIN, Seán, Paul MC KEVITT and Eoghan MAC AOGÁIN (eds): *Two Sciences of Mind. Readings in cognitive science and consciousness.* 1997.
10. NEWTON, Natika: *Foundations of Understanding.* 1996.
11. PYLKKÖ, Pauli: *The Aconceptual Mind. Heideggerian themes in holistic naturalism.* 1998.
12. STAMENOV, Maxim I. (ed.): *Language Structure, Discourse and the Access to Consciousness.* 1997.
13. VELMANS, Max (ed.): *Investigating Phenominal Consciousness. Methodologies and Maps.* n.y.p.
14. SHEETS-JOHNSTONE, Maxine: *The Primacy of Movement.* 1999.
15. CHALLIS, Bradford H. and Boris M. VELICHKOVSKY (eds.): *Stratification in Cognition and Consciousness.* n.y.p.
16. ELLIS, Ralph D. and Natika NEWTON (eds.): *The Caldron of Consciousness. Motivation, affect and self-organization - An anthology.* n.y.p.
17. HUTTO, Daniel D.: *The Presence of Mind.* n.y.p.
18. PALMER, Gary B. and Debra J. OCCHI (eds.): *Languages of Sentiment. Cultural constructions of emotional substrates.* n.y.p.

The Presence of Mind argues that beliefs and desires have a legitimate place in the explanation of action. Eliminativist arguments mistakenly focus on the vehicles of content not content itself. This book asks whether a naturalistic theory of content is possible. It is argued that a modest biosemantic theory of intentional, but non-conceptual, content is the naturalist's best bet. A theory of this kind complements connectionism and recent work on embodied and embedded cognition. But intentional content is not equivalent to propositional content.

In order to understand propositional content we must rely on Davidsonian radical interpretation. However, radical interpretation is shown to be at odds with physicalism. Consequently, if the best naturalised theory of content we are likely to get from cognitive science is only a theory of intentional content, then a naturalistic explanation of scientific theorising is not possible. It is concluded that eliminativism is intellectually incoherent.

John Benjamins Publishing Company