

1-1-1984

# A new perspective on the mind-body problem.

Jesse L. Yoder

*University of Massachusetts Amherst*

Follow this and additional works at: [https://scholarworks.umass.edu/dissertations\\_1](https://scholarworks.umass.edu/dissertations_1)

---

## Recommended Citation

Yoder, Jesse L., "A new perspective on the mind-body problem." (1984). *Doctoral Dissertations 1896 - February 2014*. 2055.  
[https://scholarworks.umass.edu/dissertations\\_1/2055](https://scholarworks.umass.edu/dissertations_1/2055)

This Open Access Dissertation is brought to you for free and open access by ScholarWorks@UMass Amherst. It has been accepted for inclusion in Doctoral Dissertations 1896 - February 2014 by an authorized administrator of ScholarWorks@UMass Amherst. For more information, please contact [scholarworks@library.umass.edu](mailto:scholarworks@library.umass.edu).



UMASS/AMHERST



312066013539860

A NEW PERSPECTIVE ON THE MIND-BODY PROBLEM

A Dissertation Presented

By

Jesse L. Yoder

Submitted to the Graduate School of  
University of Massachusetts in partial fulfillment  
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

May 1984

Philosophy



A NEW PERSPECTIVE ON THE MIND-BODY PROBLEM

A Dissertation Presented

By

JESSE L. YODER

Approved as to style and content by:

Gareth B. Matthews  
Gareth Matthews, Chairperson of Committee

Vere Chappell  
Vere Chappell, Member

Cynthia Freeland  
Cynthia Freeland, Member

M. Arbib  
Michael Arbib, Member

Michael Jubien  
Michael Jubien, Department Head  
Department of Philosophy



## DEDICATION

To Feroline Whitehead, Gareth Matthews, and Tech. Pubs. at Wang.

## ACKNOWLEDGEMENTS

Many people deserve mention for their help at various points in my career. I will mention them in roughly chronological order.

While I received much valuable help from many of my philosophy teachers at the University of Maryland, several deserve particular mention. Peter Goldstone was my first philosophy teacher and spent many hours explaining philosophy to me. I learned a lot about writing and criticism from Alan Pasch in his seminar on Necessity at the University of Maryland. I learned about Wittgenstein and other contemporary topics from Jack Odell. Many other members of the University of Maryland philosophy faculty and graduate students deserve credit for the many hours they spent teaching me philosophy and discussing philosophy with me during my undergraduate career. It was at Maryland during my sophomore year in college that I first became interested in the mind-body problem and decided to write my Ph.D. dissertation on this subject.

At The Rockefeller University I had the opportunity to discuss philosophy freely with many teachers and students. I am especially grateful to Joel Feinberg for his Social Philosophy and Philosophy of Law tutorials at The Rockefeller University. I believe I learned more about careful writing and criticism from him than from any other person. I would like to thank Harry Frankfurt for the many

hours we spent in tutorials on Philosophy of Mind and Descartes. I also received much insight into the mind-body problem from discussions with Donald Davidson. Norton Batkin, Alan Berger, Adina Schwartz, and Jonathan Lear as fellow students and as friends provided much encouragement, support, and helpful criticism.

I am grateful to the members of my dissertation committee with whom I also took courses at the University of Massachusetts. These members include Gareth Matthews, Vere Chappell, Cynthia Freeland, and Michael Arbib. I could not have completed my dissertation in so timely a fashion without Gary Matthews' quick and helpful responses to every chapter I submitted. I have made every effort to incorporate comments made by the members of my dissertation committee in the final version.

I would also like to thank my friends in the Technical Publications department at Wang Laboratories for their encouragement and support during the last year while I was writing my dissertation. Jack Cahalan provided sage philosophical comments. I would particularly like to mention Joe Quigley and Mike Albert, my racquetball partners, and Nancy Kraus. I believe that my experience as a technical writer at Wang Laboratories has significantly improved my philosophical writing.

I am grateful to Commercial Union Insurance Companies and to Wang Laboratories for providing me with full-time employment as a technical writer during the three years in which I wrote my



dissertation. My work with computers has given me a better understanding of philosophical arguments relating to artificial intelligence and computer models of the mind.

Other friends I would like to mention are Steve Shannon, Barbara Winterson, Carl Sosebee, David Morrissey, David Hersh, Tim Bunnell, Larry Hohm, and Alan McMichael. I would especially like to thank Feroline Whitehead for her support, encouragement, and assistance during the entire time I wrote my dissertation.

Finally, I would like to mention my cat Pascal, who always remained a friend and has served as an example of someone with a radically different point of view.

## PREFACE

I have reproduced all emphasis in the form of italics or underlining in quotations in the form of underlining. Unless otherwise indicated, all such underlining throughout represents the emphasis of the original author.

ABSTRACT

A NEW PERSPECTIVE ON THE MIND-BODY PROBLEM

May 27, 1984

Jesse L. Yoder, B.A., University of Maryland

M.A., University of Massachusetts, Ph.D., University of Massachusetts

Directed by: Professor Gareth Matthews

The principal critical objective of this dissertation is to examine three contemporary theories about the mind-body problem: dualism, anomalous monism, and functionalism. The dualism examined is closely linked to Cartesian dualism, while functionalism is a form of materialism. Anomalous monism is a kind of dual aspect view. All these theories have a long tradition with different formulations and exponents. I examine three contemporary exponents of these views: Saul Kripke, a dualist, Donald Davidson, an anomalous monist, and Daniel Dennett, a functionalist.

I argue that each of these theories is incorrect or has serious difficulties. The chief difficulty for dualism is in giving an account of mind-body interaction. The chief difficulty for anomalous monism is that internal tensions within the theory relating to Davidson's conception of physical events and scientific laws have the consequence that anomalous monism applies to the



entities of physics and chemistry but not to events in the brain. The most serious difficulty for functionalism is in accounting for the qualities of subjective experience.

I then discuss two criteria for identifying mental phenomena: incorrigibility and intentionality. I consider the problem of mind-body interaction as a problem for dualism. I argue that the mind-body problem is misconceived in that it is supposed that a univocal definition of 'physical' can be given. I state three different definitions of this term: one containing a scientific conception, one an objective conception, and one a personal conception.

In my final chapter, I formulate a plural aspect view that accounts for mind-body interaction. On this view, mind is an aspect of body and body is an aspect of mind. This plural aspect view can be derived from a more general theory I call Viewpoint Pluralism.

## TABLE OF CONTENTS

ACKNOWLEDGEMENTS.....	iv
PREFACE.....	vii
ABSTRACT.....	viii
INTRODUCTION.....	1
Chapter	
I. THE MIND-BODY PROBLEM: A HISTORICAL PERSPECTIVE.....	3
1.1 Introduction.....	3
1.2 The Mind-Body Problem as a Philosophical Problem.....	4
1.3 Dualism.....	8
1.3.1 Types of dualism.....	9
1.3.2 Early Greek conceptions.....	12
1.3.3 Plato.....	16
1.3.4 Descartes.....	23
1.3.5 Summary of dualist positions.....	27
1.3.6 Conclusion.....	29
1.4 Materialism.....	29
1.4.1 Aristotle.....	30
1.4.2 Hobbes.....	34
1.4.3 La Mettrie.....	37
1.4.4 Five types of materialism.....	41
1.4.5 Summary.....	45
1.5 Plural Attribute Monism.....	45
1.5.1 Spinoza.....	47
1.6 Idealism.....	50
1.7 Pluralism.....	51
1.7.1 Leibniz.....	53
1.8 Conclusion.....	54
II. SAUL KRIPKE'S DEFENSE OF DUALISM.....	61
2.1 Introduction.....	61
2.2 Kripke's Theory of Proper Names.....	62
2.2.1 Names and some general terms are rigid designators.....	63
2.2.2 Identity statements between rigid designators are necessary.....	65

2.2.3	Origin is an essential property.....	66
2.2.4	Summary of Kripke's metaphysical views.....	67
2.3	Kripke's Arguments Against the Identity Thesis.....	68
2.3.1	Kripke's argument that a person is distinct from his body.....	68
2.3.2	Event identity.....	70
2.4	Feldman's Criticisms.....	71
2.4.1	Feldman's criticism of Kripke's argument that Descartes is not identical to his body.....	72
2.4.2	Feldman's criticism of Kripke on event identity..	79
2.5	Feldman's Three Conceptions of Events.....	83
2.6	Other Issues Relating to Dualism.....	87
2.6.1	The problem of interaction.....	88
2.7	One Consequence of Kripke's View on Origin.....	91
III.	DONALD DAVIDSON'S ANOMALOUS MONISM.....	98
3.1	Introduction.....	98
3.2	Davidson's Views in "Mental Events".....	101
3.3	Davidson's Argument for Anomalous Monism.....	103
3.4	Davidson's View of Mental and Physical Events.....	107
3.5	Davidson's Mental and Physical Domains.....	115
3.5.1	Davidson's claim that every scientific law must draw on a homonomic domain.....	126
3.6	Davidson's Argument for Anomalous Monism.....	130
3.7	Davidson's Demonstration of Identity.....	132
3.8	Conclusion.....	135
IV.	DANIEL DENNETT'S DEFENSE OF FUNCTIONALISM	
4.1	Introduction.....	138
4.2	Dennett's Flowchart of Pain.....	139
4.2.1	The personal/sub-personal distinction.....	145
4.2.2	Dennett's account of pain.....	149
4.2.3	Perceptual analysis.....	154
4.3	Are Our Intuitions About Pain Inconsistent?.....	159
4.4	Robot Emotions.....	164
4.5	Summary.....	166
V.	INCORRIGIBILITY AND INTENTIONALITY: TWO CRITERIA FOR MENTAL PHENOMENA.....	169
5.1	Introduction.....	169
5.2	Incorrigibility.....	170
5.2.1	Mental events and mental features.....	171
5.2.2	Rorty's definition of incorrigibility.....	175



5.2.3	Mental features.....	185
5.2.4	Conclusion about Rorty's defense of incorrigibility.....	188
5.2.5	Rorty's argument for materialism.....	189
5.3	Intentionality.....	194
5.3.1	Chisholm's criteria for intentional sentences...	195
5.3.2	Criticisms of Chisholm's criteria.....	199
5.3.3	Chisholm's reformulated criterion.....	201
5.3.4	Marras' revisions of Chisholm's criteria.....	203
5.3.5	Conclusion about intentionality.....	211
5.4	Conclusion.....	213
VI.	MIND-BODY INTERACTION AND THREE CONCEPTIONS OF THE PHYSICAL.....	220
6.1	Introduction.....	220
6.2	Dualism and Mind-Body Interaction.....	222
6.3	Three Conceptions of the Physical.....	237
6.3.1	The scientific conception.....	237
6.3.2	The personal conception.....	243
6.3.3	The objective conception.....	248
6.4	Conclusion.....	254
VII.	AN ACCOUNT OF MIND-BODY INTERACTION.....	258
7.1	Introduction.....	258
7.2	Levin's Argument Against Mind-Body Interaction.....	260
7.2.1	Criticisms of Levin's argument.....	261
7.2.2	Levin's argument applied to (s)physical, (o)physical, and (p)physical entities.....	263
7.3	An Account of Mind-Body Interaction.....	277
7.4	Similar Accounts.....	286
7.5	Viewpoint Pluralism.....	294
7.6	Summary.....	298
7.7	Conclusion.....	299
.....		
	SELECTED BIBLIOGRAPHY.....	303

## FIGURES

### Figure

4-1.	Dennett's Flowchart of Pain.....	141
6-1.	An Illustration of a Neuron.....	230
7-1.	A Schematic Diagram of the Visual System.....	267
7-2.	A Diagram of Mental-Physical Interaction.....	278
7-3.	Potentials Recorded Before a Voluntary Finger Movement....	281
7-4.	Cortical Activity Prior to Voluntary Movement.....	282



Jesse L. Yoder  
All Rights Reserved



## INTRODUCTION

What is the relation between mind and body? This is one formulation of a problem which has perplexed both philosophers and non-philosophers for many centuries. This problem has come to be known as "the mind-body problem", and is not lacking in attempted solutions. Fortunately, each solution is not completely unique; different solutions fall into categories and types so that a survey of the main solutions is not an impossible task. While such a survey could not hope to examine every variation on a type of solution, many comments on a theory which offers a type of solution also apply to other theories offering a solution of the same or similar type.

The principal objective of this dissertation is to critically examine three contemporary theories about the mind-body problem: dualism, anomalous monism, and functionalism. The dualism examined is closely linked to Cartesian dualism, while functionalism is a form of materialism. Anomalous monism, on the other hand, is a kind of dual aspect view. All these theories have a long tradition with different formulations and exponents. I have chosen contemporary exponents of these views to focus on since my purpose is primarily critical and not primarily historical. However, in this first chapter I give a historical overview of the mind-body problem. I will now outline the chapters in the order in which they occur.

The critical examination of three contemporary theories about the mind-body problem takes place in Chapters II, III, and IV. In Chapter II, I consider dualism as represented in the writings of Saul Kripke. In Chapter III, I discuss the anomalous monism of Donald Davidson. In Chapter IV, I consider the functionalism of Daniel Dennett. These three theories are contemporary formulations of three traditional solutions to the mind-body problem: dualism, dual aspect views, and materialism. I argue that each of these three theories is incorrect or has serious difficulties. In Chapter V, I discuss two criteria for identifying mental phenomena: incorrigibility and intentionality. In Chapter VI, I present the problem of mind-body interaction as a problem for dualism. I also distinguish three conceptions of what is physical. In Chapter VII, I present an account of mind-body interaction, utilizing the distinctions developed in Chapter VI.

## C H A P T E R I

### THE MIND-BODY PROBLEM: A HISTORICAL PERSPECTIVE

#### 1.1 Introduction

The main purpose of this chapter is to introduce the mind-body problem as a philosophical problem and then to provide a historical overview of the problem. I introduce the mind-body problem as a philosophical problem in the next section. The historical overview follows, and is divided into five main sections: Dualism, Materialism, Plural Attribute Monism, Idealism, and Pluralism. The purpose of placing the "Dualism" section first is to allow me to begin with a consideration of early Greek conceptions of mind and body, which are dualistic, and also to allow me to discuss Plato before Aristotle. I begin the "Dualism" section with a consideration of early Greek conceptions of soul, mind, and body, particularly as they occur in the works of Homer. I then consider the views of Plato and Descartes. In the "Materialism" section I mainly discuss materialism as represented by Aristotle and Hobbes. In the "Plural Aspect Monism" section I discuss Spinoza's dual aspect view as an ancestor to anomalous monism. Because I do not consider idealism in later chapters, it receives only brief attention here. In the "Pluralism" section I describe the views of

Leibniz, mainly in relation to his notion of a point of view. I conclude with a brief summary of the chapter.

## 1.2 The Mind-Body Problem as a Philosophical Problem

Sometimes when I tell my non-philosophical friends that I am writing about the mind-body problem, they say "Oh, I didn't know there was a problem!". What, then, is the problem about the relation between mind and body? The fact that many people who have not reflected on this issue think it odd that someone should think there is a problem about the mind-body problem reflects how unproblematical this relation is in daily life. We have no problem in acting on our decisions, in intentionally raising our arms, in feeling pain due to bodily injury, and in seeing an oncoming automobile. These examples of mind-body interaction are so natural that it seems pointless from a practical viewpoint to raise a problem about them. These actions would be problematical only to someone with paralysis or a mental disorder. If there is a mind-body problem, then, it is a theoretical or philosophical one, not a practical one.

The philosophical problem here is not the practical question "How can I raise my arm?" but the philosophical question "What coherent account or explanation can be given of mind, of body, and of their relation?" Some solutions to this problem have held that mind and body are of the same type. These solutions are monistic,



and both materialism and idealism are forms of monism: the view that there is only one type of entity or phenomenon in the universe. Other solutions are dualistic: there are two types of entity or phenomenon in the universe and these two types are mind and body or mental and physical phenomena. There are also pluralistic solutions: minds and bodies are two of a wide variety of types of entities and phenomena in the universe.

It should be apparent that the problem is already becoming more complex than it appeared at first. To answer the philosophical question "What coherent account or explanation can be given of mind, of body, and of their relation?", at least the following questions need to be answered:

1. What is mind?
2. What is body?
3. What is the relation between mind and body?

Even questions 1 and 2 can be interpreted conceptually or ontologically, as can question 3:

- 1(c). What does the term 'mind' mean, from a conceptual viewpoint?
- 1(o). What type of entity is mind, from an ontological viewpoint?
- 2(c). What does the term 'body' mean or refer to, from a conceptual viewpoint?
- 2(o). What type of entity is body, from an ontological viewpoint?

3(c). What conceptual relations obtain between mind and body?

3(o). What ontological relations obtain between mind and body?

Question 3 could also be taken to mean either of these two questions:

3(a). Are mind and body related by identity or distinctness?

3(b). Are mind and body related by interaction, and if so how?

This list of questions shows that the mind-body problem is not just one problem but rather a set of related problems. Probably the way the mind-body problem is most commonly formulated as a philosophical problem is:

I. What is the relation between mind and body?

Answers to this question do implicitly or explicitly contain a theory about mind and body. For example, materialism asserts that mind is identical to, can be reduced to, or is a function of body, while dualism asserts that they are of different (usually dichotomous) types and they are distinct. In this and later chapters I shall take question I as being the formulation of the mind-body problem with which I shall be concerned; if there are other ways of formulating this problem, then what I say may or may not apply to these other formulations, depending on what they are.

The only ways to show that the mind-body problem is a philosophical problem other than what I have just said is to argue that it is traditionally discussed by philosophers in philosophical contexts<sup>1</sup> or to appeal to a general conception of what a

"philosophical problem" is. The rest of this chapter shows that this problem is traditionally discussed by philosophers. I have already done the second implicitly by saying that the mind-body problem should be interpreted as a question about what coherent account or explanation can be given of mind and body. Rather than defend the conception of "philosophical problem" which is implicit in these remarks, I will simply say that it is a traditional conception of a "philosophical problem" that it involves giving a coherent account or explanation of some type of entity or phenomenon, or a coherent account or explanation of the terms referring to some type of entity or phenomenon. One of the challenges in solving the mind-body problem is to construct a single coherent account or explanation that will account for both minds and bodies, or both mental and physical phenomena, as well as for their relations. Thus, materialism is not usually faulted for its account of body (although I do so in later chapters) but for its account of mind. Another way of putting this is to say that a solution will involve constructing a single coherent viewpoint in terms of which mind, body, and their relations can be accounted for.

I now begin the historical survey of views on the mind-body problem. This survey is divided into five sections: Dualism, Materialism, Plural Aspect Monism, Idealism, and Pluralism. While such a survey cannot be exhaustive, its main point is to show that there is a link between those contemporary theories discussed in later chapters and traditional mind-body theories. This link is

that those contemporary theories discussed in later chapters are contemporary formulations of traditional dualism, materialism, dual aspect views, or pluralism. Hence, many criticisms of and arguments for these contemporary formulations also apply to traditional formulations of these theories. Therefore the critical discussion of contemporary theories in Chapters II, III, and IV should be conceived as a critical discussion of the traditional theories which these contemporary theories formulate and not merely as a discussion of single isolated theories.

### 1.3 Dualism

Dualism is the view that there are two irreducibly distinct types of entities or phenomena in the universe. Dualism can be formulated as an ontological claim or as a conceptual claim. Ontologically, dualism is a claim which can apply at least to entities or substances, properties, events, states, conditions, objects, and subjects; and this list is not complete. Sometimes a distinction is made between substantial dualism, or dualism of substance, and dualism of properties. On the former view, there are two irreducibly distinct kinds or types of substance in the world<sup>2</sup> or universe: minds and bodies. On the latter view, there are two irreducibly distinct kinds of properties in the the world or universe: mental properties and physical properties.

### 1.3.1 Types of Dualism

Mind-body discussion also often involves discussion of what types of events or states there are: mental and/or physical. An event or state is not a property, but neither events nor states seem to fit the criteria for being a substance either. Substances such as chairs and persons have continuity over time while most events occur at a particular time and then cease to exist. Even the term 'entity', which might seem applicable to almost any thing, does not seem clearly applicable to an event such as the blinking of an eye or to a state such as the state of confusion in which someone might find the kitchen after a Chinese meal has been cooked. Besides dualism of substance and dualism of properties, then, there is also dualism of events and dualism of states. Similar reasoning could also show the existence of dualism of conditions, dualism of objects, and dualism of subjects.

What distinctions are significant here? Certainly one of the most important distinctions is between the term 'mind' used as a noun and the term 'mental' used as an adjective. Used as a noun, 'mind' contrasts with 'body' and refers to a type of substance or entity. Used as an adjective, 'mental' can refer to a type of property, event, state, condition, subject, or object, to name some but not all of the types of phenomena which can be either mental or physical. I might say "My mind is playing tricks on me" or "A feeling is a mental event". Rather than trying to specify all the

different types of phenomena which can be either mental or physical each time this issue arises, I use the term 'phenomena' as a general term to cover at least properties, events, states, conditions, abilities, capacities, and phenomena other than entities that can be either mental or physical. This is not because I think there is no important difference between, for example, dualism of properties and dualism of events. Furthermore, there are times such as in the discussion of Davidson in Chapter III when the discussion will center on one particular type of phenomenon; in Davidson's case it centers on events. Only when the differences in types of phenomena are unimportant in a particular context or I wish to make a general statement do I use the term 'phenomenon'.

One distinction I maintain throughout future discussion is the distinction between questions about the relation between mind and body and questions about the relation between mental phenomena and physical phenomena. Much contemporary discussion of the mind-body problem centers on the relation between mental and physical events or states rather than the relation between mind and body. The implicit (and sometimes explicit) assumption of those conducting this discussion seems to be that if one could state the relation between mental and physical events, states, or other phenomena, then it would be a simple matter to state the mind-body distinction. A person's mind could be defined as a collection of mental events or states, or some other such definition could be given. While this may be true, it is not obviously true, and it is a mistake to make



such an assumption without examining its merits. Furthermore, the notion of a person is central to the mind-body problem and it is more important to say what a person's body and mind are and what their relation is to the person than to say what the relation is between a person and his brain states and mental states. And even if it is not more important, the issues are distinct and should be distinguished. Hence I maintain the distinction between mind-body questions as opposed to mental-physical phenomena questions throughout future discussion.

In general, I use the term 'entity' rather than the terms 'subject', 'object', or 'substance' to denote an individual thing. A subject is an individual thing which is the subject of perception or of properties, while an object is an individual thing that exists independent of the mind or that is the object of perception. Something can be both a subject and an object: a subject of perception can also be an object of perception, while a physical object can also be described as a subject of properties. The term 'substance' has philosophical difficulties of its own. The term 'entity', on the other hand, simply refers to an individual thing and does not refer to a particular type of individual thing. In general, then, I use the term 'entity' to refer to persons, human beings, animals, minds, bodies, physical objects such as rocks and tables, and also mental objects such as beliefs and feelings. The mind-body issue, then, is one of the relation between minds and bodies, and of the relation between mental and physical phenomena.

Let us return to the distinction between ontological and conceptual dualism. Ontologically, dualism can be stated as follows:

- D(o). There are two irreducibly distinct types of entities in the universe: minds and bodies.
- D(o)'. There are two irreducibly distinct types of phenomena in the universe: mental and physical phenomena.

Conceptually, dualism can be stated as follows:

- D(c). The terms 'mind' and 'body' represent irreducibly distinct categories.
- D(c)'. The terms 'mental' and 'physical' as applied to phenomena represent irreducibly distinct categories.

These distinctions become especially important later when I discuss philosophers such as Spinoza who combine conceptual dualism with ontological monism.

### 1.3.2 Early Greek Conceptions

The history of the concepts of mind, body, and soul which we have today goes back to the earliest recorded part of Western civilization. Homeric man had a very different view of himself and his neighbor than we do today. According to Bruno Snell,<sup>3</sup> "the early Greeks did not, either in their language or in the visual arts, grasp the body as a unit". The Homeric man knew the human body not as a unity but as a sum total of the limbs. Likewise, there are at least three words that characterize the mind or soul. For Homer psyche is the force which keeps the human being alive. Thymos is the generator of motion or agitation in Homer, while noos

is the cause of ideas and images. Thymos, noos, and psyche are conceived as separate organs, each with its own individual function. Just as the limbs of the body are not conceived as part of a whole called "body", so thymos, noos, and psyche are not conceived as elements of a whole called "soul" or "mind".

Among the three terms 'thymos', 'noos', and 'psyche', 'thymos' and 'noos' are most closely associated with our concept of "mind", while 'psyche' is most closely associated with our concept of "soul". Snell describes the thymos as the organ of (e)motion and noos as the organ of perception.<sup>4</sup> We also might describe this as the distinction between feeling and thought, certainly a fundamental distinction today within the mind. The word 'psyche' is akin to the Greek word meaning "to breathe" and denotes the breath of life which departs through the mouth. Homer says that the psyche leaves man at the moment of death and goes down to Hades; hence, this is one of the original sources for the conception of the soul as an immaterial entity that survives death.

Snell's account of how the terms 'soma' (body) and 'psyche' (soul) evolved from these very limited applications to terms more closely resembling our terms 'body' and 'soul' is very interesting:<sup>5</sup>

That Homer's conception of thymos, noos and psyche still depended to a large extent on an analogy with the physical organs becomes a matter of absolute certainty if we turn to that era of transition when his conception began to be abandoned. To be sure, the evidence for the use of the words soma and psyche during the period extending from Homer to the fifth century is not full enough to allow us to trace the origin of the new terms 'body' and 'soul' in

every detail. Apparently they were evolved as complementary terms, and more likely than not it was psyche which first started on its own course, perhaps under the influence of notions concerning the immortality of the soul. The word denoting the eschatological soul was put to a new use, to designate the soul as a whole, and the word for corpse came to be employed for the living body; the reason for this must be that the element which provided man during his living days with emotions, perceptions and thoughts was believed to survive in the psyche. Presumably people felt that animate man had within him a spiritual or intellectual portion, though they were unable to define this element by one term sufficiently accurate and inclusive. As a matter of fact, this is the state of affairs which we shall meet among the early writers of lyric poetry. And it may be inferred that, because the eschatological psyche had been correlated with the soma of the dead, the new psyche, the 'soul', demanding a body to suit it, caused the term soma to be extended so that it was ultimately used also of the living body. But whatever the details of this evolution, the distinction between body and soul represents a 'discovery' which so impressed people's minds that it was thereafter accepted as self-evident, in spite of the fact that the relation between body and soul, and the nature of the soul itself, continued to be the topic of lively speculation.

Snell credits Heraclitus with being the first writer to introduce this new concept of the soul:<sup>6</sup>

The first writer to feature the new concept of the soul is Heraclitus. He calls the soul of living man psyche; in his view man consists of body and soul, and the soul is endowed with qualities which differ radically from those of the body and the physical organs. We can safely say that these new qualities are irreconcilable with the categories of Homer's thought; he does not even dispose of the linguistic prerequisites to describe what Heraclitus predicates of the soul. The new expressions were fashioned in the period which separates Heraclitus from Homer, that is to say the era of the lyric. Heraclitus says: 'You could not find the ends of the soul though you travelled every way, so deep is its logos'. This notion of the depth or profundity of the soul is not unfamiliar to us; but it involves a dimension which is foreign to a physical organ or its function.... Homeric speech does not yet know this aspect of the word 'deep'.

Snell goes on to recount several other important ways in which the psyche of Heraclitus is different from the Homeric psyche. It has a logos which is a "common" thing, so that two beings can be of the same spirit or mind, and the psyche of Heraclitus also has a logos capable of extending and adding to itself of its own accord. For Homer, any augmentation in mental powers is effected from without, not from within. We see, then, that the development of the concepts of soul and body was like an organic process which took many centuries and did not occur by sudden stipulation.

In the course of his discussion of 'psyche', 'thymos' and 'noos', Snell offers some interesting insight into the origin of 'to know':<sup>7</sup>

Never does Homer, in his descriptions of ideas or emotions, go beyond a purely spatial or quantitative definition; never does he attempt to sound their special, non-physical nature. As far as he is concerned, ideas are conveyed through the noos, a mental organ which in turn is analogous to the eye; consequently 'to know' is εἶδέναι which is related to ἰδεῖν 'to see', and in fact originally means 'to have seen'. The eye, it appears, serves as Homer's model for the absorption of experiences. From this point of view the intensive coincides with the extensive: he who has seen much sufficiently often possesses intensive knowledge.

We often speak of "intellectual perception" and of seeing that something is true. Also the term 'view' as it occurs in "visual or intellectual point of view" is related to this analogy. If Snell is right, the idea of intellectual perception or having an intellectual view originated in a conception that there is a mental organ like the eye which is responsible for understanding and perceiving in a way similar to the way the eye is responsible for vision.

Furthermore, this "mental organ" was conceived as being quite different from any conception we might have today of such an "organ", since it was not conceived as being an element of a greater whole known as "mind". This conception of a "mental organ" or something like it is in the Aristotelean doctrine of mental faculties and in the "faculty psychology" of today. It also survives today in terms such as 'faculty', 'view', and in phrases such as "I can see what you mean".

We can see from the preceding discussion that it was not possible to even formulate dualism as a philosophical doctrine until the elements of psyche, thymos, and noos could be seen as elements of a greater whole. However, it is important to notice that even as early as Homer's time people found it useful and important to distinguish between their body and other aspects which today we call emotion, perception, and life, spirit, or soul. Materialists who wish to deny the validity of these distinctions would do well to try to explain why peoples in the earliest part of Western civilization found it useful and important to develop a language which recognizes a distinction between the body and non-physical elements.

### 1.3.3 Plato

Two main elements of Plato's thought are of interest in relation to Plato's views on the mind-body problem: his doctrine that the soul exists before the birth of the person and also



survives death and his doctrine that the soul consists of three elements. I now discuss both of these doctrines, beginning with the former.

G. M. A. Grube cites a source for Plato's conception of the soul which is different from the Homeric source we have just discussed:<sup>8</sup>

The conception of the soul as the highest part of man seems to have been imported into Greece by those mystical teachers and prophets who are usually somewhat summarily lumped together as the Orphics. Their doctrines came from the East; they seem to have taught an immortality that was no longer a pale reflection of earthly life, but a release from the body and a deliverance. The body to them was the prison or tomb of the soul--*ψυαδ κμα*, as they pithily expressed it. Man then aims at the purification of this soul, and after many incarnations rises to perfection and is absorbed, or reabsorbed, into the divine. It was probably under Orphic influence that the Pythagoreans developed their way of life as a gradual process of purification. But to them this immortal psyche was the intellectual power of man and purification lay to a large extent through a strictly scientific, which to them meant a mathematical, training; though some of them seem to have laid a great deal of emphasis on number-magic and on ritual. From them must have come the conception of the intellect as the noblest and immortal part of man, of salvation through knowledge, a conception so magnificently expressed in the Phaedo, and it remained with Plato to the end.

This account and the view that Plato's conception also originated in Homeric concepts are not inconsistent: the Homeric conception of psyche was that its survival in Hades was a pale reflection of human life. This original weak conception of psyche had by Plato's time become strengthened through the work of philosophers such as Heraclitus and Pythagoras into a conception of a single entity which survives through time and which unites different mental elements



into one. If Grube is right, this conception became identified in Plato with the rational and intellectual part of man through the influence of the Orphics. What was the soul now acquired mental properties, but still survived death.

This account should shed some light on a fundamental ambiguity in dualism which survives even today: the ambiguity between "soul" and "mind". Philosophers today prefer to discuss "the mind-body problem" and take dualism to assert the distinctness of mind and body. Talk of the soul is left to theologians and ministers, and to those who speak metaphorically. Phrases such as "a spirited display", "my spirit was set free", or "a feeling in my soul" are dismissed as metaphorical and/or anachronistic. The preceding account shows, however, that the original conception of psyche or soul was developed long before Christianity and as developed as a way of conceptualizing life itself, not in connection with a religious theory. In Plato, this conception of psyche became identified with the intellectual part of man; this intellectual part or element then was conceived as having immortality but lost some of its association with life. Today, philosophers generally focus on the intellectual mind which they presume not to be immortal and discuss life or the soul seldom or not at all.<sup>9</sup>

One principal source for Plato's views and arguments for the preexistence and postexistence of the soul is the Phaedo. The purpose of this dialogue is to show that virtue has its own reward, that the man who lives a virtuous life is rewarded after he dies.

In order to demonstrate this, Plato must show that the soul survives death. Plato has Socrates argue for the existence of the soul after death by arguing that it follows from the existence of the soul before birth:<sup>10</sup>

For if the soul exists before birth, and in coming to life and being born can be born only from death and dying, must she not after death continue to exist, since she has to be born again?

Socrates offers several distinct arguments for the claim that the soul exists before birth. One is based on the idea that things having opposites are generated out of their opposites (the opposites here are life and death). Another argument is based on the famous Platonic doctrine that learning is recollection of knowledge gained prior to our birth as human beings.

My purpose here is not to critically evaluate Plato's argument but rather to describe the type of dualism he advocated. However, there are two comments worth making in relation to his arguments. One is that it is important to distinguish questions about (mere) survival of death from questions about the immortality of the soul. If I am to live forever, I must at least survive my own death. But I might survive my own death yet not live forever. Perhaps, as Cebes proposes,<sup>11</sup> the soul enters into one body after another, but eventually perishes at the end of one of these lives.

Secondly, if Plato wishes to prove that a virtuous man is rewarded in the after-life, it is crucial that a person be identified with his soul. If my soul survives death, but I am not identified with my soul, then it is difficult to see why this is of

more interest than the claim that my liver will survive my death. This same point applies to preexistence. Suppose, as Plato says, learning is recollection of something known in an earlier life. If some person A can be truly said to be remembering that so-and-so, then it must be the case that at some earlier time A has at least believed (and perhaps known) that so-and-so is the case. This also applies to experiences and events. If I am to be truly said to remember an experience or event, then it must be an experience I had or an event I witnessed. Hence if Plato's theory that learning is recollection of knowledge gained before birth is to succeed, he must claim that the person existed before his birth in order to acquire this knowledge. Thus Plato's arguments for preexistence and postexistence both require that a person be identified with his soul.

The second main element of Plato's thought I will discuss is his tripartite division of the soul. The main sources of this doctrine are The Republic and the Phaedrus.<sup>12</sup> In the Phaedo, Plato considers the soul as a simple unity, identifying it with the intellect. The tripartite division of the soul retains the idea that the soul is a unity, but adds the idea that it is complex, not simple. Plato's argument that there are three elements of the soul is based in part on an analogy with his division of the state into three classes and also on the fact that we experience internal conflicts:<sup>13</sup>

"Are we to say, then, that some men sometimes though thirsty refuse to drink?" "We are indeed", he said, "many and often." "What then," said I, "should one affirm about them? Is it not that there is a something in the soul that bids them drink and something that forbids, a different something that masters that which bids?" "I think so." "And is it not the fact that that which inhibits such actions arises when it arises from the calculations of reason, but the impulses which draw and drag come through affections and diseases?" "Apparently." "Not unreasonably," said I, "shall we claim that they are two and different from one another, naming that in the soul whereby it reckons and reasons the rational and that with which it loves, hungers, thirsts, and feels the flutter and titillation of other desires the irrational and appetitive--companion of various reflections and pleasures." "It would not be unreasonable but quite natural," he said, "for us to think this." "These two forms, then, let us assume to have been marked off as actually existing in the soul. But now the Thumos or principle of high spirit, that with which we feel anger, is it a third, or would it be identical in nature with one of these?

Plato has Socrates, who is speaking here with Glaucon, conclude that the third element, the "spiritual" element, is a distinct element from the rational or appetitive elements. These divisions resemble the Homeric mental organs of perception, emotion, and life except that the "spirited" element has lost much of its connection to the "breath of life" and is a more purely psychological concept. Also, the Homeric divisions were not elements of a whole; they existed independently, while in Plato the three elements are elements of a unity known as the soul.

In conclusion, Plato did not seem overly concerned to prove the distinctness of soul from body; perhaps he thought this did not require proof. Instead, he was concerned to demonstrate much bolder claims: that the soul exists before birth and also after death. He

also provided a tripartite analysis of the soul which preserved the idea that the soul is a unity and yet accounts for some of the complexity of our mental and emotional experience. Although the idea that the soul exists before birth is not much discussed or widely held in Western cultures, yet the idea that the soul survives death is standard Christian doctrine. Philosophers today, however, tend to read "soul" as "mind" and concentrate on Plato's identification of the soul with the intellect or reason at the expense of those aspects of the soul connected with life or after-life.

Plato also made another important contribution to dualism; it is in Plato that the soul appears as an independent substance which can exist apart from the body. It is this conception that was later taken over by Descartes who referred to this substance sometimes as "mind" and sometimes as "soul". At the same time, Plato made dualism possible rather than triadism or some other form of pluralism by identifying the soul with intellect or reason and by treating the soul as a unity. It is not an oversimplification to say that in Plato the concepts of "soul" and "mind" merged; a merger made possible by incorporating a psychological interpretation of "spirit" into the soul-mind. In this way the concept of "spirit" or "life" was legislated out of the mind-body problem and out of dualism. Hence Plato should be considered the architect of dualism.

#### 1.3.4 Descartes

Descartes did more than defend a robust form of dualism: he also laid down the ground rules for discussion of the mind-body problem in modern philosophy. Descartes divided the universe into mental and physical entities and phenomena. He treated the terms 'soul' and 'mind' as being interchangeable and identified himself with his soul-mind in parenthetical remarks. Descartes' determination to bring the study of animals and the human body under the control of natural law together with his mental-physical dichotomy resulted in his claim that animals do not have souls or mind; that they have no feelings. Even today Descartes' mental/physical dichotomy is written into the formulation of "the mind-body problem".

Descartes, unlike Plato, is concerned to prove the distinctness of mind and body; he is less concerned to prove the survival of death by the soul, although he believed in this. Descartes gives his proof of distinctness of mind and body in the following famous passage:<sup>14</sup>

And first of all, because I know that all things which I apprehend clearly and distinctly can be created by God as I apprehend them, it suffices that I am able to apprehend one thing apart from another clearly and distinctly in order to be certain that the one is distinct from the other, since they may be made to exist in separation at least by the omnipotence of God; and it does not signify by what power this separation is made in order to compel me to judge them to be different: and, therefore, just because I know certainly that I exist, and that meanwhile I do not remark that any other thing necessarily pertains to my nature or essence, excepting that I am a thinking thing, I rightly conclude that my essence consists solely in the fact that I am a thinking thing (or a substance



whose whose essence or nature is to think). And although possibly (or rather certainly, as I shall say in a moment) I possess a body with which I am very intimately conjoined, yet because, on the one side, I have a clear and distinct idea of myself inasmuch as I am only a thinking and unextended thing, and as, on the other, I possess a distinct idea of body, inasmuch as it is only an extended and unthinking thing, it is certain that this I (that is to say, my soul by which I am what I am), is entirely and absolutely distinct from my body, and can exist without it.

In the preceding passage, Descartes states that it is sufficient for two things A and B to be ontologically distinct that he can form a clear and distinct apprehension of A apart from B. He then states that he can form a clear and distinct apprehension of himself as a thinking and unextended thing and also can form a clear and distinct apprehension of body as an extended and unthinking thing. From this he concludes that he is ontologically (entirely and absolutely) distinct from his body, and can exist without it.

Descartes' classic argument is worth examining because it illustrates so well the strengths and weaknesses of dualism. One strength it illustrates is that dualists have a fairly easy time demonstrating the distinctness of mind and body. Here Descartes picks dichotomous properties: thinking and unthinking, extended and unextended, as essential properties of mind and body. Picking dichotomous properties makes dualism fairly easy to prove since the same thing cannot be both thinking and unthinking or both extended and unextended. On the other hand, Descartes really proves the distinctness of himself from his body, not the distinctness of his mind from his body. Only in the last sentence does he mention his



soul, and there he says "My soul by which I am what I am". But this is hardly a claim of identity, and Descartes did not really think he was identical to his soul: he identified himself with the union of mind and body. The moral here is that even Descartes could not give a clear and consistent account of the relation between himself and his mind or soul.

Notice also that Descartes slips from a consideration of "my body" to "a distinct idea of body". Descartes held that body as such is a substance, but he still speaks of his body as "a body". This ambiguity is paralleled in the question "What is the relation between mind and body?". This sounds as if it asks about the relation between mind-as-a-whole and body-as-a-whole, but it is usually taken to ask about the relation between an individual mind and an individual body. Descartes also says that he is essentially unextended; this may be truer of his mind than of him but we do know that he had height and that he could be touched, that he took up space, and that he had weight. Descartes would not have to deny this if he maintained his claim that he is the union of body and mind, not that he is identical to mind. He would no doubt deny that these predicates apply to his soul or mind.

Another interesting feature of Descartes' argument is that he needs to find a generic term which will apply both to him and to physical objects. He chooses the French word 'chose', which we translate as "thing", for this purpose, and says that he is essentially a thinking and unextended thing while body is

essentially an extended and unthinking thing. In this way he sidesteps the question whether or not it is essential to him that he is a living thing i.e. a being. By using the term 'thing' to refer to himself, other persons, and physical objects, Descartes avoids the question about the status of life and maintains a strict mental/physical dualism.

In Chapter II I discuss an argument proposed by Saul Kripke which is similar to Descartes' "conceivability" argument, so I will not comment further on this argument here. One final comment about Descartes is that he identified mind with consciousness. While Plato identified mind with the intellect or reason, Descartes identified mind with consciousness and this identification has influenced discussion of the mind-body issue since then. Descartes also identified body with extension. Perhaps equally significant is the fact that Descartes made mental and physical a dichotomy, thereby making interaction between mind and body impossible to explain. Since Descartes' time, the mind-body problem has been discussed in the terms laid down by Descartes. It would not be correct to say that Descartes invented dualism - that distinction belongs to Plato - but it would not be incorrect to say that Descartes invented the mind-body problem.

### 1.3.5 Summary of Dualist Positions

Now that I have briefly summarized the views of the two most important dualists, Plato and Descartes, I conclude my discussion of dualism with a summary of several important dualistic theories. All these theories are defined by different accounts they give of mind-body interaction.

Interactionism. Interactionism is the view that mind and body interact; that mental phenomena sometimes cause physical phenomena and physical phenomena sometimes cause mental phenomena. Descartes, who set up the problem of interaction, was himself an interactionist. Interactionism has been defended in this century by C. D. Broad<sup>15</sup> and has been advocated by C. J. Ducasse.<sup>16</sup> I discuss the problem of interaction as a problem for dualism in some detail in Chapter VI.

Occasionalism. On the occasionalist view, causal interaction between mental and physical does not occur, but there are recurrent sequences. Occasionalists such as Arnold Geulinx and Nicolas de Malebranche proposed that God is the intermediary between mental and physical phenomena. When I will to move my arm, this is the occasion for God to move my arm. Occasionalists held that God's intervention is necessary for causal interaction, even between physical objects. They used the analogy of two clocks that keep in phase not because they have direct causal connection but because they have the same maker.

Parallelism. According to psychophysical parallelism, mental and physical phenomena are correlated in a regular way but without any causal connection. Leibniz took the occasionalist "two clocks" analogy and made them perfect mechanisms synchronized by God at their origin so that they remain forever in phase without intervention in virtue of the pre-established harmony. While parallelists relieve themselves of the burden of explaining causal interaction, they now have the burden of explaining constant correlation.

Epiphenomenalism. According to epiphenomenalism, the causal connection between body and mind goes only in one direction: from body to mind, or from physical to mental phenomena. This theory avoids the problems presented by accounting for the action of mind on body, but still has to account for how bodily action can produce mental phenomena. Mental events are sometimes compared to the smoke of a steam engine or to a movie on a screen. One way to explain the causal efficacy of mental events is to say that when a mental event seems to cause a physical event, as when a sensation of pain causes a wince, the same brain event which causes the pain also causes the wince. Epiphenomenalism has been defended by T. H. Huxley<sup>17</sup> and George Santayana.<sup>18</sup>

### 1.3.6 Conclusion

This concludes my discussion of dualism. First, I looked at Greek conceptions. Especially noteworthy here is the fact that the early Greeks did not have a word corresponding to our 'mind' but they did distinguish what today we would call several elements of the mind from bodily elements. I next discussed Plato, saying that he was the architect of dualism in that he identified the soul with the intellect of man or mind and argued that the soul exists before birth and also survives death. The discussion of Descartes took note of Descartes' unclarity about the relation between himself and his mind. It also credits Descartes with being the inventor of the mind-body problem in that he laid down the categories of mind and body in terms of which this problem is still discussed today. Finally, I summarized four types of dualistic theories which are distinguished based on the account each gives of interaction between mind and body. I now discuss materialism.

## 1.4 Materialism

Another important view on the mind-body problem is monism. Monism is the view that there is fundamentally one type of entity or phenomenon in the universe. Materialistic monism, or materialism, holds that this type of entity or phenomenon is material or physical

in nature. Like dualism, however, materialism has a conceptual and an ontological interpretation. Ontologically, materialism can be stated as follows:

- M(o). There is fundamentally one type of entity in the universe and that type of entity is physical or material.
- M(o)'. There is fundamentally one type of phenomenon in the universe and that type of phenomenon is physical or material.

Conceptually, materialism can be stated as follows:

- M(c). The term 'mind' as applied to phenomena represents a category of the same type as is represented by the term 'body'.
- M(c)'. The term 'mental' as applied to phenomena represents a category of the same type as is represented by the terms 'physical' and 'material'.

I now discuss three philosophers who advocated some version of materialism: Aristotle, Hobbes, and La Mettrie.

#### 1.4.1 Aristotle

It should not be surprising that in a philosopher whose works are as extensive as those of Aristotle that we should find that he advocates more than one conception of "soul". However, it is the views Aristotle defends in De Anima that are generally identified as representing his mature and most characteristic thought. Consequently, I mainly discuss Aristotle's De Anima views in this section.

Aristotle states the view that the soul is the form of the body in the following passage:<sup>19</sup>



But since it is also a body of such and such a kind, viz., having life, the body cannot be soul; the body is the subject or matter, not what is attributed to it. Hence the soul must be a substance in the sense of the form of a natural body having life potentially within it. But substance is actuality, and thus soul is the actuality of a body as above characterized. Now the word actuality has two senses corresponding respectively to the possession of knowledge and the actual exercise of knowledge. It is obvious that the soul is actuality in the first sense, viz. that of knowledge as possessed, for both sleeping and waking presuppose the existence of soul, and of these waking corresponds to actual knowing, sleeping to knowledge possessed but not employed, and, in the history of the individual, knowledge comes before its employment or exercise....

We have now given an answer to the question, What is soul?--an answer which applies to it in its full extent. It is substance in the sense which corresponds to the definitive formula of a thing's essence. That means that it is assigned the 'essential whatness' of a body of the character just assigned. Suppose that what is literally an 'organ', like an axe, were a natural body, its 'essential whatness', would have been its essence, and so its soul; if this disappeared from it, it would have ceased to be an axe, except in name. As it is, it is just an axe; it wants the character which is required to make its whatness or formulable essence a soul; for that, it would have had to be a natural body of a particular kind, viz. one having in itself the power of setting itself in movement and arresting itself. Next, apply this doctrine in the case of the 'parts' of the living body. Suppose that the eye were an animal--sight would have been its soul, for sight is the substance or essence of the eye which corresponds to the formula, the eye being merely the matter of seeing; when seeing is removed the eye is no longer an eye, except in name--it is no more a real eye than the eye of a statue or of a painted figure. We must now extend our consideration from the 'parts' to the whole living body; for what the departmental sense is to the bodily part which is its organ, that the whole faculty of sense is to the whole sensitive body as such.

Aristotle here is espousing a view quite different from the identity theory; in fact, he dismisses the mind-body question as we understand it today as meaningless:<sup>20</sup>



We can wholly dismiss as unnecessary the question whether the soul and the body are one: it is as meaningless as to ask whether the wax and the shape given to it by the stamp are one, or generally the matter of a thing and that of which it is the matter.

Aristotle's view is materialistic in that he does not regard the soul as a substance existing apart from body and does regard it as incapable of existence apart from body. The form of a body is for Aristotle the essence which makes it the type of body it is. Just as sight is the form or essence of the eye, so the faculty of sense is the form or essence of the body. The soul, then, is dependent on the body for its existence and is not an independent substance but it is the soul which makes the body the type of body it is.

It should be clear from this discussion that the type of contemporary materialism Aristotle's views are closest to is functionalism. Today we would say that sight is the function of the eye instead of the form of the eye. Because functionalism is the subject of Chapter IV, there is no need to critically evaluate Aristotle's views here. Before leaving Aristotle, however, I make several more comments.

In the above paragraph, Aristotle distinguishes between a natural and an artificial body. In Chapter VI I draw the same distinction and say that artificial objects (artefacts) have an essential property which is mental since it is essential to artefacts that they have a certain function which makes reference to a mental phenomenon such as an intention. For example, it is essential to a cup that it be used for drinking or that someone have

intended that it be used for drinking. Clearly, this is an Aristotelian claim. However, I do not take it as a materialistic claim for the following reason. If a statement of the function essentially involves reference to a mental phenomenon such as an intention, then whether this function is mental or physical depends on how one analyzes the mental element. Aristotle would analyze this intention as the form or function of some bodily state whereas I regard an intention as irreducibly distinct in kind from a bodily state. Hence a functional analysis of artefacts or even of neural states or mental states can be materialistic, dualistic, or even pluralistic depending on the ontological status of the entities involved.

Aristotle's concern with living organisms and with biology gave him a very different perspective on person, mind, and body from our contemporary Cartesian perspective. Although Aristotle's views are most correctly described as materialistic, his materialism is not of the same type as the identity theory. We could even classify his theory as dualistic in that the form (soul) of the body is viewed as distinct from the body. Aristotle also was aware of the importance of accounting for the organism or person as distinct from the mind or body. He also recognized the importance of the distinction between living and non-living things. However, my purpose here is to briefly sketch Aristotle's view on the mind-body problem and to

say what contemporary theory his theory most resembles. Since I have accomplished this purpose, I proceed to a consideration of a more modern materialist: Thomas Hobbes.

#### 1.4.2 Hobbes

Hobbes' political philosophy was more influential than his philosophy of mind. However, his philosophy of mind merits discussion since it represents a simple, uncompromising form of ontological materialism that is similar in spirit to the views of those who advocate the identity theory today. Hobbes was a contemporary of Descartes, and was the author of the Third Set of Objections to the Meditations. In Objection II, he quickly objects that one cannot infer from "I am exercising thought" to "I am thought", just as one cannot infer "I am the walking" from "I am walking".<sup>21</sup> Descartes replies that he meant what is endowed with the faculty of thinking i.e. mind, but it still does not follow that Descartes has shown his self to be identical to his mind. Hobbes' next conclusion is less sagacious. He says that Descartes' demonstration that he is essentially a thinking thing shows that that which thinks is material rather than immaterial. His reasoning is that knowledge of "I exist" depends on knowledge of "I think", which in turn depends upon the fact that we cannot separate thought from a matter that thinks<sup>22</sup>.

There are several sources for Hobbes' psychology.<sup>23</sup> However, Book I of the Leviathan gives a good idea of his basic principles.

He begins the Introduction to the Leviathan with a statement more characteristic of mechanism than of ontological materialism.<sup>24</sup>

Nature (the Art whereby God hath made and governes the World) is by the Art of man, as in many other things, so in this also imitated, that it can make an Artificial Animal. For seeing life is but a motion of Limbs, the beginning whereof is in some principall part within; why may we not say, that all Automata (Engines that move themselves by springs and wheeles as doth a watch) have an artificial life? For what is the Heart, but a Spring; and the Nerves, but so many Strings; and the Joynts, but so many Wheeles, giving motion to the whole Body, such as was intended by the Artificer? Art goes yet further, imitating that Rationall and most excellent worke of Nature, Man.

In the first chapter he speaks of sensible qualities and makes clear that his view is that they are nothing but bodies in motion:<sup>25</sup>

All which qualities called Sensible, are in the object that causeth them, but so many several motions of the matter, by which it presseth our organs diversely. Neither in us that are pressed, are they any thing else but divers motions; (for motion, produceth nothing but motion.) But their appearance to us is Fancy, the same waking, that dreaming. And as pressing, rubbing, or striking the Eye makes us fancy a light; and pressing the Eare, produceth a dinne; so do the bodies we see, or hear, produce the same thing by their strong, though unobserved action.

Finally, in Chapter VI he reiterates his view that sensation is bodily motion. Today, materialists would say that sensation is brain processes:<sup>26</sup>

There be in animals, two sorts of Motions peculiar to them: one called Vitall; begun in generation, and continued without interruption throughout their whole life; such as are the course of the Bloud, the Pulse, the Breathing, the Concoction, Nutrition, Excretion, &c.; to which Motions there needs no help of Imagination. The other is Animall motion, otherwise called Voluntary motion; as to go, to speak, to move any of our limbes, in such manner as is first fancied in our minds. That Sense, is Motion in the organs and interiour parts of mans body,

caused by the action of the things we See, Heare, &c; And that Fancy is but the Reliques of the same Motion, remaining after Sense, has been already sayd in the first and second Chapters.

Part I of the Leviathan is an attempt to give an account of man which prepares the way for his account of the state in Parts II and II. It is interesting to note the contrast with Plato: Plato based his account of the tri-partite nature of the soul in part on his analysis of the state and society. Hobbes, by contrast, uses his account of the psychology of human beings as the basis for his account of the state. Hobbes thought that men are naturally competitive, diffident, and glory-seeking, and this leads to a state of nature in which it is to everyone's advantage to lay down some liberty and form a commonwealth.<sup>27</sup> The psychology Hobbes develops in Part I is somewhat like that developed by Descartes in Passions of the Soul in that it is an attempt to classify some of the passions and other mental phenomena and show how they relate to each other. Much of this is independent of Hobbes' materialism, and is of interest in its own right.

I have already said that Hobbes' views most resemble those of today's identity theorists. Because I discuss the identity theory in Chapter II and VI, I move on to briefly consider a different sort of materialist: the mechanist La Mettrie.

### 1.4.3 La Mettrie

In the next section I distinguish five type of materialistic theory. One of these, mechanism, seems to be of a different type from the other four. Mechanism is the view that human beings are machines or machine-like, and that their behavior can be described, explained, and predicted using mechanical principles. The difference between mechanism and other forms of materialism is that these other forms of materialism are theories about what type of entity or phenomenon a mental entity or phenomenon is. These theories hold that a mental entity or phenomenon is identical to a physical entity or phenomenon, or can be reduced to or is a function of some physical entity or phenomenon. Mechanism is a theory about the explanation, description, and prediction of mental phenomena which does have implications about their nature. However, one might be an identity theorist or a functionalist and still hold that mental phenomena are to be explained on organic principles or even psychological laws - maintaining that these are reducible to physical but not necessarily mechanical laws.

Those who today argue for computer models of the mind would do well to read Julien La Mettrie. La Mettrie's L'Homme Machine, first published in 1747, is a classic defense of the mechanistic view which today is associated with cybernetics and computer models of the mind. La Mettrie's view was that man is a machine and that his behavior can be explained on mechanical principles. It is surely



one of the great ironies of history that Descartes should be the source of a doctrine that is quite contrary to his own philosophy. However, Descartes' view on animals was that they are machines, and it is not difficult to see how this view could be extended to human beings. This is what we find in the following quotation:<sup>28</sup>

Let us then permit the so-called Mr. Charp to laugh at the philosophers who regard animals as machines. Certainly I think differently! I consider Descartes as a man who is respectable in every regard. If he had not been born in a century so in need of enlightenment he would have earned recognition for his emphasis on experience, observation, and the danger of setting them aside. But let me make a small correction to this great man's work, for the sake of all those little philosophers, bad jokers, and poor imitators of Locke, who instead of laughing impudently at the birth of Descartes, would be far better off recognizing that without him the field of philosophy, like that of the spirit without Newton, might be largely undeveloped.

It is true that this celebrated philosopher made many errors; no one denies that. Yet he did come to know the nature of animals, and was the first to perfectly demonstrate that animals are pure machines. After a discovery of such importance, and one that requires such sagacity, the least we can do is forgive him all his errors!

In my opinion he compensates for all his errors with this major recognition. For while he makes much of the distinction between the two substances, it is obviously simply a skillful device, a stylistic ruse, to make palatable to the theologians a poison hidden in the shadow of an analogy which strikes everyone, and which only they fail to see. For it is truly this striking analogy that forces all intellectuals and true judges of mankind to admit that these proud and vain creatures, more distinguished by their pride than their name of Man despite attempts to raise themselves, are fundamentally only animals, and machines that crawl perpendicularly. They all have this marvelous instinct that becomes spirit through education, and which is seated in the brain, or in its place, if this is lacking, or becomes ossified, in the



elongated marrow, and never in the cerebellum; for I have seen it considerably damaged...without causing the soul to cease its functions.

Here La Mettrie credits Descartes with the view that animals are machines, and extends this view to human beings. His concluding remarks begin as follows:<sup>29</sup>

We therefore boldly conclude that man is a machine; and there is in the entire universe only a single substance that is diversely modified.

La Mettrie, then, was ontologically a materialist, but his form of materialism was mechanism.

I have already said that La Mettrie's mechanism has implications for the nature of mental phenomena. However, it is far from clear that these implications take the form of logical entailments. Aram Vartanian, a recent interpreter of La Mettrie, believes that La Mettrie has provided a picture or mechanical model of the human mind but not its essence.<sup>30</sup>

La Mettrie himself neglects, it is true, to discuss directly or to fix with care the scope and meaning of the materialism present in l'homme machine--a fact that has helped many of his critics to go astray. Nevertheless, he has scattered in the text a sufficient number of clues to reveal to a circumspect reader the spirit of his thought. Perhaps the most telling of these is the assertion that "la Nature du mouvement nous est aussi inconnue que celle de la matiere." If such is the case, then the nature of the "machine" with which man is hypothetically equated would likewise be unknown, and there could be no question of raising the correspondence between organic and psychic phenomena to the level of identity--an opinion which, in fact, La Mettrie nowhere attempts to establish....

The homme machine idea gives us, therefore, no more than a mechanical model, or a picturable analogy, of the mind--not its essence.

La Mettrie should not be thought of as a forerunner of identity theory but rather of contemporary philosophers and psychologists who construct mechanical models of the mind, particularly computer models.

There also seems to be some unclarity on the part of those who construct or argue for computer models of the mind on whether they are stating the essence of mind, saying what mind is, or merely providing analogies and models for the mind, saying what mind is like. Participants in these discussions seem to regard theories according to which the mind works like a computer as being materialistic; yet it does not seem contradictory to be ontologically a dualist and yet hold that the mind works like a computer or employs computer-like methods. One might hold, for example, that the brain stores memories in a way similar to the way a computer stores memories on a disk, and that the search procedure for recalling memories is similar: the brain is mapped out into addressable locations and a particular location is accessed according to its address. Such a model or picture does not logically require ontological materialism: the states which result from these searches do not have to be brain states. One might even be a dualist for computers and hold that certain internal computer states such as the computer's analyzing this week's payroll is not identical to a state of the hardware but rather is a distinct type of state. Hence, the association of mechanism with ontological materialism may be due more to the fact that most advocates of

mechanism and computer models of the mind are materialists rather than due to any logical entailment of one by the other.

Mechanism as a philosophy of mind and computer models of the mind do not play a major role in later chapters. Nevertheless, La Mettrie's work is important to take note of in this historical survey, both in virtue of its links with present day discussion of computer models of the mind and in virtue of its links with Descartes. We have already seen that the idea of l'homme machine originated in Descartes' view that animals are machines.

#### 1.4.4 Five Types of Materialism

The discussion of La Mettrie concludes my historical discussion of materialism. The purpose of this discussion has obviously not been to trace the history of materialism from Homer to the present, but rather to pick out several important philosophers from earlier periods and show how their views resemble views being discussed today. Of the three I have discussed, Aristotle's theory is closest to what today we call functionalism, while Hobbes' theory most closely resembles the identity thesis. Hobbes was apparently also a mechanist. La Mettrie was a mechanist, as we have seen, and his views most closely resemble those of philosophers today who construct and defend computer models of the mind. I said that La Mettrie's mechanism, and mechanism in general, does not logically entail materialism; I have classified it as a materialist view because it is generally regarded as a form of materialism and

because most mechanists are materialists. Even if a mechanical model of the mind does not give the essence of mind, it does give a material model.

The three forms of materialism discussed in detail in later chapters are the identity theory, anomalous monism, and functionalism. I have discussed Aristotle as an antecedent to functionalism and Hobbes as an antecedent to the identity theory. My survey seems to provide no historical precedent for anomalous monism, Donald Davidson's materialistic theory which is the subject of Chapter III. Spinoza is the closest historical antecedent to Davidson, and while his theory is monistic, it is not a form of materialism. Hence I discuss Spinoza's theory as a separate type of monism in the next section. First, however, I distinguish five types of contemporary materialism.

The Identity Theory. According to the identity theory, mental events are identical to brain events. This thesis can also be formulated as the view that any individual mind is identical to some brain, or that the person is identical to his body. This theory was popular in the 1960's, and was defended by philosophers such as J. J. C. Smart<sup>31</sup> and D. M. Armstrong.<sup>32</sup> Philosophers advocating this view used a notion they called "strict" identity<sup>33</sup> according to which mental-physical identities satisfy the condition of Leibniz's Law.<sup>34</sup> Criticisms of the identity theory led some philosophers to retreat to theoretical materialism, functionalism, or eliminative materialism.

Theoretical Materialism. According to theoretical materialism, statements of identity between mental and physical events resemble certain identifications found in physical theories. One example sometimes used is the identification of temperature with mean molecular energy. This type of identity is sometimes called contingent identity since these statements are supposed to be contingently true, not necessarily true. Thomas Nagel defends this type of view in "Physicalism"<sup>35</sup> and Jaegwon Kim critically discusses it in "On the Psycho-Physical Identity Theory".<sup>36</sup> This theory could be considered a different formulation of the identity theory, with a weaker sense of identity. It is this form of materialism that Saul Kripke argues against in Naming and Necessity.<sup>37</sup>

Functionalism. According to functionalism, mental states are correctly accounted for by giving a functional analysis or account of physical (brain) states. Functionalism was developed in part in response to the objection that a creature unlike human beings might have similar mental states even though its neural structure is quite unlike ours. The functions of its brain states might still be similar. Functionalism was developed also in response to computer models of the mind, and has been defended by Hilary Putnam<sup>38</sup> and Daniel Dennett.<sup>39</sup> I discuss Dennett's version of functionalism in Chapter IV.

Eliminative Materialism. According to eliminative materialism, terms which refer to apparently mental phenomena in fact refer to physical phenomena. Just as the progress of science revealed that there are no witches, so progress in describing and explaining mental phenomena will reveal that phenomena which were thought to be mental are really physical. In this way, talk of mental phenomena will be eliminated from our vocabulary. Probably the chief defender of this view is Richard Rorty.<sup>40</sup> Since I do not discuss this version of materialism in later chapters, I will point out here that the analogy it draws between mental phenomena and witches and demons does not withstand serious examination. This does not mean that there is never any circumstance in which a mental term is replaced or is found to be inapplicable. We have already seen that the Homeric conceptions of thymos, noos, and psyche were replaced in Plato's time by a conception according to which these elements are elements of a unity called "soul" or "mind" rather than being separate organs which are no part of a greater whole. But to admit this analogy is quite different from admitting an analogy whereby all our mental terminology could be eliminated.

Mechanism. Mechanism is the view that human and animal behavior is to be explained, predicted, and described according to mechanical principles. In La Mettrie's time philosophers used clocks as models of human behavior and mental phenomena; today they use computers. Because this theory is about explanations and models of what human beings do rather than about the nature of mental



phenomena, it is not logically inconsistent with dualism. However, it is generally considered to be a materialistic theory, perhaps because most of its advocates are materialists. We saw that Hobbes was a mechanist, and that La Mettrie is the classic exponent of mechanism. A mechanical concept of mind has been defended in recent times by A. M. Turing<sup>41</sup> and by Hilary Putnam.<sup>42</sup>

#### 1.4.5 Summary

This completes my discussion of materialism. I discussed Aristotle as a forerunner of contemporary functionalism, Hobbes as a forerunner of the identity theory, and La Mettrie as a mechanist and a forerunner of theories which involve constructing computer models of the mind. I also distinguished five different materialist theories: the identity theory, theoretical materialism, functionalism, eliminative materialism, and mechanism. I now discuss the views of another monist whose form of monism is not materialistic: the plural aspect monism of Benedict de Spinoza.

### 1.5 Plural Attribute Monism

In my discussion of dualism, I made a distinction between dualism of substance and dualism of properties.<sup>43</sup> I then went on to state my intention to use the term 'phenomena' as a general term to cover properties, events, states, conditions, abilities, capacities, and other phenomena other than entities that can be



either mental or physical. Of course, there are still important distinctions within the class of phenomena as I have described it. There will still be occasions when we need to pick out some subclass of this class for special attention. We still may wish to single out dualism of properties or dualism of events, for example.

Such is the case in relation to Spinoza, who held a view best described as plural attribute monism. I have selected Spinoza for discussion because his view is an antecedent of Donald Davidson's view, which is discussed in Chapter III. Though Spinoza defended an ontological form of his theory, I state both an ontological and a conceptual version of plural attribute monism:

A(o). Minds and bodies are modes of the attributes of thought and extension. They are two of many attributes, all of which attributes are attributes of a single underlying substance.

A(o)'. Mental and physical attributes are two of many types of attributes, all of which attributes are attributes of a single underlying substance.

Conceptually, plural attribute monism can be stated as follows:

A(c). The terms 'mind' and 'body' represent two irreducibly distinct conceptual categories of attribute modes. These are two of many such irreducibly distinct conceptual categories of attribute modes. There is a single conceptual category of substance which these conceptual categories of attribute modes apply to.

A(c)'. The terms 'mental' and 'physical' represent two irreducibly distinct conceptual categories of attributes. These are two of many such irreducibly distinct conceptual categories of attributes. There is a single conceptual category of substance which these conceptual categories of attributes apply to.

### 1.5.1 Spinoza

Spinoza lived from 1632 - 1677. placing him between Descartes and La Mettrie. He was heavily influenced by Descartes and, like many philosophers of his time, developed many of his views in response to those of Descartes. His form of monism cannot be described either as materialism or idealism, but it is only a form of ontological monism. He did believe that there is only one substance in the universe, and he identified this substance with God. He held that this one substance, God, has infinitely many attributes, including mind and body. It is his view that God has infinitely many attributes that makes him a plural aspect monist.

Spinoza's view of substance is central to his philosophy. He held that the essence of a substance involves existence:<sup>44</sup>

PROPOSITION VII. It pertains to the nature of substance to exist.

Demonstration. There is nothing by which substance can be produced....It will therefore be the cause of itself, that is to say...its essence necessarily involves existence, or, in other words, it pertains to its nature to exist.--Q.E.D.

He denies that man is a substance in these remarks:<sup>45</sup>

PROPOSITION X. The being of substance does not pertain to the essence of man, or, in other words, substance does not constitute the form of man.

Demonstration. The being of substance involves necessary existence.... If, therefore, the being of substance pertained to the essence of man, the existence of man would necessarily follow from the existence of substance..., and consequently he would necessarily exist, which...is an absurdity. Therefore the being of substance does not pertain, etc.--Q.E.D.

Finally, he makes it clear that God is the only Substance:<sup>46</sup>

PROPOSITION XIV. Besides God no substance can be nor can be conceived.

Demonstration. Since God is Being absolutely infinite, of whom no attribute can be denied which expresses the essence of substance..., and since He necessarily exists..., it follows that if there were any substance besides God, it would have to be explained by some attribute of God, and thus two substances would exist possessing the same attribute, which is absurd; and therefore there cannot be any substance except God, and consequently none other can be conceived. For if any other could be conceived, it would necessarily be conceived as existing, and this (by the first part of this demonstration) is absurd. Therefore, besides God no substance can be nor can be conceived.--Q.E.D.

Spinoza not only held that an individual mind, an individual body, and an individual person are not substances, he also held that mind as a whole and body as a whole are not substances but rather attributes of the one infinite substance: God. This is clear from the following propositions:<sup>47</sup>

PROPOSITION I. Thought is an attribute of God, or God is a thinking thing.

PROPOSITION II. Extension is an attribute of God, or God is an extended thing.

Spinoza is probably the only philosopher I discuss who would understand the question "What is the relation between minds and body?" as a question about the relation between mind-as-a-whole and body-as-a-whole rather than as a question about the relation between an individual mind and an individual body.<sup>48</sup> He says that "the human mind is a part of the intellect of God".<sup>49</sup> Although Spinoza does seem to recognize the existence of individual minds and bodies,

these individuals are not substances and are correctly understood as modes of the attributes of thought and extension, which are attributes of God.

Spinoza's definition of 'attribute' is as follows:<sup>50</sup>

IV. By attribute I understand that which the intellect perceives of substance as constituting its essence.

He applies this definition to thought and extension in these remarks:<sup>51</sup>

Substance thinking and substance extended are one and the same substance, which is now comprehended under this attribute and now under that....Whether we think of Nature under the attribute of extension or under the attribute of thought or under any other attribute whatever, we shall discover one and the same order or one and the same connection of causes, that is to say, in every case the same sequence of things....When things are considered as modes of thought we must explain the order of the whole of Nature or the connection of causes by the attribute of thought alone, and when things are considered as modes of extension, the order of the whole of Nature must be explained through the attribute of extension alone, and so with other attributes.

Spinoza's view, then, is that God can be conceived either under the mode of thought or under the mode of extension. Extension is an attribute of God as conceived under this mode and thought is an attribute of God as conceived under the mode of thought. Today we might say that the categories of mental and physical represent two points of view in terms of which persons or objects can be described.

I maintain during discussion of Davidson that Spinoza's view is an ancestor of Davidson's anomalous monism. This is so for the following reason. Davidson has referred to his view as a combination of conceptual dualism and ontological monism.<sup>52</sup> He

does not employ the notion of a point of view, but he does introduce the idea of a conceptual domain. Davidson thinks that the mental conceptual domain is irreducible to the physical conceptual domain, and vice versa. This is similar to Spinoza's view that extension and thought are two distinct modes for conceiving substance.

However, Davidson is a conceptual dualist, not a conceptual pluralist: he does not recognize the existence of conceptual domains other than the mental and physical conceptual domains. Furthermore, his form of monism is materialism; he does not believe that his conceptual dualism implies ontological dualism and he argues for the identity theory. For Spinoza, on the other hand, his one substance has infinitely many attributes and it is not just extended and thinking but also necessarily existing, indivisible, etc. The true sense, then, in which Spinoza is a monist is that he held that there is only one substance but he is a pluralist in that he held that this one substance has infinitely many attributes.

## 1.6 Idealism

Idealism is a radically different form of monism in that according to idealism the single type of substance in the universe is mental rather than physical. The classic defense of this view is given in Berkeley's Principles of Human Knowledge.<sup>53</sup> The spirit of idealism is still alive today in the form of phenomenism, an epistemological theory which attempts to reduce material objects to

sense-data. Phenomenalism has been defended in this century by Bertrand Russell<sup>54</sup> and by A. J. Ayer.<sup>55</sup> Because I do not discuss either idealism or phenomenalism in later chapters, I do not discuss them further here. Notice, however, that both materialism and idealism depend on the existence of the mental-physical distinction for the very statement of their respective theories.

### 1.7 Pluralism

Pluralism is sometimes identified with the view that there are many irreducibly distinct individuals in the universe. While few philosophers would be inclined to deny this today, we have just seen that Spinoza held that there is only one substance in the universe, and this substance is God. Descartes held that what we might think of as individual bodies are modes of one extended substance. The view that there are many irreducibly distinct individuals in the universe is pluralism of individuals. This view might also be described as pluralism of entities or substances. Even a materialist who believes that there is only one type of entity in the universe, viz. material entities, might still hold that there are many such individual entities. Such a person would be a monist in types of entities and a pluralist of individual entities.

The type of pluralism which is of most interest in relation to this discussion of the mind-body problem is pluralism in types of entities and phenomena. This is the view that there are many



fundamentally distinct types of entity or phenomena in the universe, where this number is greater than two. Like monism and dualism, pluralism of types has an ontological and a conceptual interpretation:

P(o). There are many fundamentally distinct types of entities in the universe.

P(o)'. There are many fundamentally distinct types of phenomena in the universe.

Conceptually, this view can be formulated as follows:

P(c). The terms 'mind' and 'body' represent two of many irreducibly distinct categories or conceptual domains which apply to entities in the universe.

P(c)'. The terms 'mental' and 'physical' represent two of many irreducibly distinct categories or conceptual domains which apply to entities in the universe.

P(c)''. The terms 'mental' and 'physical' represent two of many irreducibly distinct categories or conceptual domains which apply to types of phenomena in the universe.

We have just seen that Spinoza would deny P(o) since he held that there is only one substance in the universe. He would ascribe to P(o)', however, in light of his view that there are infinitely many attributes in the universe. Just as he would deny P(o), he would deny P(c), reading "substance" for 'entity'; however, he would accept P(c)'. We can say, then, that Spinoza accepted conceptual and ontological plurality of phenomena but denied conceptual and ontological plurality of substance. Spinoza would say that he accepted plurality of attributes.

### 1.7.1 Leibniz

I have already made note of the fact that Leibniz was a parallelist in section 1.3.5. Since Leibniz held that bodies are phenomenal in nature he is often counted as a phenomenalist or idealist. My concern here is not with either of these strains of his thought, but with his view on points of view: Leibniz was a pluralist with respect to points of view or perspectives.

Leibniz held that each monad has a different point of view:<sup>56</sup>

56. This connection of all created things with every single one of them and their adaptation to every single one, as well as the connection to every single one, as well as the connection and adaptation of every single thing to all others, has the result that every single substance stands in relations which express all the others. Whence every single substance is a perpetual living mirror of the the universe.

57. Just as the same city regarded from different sides offers quite different aspects, and thus appears multiplied by the perspective, so it also happens that the infinite multitude of simple substances creates the appearance of as many different universes. Yet they are but perspectives of a single universe, varied according to the points of view, which differ in each monad.

58. This is the means of obtaining the greatest possible variety, together with the greatest possible order; in other words, it is the means of obtaining as much perfection as possible.

Leibniz's view that each monad has a different point of view both provides a principle of individuation for monads and provides a basis for plurality of monads and substances. Leibniz says that the same city viewed from different sides offers different aspects so he, like Spinoza, can be described as having a plural aspect view.

Unlike Spinoza, he held that there is an infinity of individual substances, so he held that there are infinitely many individual substances each having infinitely many aspects.

### 1.8 Conclusion

To summarize, I began this historical survey with an introduction to the mind-body problem as a philosophical problem. The main body of the chapter was concerned with an examination of dualism, three types of materialism, and pluralism. In section 1.3 I described the views of Plato and Descartes as examples of conceptual and ontological dualism. In section 1.4 I identified Aristotle as a historical antecedent of contemporary functionalism, and Hobbes as a historical antecedent of the identity theory. I discussed the views of both these materialists. I also considered La Mettrie as the classic exponent of mechanism. In section 1.5 I considered Spinoza as a historical antecedent of anomalous monism. After taking note of idealism, I considered Leibniz in section 1.7 as a pluralist in points of view.

This concludes my historical survey of the mind-body problem. The main purpose of this survey has been to show that the four contemporary theories on the mind-body problem which I examine in succeeding chapters: dualism, the identity theory, anomalous monism, and functionalism are all contemporary formulations of traditional theories on this problem. Hence, many of my criticisms of

contemporary views should apply to these traditional theories as well. A secondary purpose has been to show where these four theories stand in relation to other theories on the mind-body problem. In some cases, criticisms of a version of one type of theory such as the identity theory apply to other versions of the same type of theory such as functionalism. Hence, I believe that the discussion in the remaining chapters represents a fairly broad discussion of the mind-body problem which has implications beyond its implications for the four particular theories I examine.

## FOOTNOTES

1. Obviously, there are many problems traditionally discussed by philosophers in non-philosophical contexts which are not philosophical problems. The main point here is to introduce the mind-body problem as a philosophical problem rather than a practical problem.
2. I take dualism to be a claim about what types of entities or phenomena there are in the universe. Philosophers sometimes use the term 'world' in what I take to be a synonym for 'universe'. For example, there is talk of "possible worlds", in which it is imagined that history remains the same except that one event or sequence of events is different. These kinds of discussions sometimes occur in relation to considering counterfactuals. Since the history of our earth includes its relations to other entities in the universe, such examples are about possible universes, not merely possible earths or possible solar systems. A mind-body theory which limited itself to saying what entities or phenomena exist on our earth but had nothing to say about the rest of the universe would be of only limited interest. In this chapter and in succeeding chapters, I use the term 'world' as a synonym for 'universe' unless I specify otherwise or unless the context makes it clear that I am using it in some other sense.
3. Bruno Snell, The Discovery of the Mind (Cambridge, Massachusetts: Harvard University Press, 1953), p. 7.
4. Ibid., p. 15.
5. Ibid., pp. 16-7.
6. Ibid., pp. 17-8.
7. Ibid., p. 18.
8. G. M. A. Grube, Plato's Thought (Boston, Massachusetts: Beacon Press, 1964), p. 121.
9. An exception to this is Gareth Matthews. In an early version of a paper entitled "Life and Death as the Arrival and Departure of the Psyche", he says "This paper is part of a rather ambitious project aimed at recommending the traditional concept of the psyche as an alternative to the Cartesian concept of mind. Here are two crucial differences between the concept of a psyche and the (Cartesian) concept of a mind: (1) whereas a psyche is an animator, a mind has nothing essential

to do with life (see my 'Consciousness and Life,' Philosophy 52 (1977), 13-26; (2) whereas every living animal has a psyche, only human beings, angels and God, according to Descartes, have minds (see my 'Animals and the Unity of Psychology' 53 (1978), 437-54). While I endorse Matthews' attempt to call attention to the notion of psyche and to its importance, I would view it as a supplementary notion to the Cartesian concept of mind rather than as an alternative.

10. Plato, Phaedo, translated by Benjamin Jowett, Vol. I, The Dialogues of Plato (New York: Random House, 1937), 77e. p. 462.
11. Ibid., 88a, p. 472.
12. It is in the Phaedrus that Plato compares the soul to a charioteer driving a pair of winged horses. See Plato, Phaedrus, 246a, p. 250. I do not discuss this analogy, but instead use Plato's discussion of the soul in The Republic as the basis for my discussion of the tripartite division of the soul.
13. Plato, The Republic, translated by Paul Shorey, Vol I, Republic (Cambridge, Massachusetts: Harvard University Press, 1969), 439c-e, pp. 397-99.
14. Rene Descartes, The Philosophical Works of Descartes, trans. Elizabeth S. Haldane and G. R. T. Ross, Vol. I (New York, Cambridge University Press, 1972), p. 190.
15. Refer to C. D. Broad, The Mind and Its Place in Nature (London: Kegan Paul, Trench, Trubner, & Co., Ltd., 1947), esp. pp. 95-133.
16. Refer to C. J. Ducasse, The Belief in a Life After Death (Springfield, Illinois: Charles C. Thomas, 1961), esp. pp. 59-117.
17. T. H. Huxley, "On the Hypothesis that Animals are Automata and Its History", in Collected Essays (London: 1893-4).
18. George Santayana, The Realm of Essence (New York: 1927) and Reason and Common Sense (New York: 1922).
19. Aristotle, De Anima, in The Basic Works of Aristotle, ed. Richard McKeon (New York: Random House, 1941), 412a-b, pp. 555-6.
20. Ibid., 412b, p. 555.



21. Descartes, 1972, Vol. II, p. 61.
22. Ibid., p. 62.
23. Besides the Leviathan (see next footnote for reference), there is also De Corpore; see The English Works of Thomas Hobbes of Malmesbury, collected and edited by Sir William Molesworth (London: John Bohn, 1839), Vol. I, Part IV, esp. pp. 389-94 and 396-99.
24. Thomas Hobbes, Leviathan, ed. Francis B. Randall (New York: Washington Square Press, 1964), p. xxvii.
25. Ibid., p. 2.
26. Ibid., p. 29.
27. Ibid., pp. 115-9.
28. La Mettrie, L'Homme Machine, in La Mettrie's L'Homme Machine. ed. Aram Vartanian (Princeton, New Jersey: Princeton University Press, 1960), pp. 191-2. I wish to thank Stanley Yoder for providing me with a translation of this passage.
29. Ibid., p. 197.
30. Ibid., pp. 15-6. This quotation is from Vartanian's Chapter II. Besides giving an interpretation of L'Homme Machine, Vartanian discusses the historical background of the work and the critical reaction of La Mettrie's contemporaries. The text of L'Homme Machine appears (in French) as the last section of the book.
31. Refer to J. J. C. Smart, "Sensations and Brain Processes", in The Philosophy of Mind, ed. Vere C. Chappell (Englewood Cliffs, New Jersey: Prentice Hall, Inc., 1962), pp. 160-72 and J. J. C. Smart, Philosophy and Scientific Realism (New York: The Humanities Press, 1963), esp. Chapters IV, V, and VI. The Mind/Brain Identity Theory, ed. C. V. Borst (New York: St. Martin's Press, 1975) also reprints "Sensations and Brain Processes" and contains a critical discussion of this article. This entire volume is devoted to a discussion of the identity theory and contains a number of other important articles.
32. Refer to D. M. Armstrong, A Materialist Theory of Mind (New York: Humanities Press, 1968).
33. Refer to J. J. C. Smart, "Sensations and Brain Processes" in The Philosophy of Mind, ed. Vere C. Chappell, p. 163 for an explanation of "strict" identity.

34. While I believe that objections to the identity theory based on Leibniz's Law are effective, I do not emphasize these objections in subsequent chapters. The reason for this is that much recent discussion of materialism focuses on functionalism, while objections based on Leibniz's Law are mainly directed against the identity theory. I discuss functionalism in Chapter IV.
35. Thomas Nagel, "Physicalism", in The Mind/Brain Identity Theory, ed. C. V. Borst, pp. 214-30.
36. Jaegwon Kim, "On the Psycho-Physical Identity Theory", American Philosophical Quarterly, 3 (1966), pp. 227-35.
37. Saul Kripke, Naming and Necessity, in Semantics of Natural Languages, Gilbert Harman and Donald Davidson (eds.) (Dordrecht: Reidel, 1971). Kripke's views on the mind-body problem as they are expressed in Naming and Necessity are the subject of the next chapter (Chapter 2).
38. Hilary Putnam, Mind, Language and Reality, Philosophical Papers (Cambridge: Cambridge University Press, 1975), Chapters 14-22.
39. Daniel Dennett, Brainstorms (Cambridge: Bradford Books, 1981).
40. Refer to Richard Rorty, "In Defense of Eliminative Materialism", in Materialism and the Mind-Body Problem, ed. David M. Rosenthal, (Englewood Cliffs, New Jersey: Prentice Hall, Inc., 1971), pp. 223-31.
41. A. M. Turing, "Computing Machinery and Intelligence", in Mind, Vol. LIX, No. 236 (1950).
42. Putnam, 1975.
43. See above, Section 1.3.
44. Spinoza, Ethics, ed. Frederick Woodbridge (New York: Hafner Publishing Co., 1953), p. 44.
45. Ibid., p. 86.
46. Ibid., p. 51.
47. Ibid., pp. 80-1.
48. Descartes also held that what we call individual bodies are modes of a single extended substance or body. However, Descartes did believe that there are individual minds.

49. Spinoza, 1953, p. 88.
50. Ibid., p. 41.
51. Ibid., p. 84.
52. In a Philosophy of Mind tutorial I had with Professor Davidson at The Rockefeller University in Spring, 1975.
53. Berkeley, A Treatise Concerning the Principles of Human Knowledge, ed. Colin M. Turbayne (New York: The Liberal Arts Press, 1957).
54. Refer to Bertrand Russell, Our Knowledge of the External World (New York: Mentor Books, 1960), esp. pp. 54-102.
55. Refer to A. J. Ayer, The Foundations of Empirical Knowledge (London: 1940) and "Phenomenalism", in Proceedings of the Aristotelian Society, Vol. 47, 1946/1947, pp. 163-96.
56. Leibniz, Monadology, in Monadology and Other Philosophical Essays, trans. Paul Schrecker and Anne Martin Schrecker (New York: Bobbs-Merrill Company, Inc., 1965), pp. 156-7.

## C H A P T E R   I I

### SAUL KRIPKE'S DEFENSE OF DUALISM

#### 2.1 Introduction

Dualism is one solution to the mind-body problem that has had many capable defenders in the past, including Descartes, Locke, and Hume. Like many other theories, dualism has been more popular at some times than others. It has not been a popular theory in this century; however, interest in dualism has increased in the past ten years, partially in response to criticisms made of materialism and behaviorism. Dualism has also recently found a capable exponent in Saul Kripke, whose views are a consequence of his views in theory of reference and proper names.

In this chapter, I state two related arguments for dualism that Saul Kripke presented recently in Naming and Necessity.<sup>1</sup> Because these arguments follow from Kripke's theory of proper names, I begin by explaining this theory. I then examine some important criticisms of these arguments advanced by Fred Feldman in several recent articles.<sup>2</sup> In the last section, I discuss what I take to be the most serious problem for dualism: the problem of interaction. I

conclude with several observations which seem to show that mind and body are more closely related than they might appear to be at first sight.

## 2.2 Kripke's Theory of Proper Names

In Naming and Necessity, Kripke presents an argument against several versions of the identity thesis. This thesis as it is usually formulated is a form of materialism which claims either that a person is identical to his body, that each mind is identical to some body, or that mental states or events are identical to brain states or bodily states or events. The claim that mental states or events are identical to brain states or bodily states is sometimes formulated in terms of particular states, and sometimes in terms of states of some kind. Kripke's argument makes use of some concepts that he developed in order to formulate his theory of proper names. Because these concepts play a key role in Kripke's argument against the identity thesis, I begin by explicating Kripke's theory of proper names and general terms.

Kripke's theory of proper names represents a break from the tradition of Gottlieb Frege<sup>3</sup> and Bertrand Russell,<sup>4</sup> according to which proper names are abbreviated descriptions; Kripke's theory is much closer to John Stuart Mill's,<sup>5</sup> who held that names have denotation but no connotation. Kripke's theory is actually somewhat complex, and for our purposes it is sufficient to discuss those

aspects that relate directly to the mind-body argument. Those related aspects particularly include the view that names and some general terms are rigid designators, that identity statements between rigid designators are, if true, necessarily true, and that the origin of a thing is an essential property of that thing.

### 2.2.1 Names and Some General Terms are Rigid Designators

I begin by discussing designators. Kripke makes the following comments:<sup>6</sup>

One of the intuitive these I will maintain in these talks is that names are rigid designators. Certainly they seem to satisfy the intuitive test mentioned above: although someone other than the U.S. President in 1970 (e.g. Humphrey might have), no one other than Nixon might have been Nixon. In the same way, a rigid designator rigidly designates a certain object if it designates that object wherever the object exists.

The distinction Kripke draws here is between a rigid (necessary) designator, which designates the same object in every world in which the object exists, and a non-rigid (accidental) designator which may designate different objects in different possible situations (or "possible worlds"). Kripke holds that proper names, such as 'Nixon', are rigid designators, while definite descriptions, such as 'The U.S. President in 1970', are non-rigid. One difference is that, while one can imagine situations in which the term 'The U.S. President in 1970' does not denote Nixon, the term 'Nixon' always refers to the same man in every imaginable situation.

Kripke also apparently believes that some general terms, such as 'gold' and 'tiger' are rigid designators. He says that:<sup>7</sup>



According to the view I advocate, then, terms for natural kinds are much closer to proper names than is ordinarily supposed. The old term 'common name' is thus quite appropriate for predicates marking out species or natural kinds, such as 'cow' or 'tiger'. My considerations apply also, however, to certain mass terms for natural kinds, such as 'gold', 'water', and the like.

Elsewhere, he says:<sup>8</sup>

The original concept of cat is: that kind of thing, where the kind can be identified by paradigmatic instances. It is not something picked out by any qualitative dictionary definition.

Kripke rejects the view that 'tiger' has as its meaning a conjunction or disjunction of properties, of which some or all must obtain for the term to be applicable to something. Rather, certain general terms refer to some kind of thing, where the kind of thing can be shown by example, and where the kind remains uniform across all possible worlds.

In order to elucidate this claim, Kripke contrasts "fixing the reference" of an expression with "giving the meaning" of it. Using the Wittgensteinian example of the standard meter, he states that someone who stipulates that a stick S which is to serve as a standard of length for the term 'meter' is not laying down a definition of 'meter', but, rather, fixing its reference. Hence, the phrase 'one meter' is not synonymous with the phrase 'the length of S at t'; rather, 'one meter' denotes the same length in all possible worlds.

The importance of Kripke's view that some general terms are rigid designators will become apparent shortly when I consider his

claim that 'pain' is a rigid designator. For now, I summarize his view as follows:

1. Names and some general terms are rigid designators.
2. A rigid designator is a term which denotes the same object in all possible worlds.
3. A rigid designator is not synonymous with a definite description denoting its reference or with some cluster of properties; but its reference can be fixed via a definite description.

#### 2.2.2 Identity Statements Between Rigid Designators are Necessary

Kripke not only holds that names and some general terms are rigid designators, but he also holds that identity statements between rigid designators, if true, are necessarily true. Kripke tells us:<sup>9</sup>

I hold the following about the general case. First, that characteristic theoretical identifications like 'Heat is the motion of molecules', are not contingent truths but necessary truths, and here of course I don't mean just physically necessary, but necessary in the highest degree - whatever that means.

Elsewhere Kripke says that 'heat' is a rigid designator, so "Heat is the motion of molecules" is apparently an example of a necessary truth between two rigid designators. The difference between necessarily true identity statements and contingently true identity statements is that, for the latter, at least one of the terms on either side is an accidental designator, and hence does not denote the same object in all possible worlds. But if the terms are rigid, then, they must designate the same object in all possible worlds, and hence the identity is necessary. An example of an accidentally

true identity statement is "Benjamin Franklin is the inventor of bifocals", while an example of a necessarily true identity statement is "Hesperus is Phosphorus".

Kripke's view that identity statements between rigid designators are necessarily true plays a key role in his argument against the identity thesis. For the moment, his view can be summarized as follows:

4. An identity statement of the form 'a = b' is, if true, necessarily true, if 'a' and 'b' are rigid designators.
5. An identity statement of the form 'a = b' is, if true, accidentally true if either 'a' or 'b' is a non-rigid designator.

### 2.2.3 Origin is an Essential Property

Another aspect of Kripke's metaphysical views that have implications for his mind-body argument is his view that the origin of a thing is an essential property of the thing. Considering the example of a wooden table, Kripke comments:<sup>10</sup>

In the case of this table, we may not know what block of wood the table came from. Now could this table have been made from a completely different block of wood, or even of water cleverly hardened into ice - water taken from the Thames River? We could conceivably discover that, contrary to what we now think, this table is indeed made of ice from the river. But let us suppose that it is not. Then, though we can imagine making a table out of another block of wood or even from ice, identical in appearance with this one, and though we could have put it in this very position in the room, it seems to me that this is not to imagine this table as made of wood or ice, but rather it is to imagine another table, resembling this one in all external details, made of another block of wood, or even of ice.

Kripke also considers the question whether Queen Elizabeth could

have been born of different parents from the parents from whom she actually came. He concludes that "Anything coming from a different origin would not be this object".<sup>11</sup> Kripke cites these cases as examples of examples of essential properties, and does not attempt to state a general theory about origin and essence. However, it does seem clear that he endorses the following principle in these (and related) passages in Naming and Necessity:

6. The origin of a thing is an essential property of the thing.

This principle will become important as I consider questions about the origin of mental and physical events. However, I will now summarize the general discussion of Kripke's metaphysical views and then turn to a consideration of his arguments against the identity thesis.

#### 2.2.4 Summary of Kripke's Metaphysical Views

I have now formulated six general principles that Kripke endorses in Naming and Necessity, and that relate to his view on the mind-body problem. These principles are as follows:

1. Names and some general terms are rigid designators.
2. A rigid designator is a term that denotes the same object in all possible worlds.
3. A rigid designator is not synonymous with a definite description denoting its reference or with some cluster of properties; but its reference can be fixed via a definite description.
4. An identity statement of the form ' $a = b$ ' is, if true, necessarily true, if ' $a$ ' and ' $b$ ' are rigid designators.

5. An identity statement of the form 'a = b' is, if true, accidentally true if either 'a' or 'b' is a non-rigid designator.
6. The origin of a thing is an essential property of the thing.

I do not discuss criticisms of these principles now, or questions about their meaning or their precise formulation. Rather, I move on to a consideration of Kripke's argument against the identity thesis, while keeping these principles in mind.

### 2.3 Kripke's Arguments Against the Identity Theses

In this section, I state two different arguments that Kripke formulates against the identity thesis. The conclusion of the first argument is that a person is distinct from his body, while the conclusion of the second argument is that a feeling of pain is distinct from a brain state. I consider some criticisms which have been made of these arguments in section 2.4.

#### 2.3.1 Kripke's Argument that a Person is Distinct from his Body

One form the identity thesis takes is the claim that a person is identical to his body. Kripke states an argument against this claim in the following passage:<sup>12</sup>

Descartes, and others following him, argued that a person or mind is distinct from his body, since the mind could exist without the body. He might equally well have argued the same conclusion from the premise that the body could have existed without the mind. Now the one response which I regard as plainly inadmissible is the response which cheerfully accepts the Cartesian premise while denying the

Cartesian conclusion. Let 'Descartes' be a name, or rigid designator, of a certain person, and let 'B' be a rigid designator of his body. Then if Descartes were indeed identical to B, the supposed identity, being an identity between two rigid designators would be necessary, and Descartes could not exist without B and B could not exist without Descartes. The case is not at all comparable to the alleged analogue, the identity of the first Postmaster General with the inventor of bifocals. True, this identity obtains despite the fact that there could have been a first Postmaster General even though bifocals had never been invented. The reason is that 'the inventor of bifocals' is not a rigid designator; a world in which no one invented bifocals is not ipso facto a world in which Franklin did not exist. The alleged analogy therefore collapses; a philosopher who wishes to refute the Cartesian conclusion must refute the Cartesian premise, and the latter task is not trivial.

Here Kripke presents the following argument:

1. If 'a' and 'b' are rigid designators, then statements of the form 'a = b' are, if true, necessarily true.
2. Let 'Descartes' be a rigid designator of Descartes.
3. Let 'B' be a rigid designator of Descartes' body.
4. Hence, if "Descartes = B" is true, it is necessarily true. (1,2, and 3)
5. (x)(y)(If x can exist at some time t and y does not exist at t, then "x = y" is not necessarily true).
6. It is possible for Descartes' body to exist at some time t even though Descartes does not exist at t.
7. Therefore, "Descartes = B" is not necessarily true. (5,6)
8. Therefore, "Descartes = B" is not true. (4,7)

The above argument is, as Kripke acknowledges, a contemporary formulation of a "conceivability-type" argument for dualism which originates with Descartes. In a famous Sixth Meditation passage, Descartes argues from his ability to form a clear and distinct



conception of himself or his mind as distinct from his body to the conclusion that he or his mind really is distinct from his body.<sup>13</sup> Here Kripke simply argues that he can exist without his body, and that his body can exist without him. While this type of argument has been variously criticized,<sup>14</sup> it does represent one of the strongest possible arguments for dualism; hence, it is worth considering in some details. Before I do so, however, I will look at a similar argument, formulated in terms of mental events.

### 2.3.2 Event Identity

The identity thesis is also formulated as a theory about the relation between types of physical events and types of mental events, as well as a theory about the relation between particular mental events and particular physical events. The identity thesis asserts that this relation is one of identity. Kripke formulates an argument against the identity thesis, conceived as a claim about the relation between particular mental and physical events:<sup>15</sup>

Let 'A' name a particular pain sensation, and let 'B' name the corresponding brain state, or the brain state some identity theorist wishes to identify with A. Prima facie, it would seem that it is at least logically possible that B should have existed (Jones's brain could have been in exactly that state at the time in question) without Jones feeling any pain at all, and thus without the presence of A. Once again, the identity theorist cannot admit the possibility cheerfully and proceed from there; consistency, and the principle of identities using rigid designators, disallows any such course. If A and B were identical, the identity would have to be necessary.

The above argument can be formulated as follows:

1. If 'a' and 'b' are both rigid designators, then any statement of the form 'a = b', if true, is necessarily true.
2. Let 'A' be a rigid designator for a particular pain sensation.
3. Let 'B' be a rigid designator for a corresponding brain state.
4. Hence, "A = B", if true, is necessarily true. (1, 2, and 3)
5. (x)(y)(If it is possible for x to exist at some time t and y does not exist at that time, then "x = y" is not necessarily true).
6. It is possible for B to exist at some time t even though A does not exist at t.
7. Therefore, "A = B" is not necessarily true. (5, 6)
8. Therefore, "A = B" is not true. (4, 7)

Because both of the above arguments will come into clearer focus when I consider criticisms of them, I now consider some criticisms of these two arguments which have been raised in the literature.

#### 2.4 Feldman's Criticisms

In this section, I consider several criticisms of Kripke by Fred Feldman. Feldman is writing from a materialist perspective, and his criticisms primarily focus on premise 6 in the above two arguments.

#### 2.4.1 Feldman's Criticism of Kripke's Argument that Descartes is not Identical to His Body

Feldman formalizes two versions of Kripke's argument that a person is distinct from his body (the first of the two arguments just presented), and refers to them as "Argument A" and "Argument B". He then goes on to comment:<sup>16</sup>

Arguments A and B are valid. Furthermore, in each case the conclusion is inconsistent with person-body materialism. Nevertheless, neither argument is very persuasive: in each case anyone predisposed to believe in person-body materialism would be right to claim that the argument either begs the question or else is based on a failure to understand what person-body materialism amounts to.

Person-body materialism is the doctrine that people are identical to their bodies. In other words, anything that is a person is identical to the body of that person:

7.  $(x)(t)(Pxt \supset (\exists y)(Byt \ \& \ Oyxt \ \& \ x = y))$

Any clear-headed proponent of this view would undoubtedly say that death usually does not make people cease to exist. Since, on this view, people are their bodies, the conditions of survival of any person must be the same as those of his or her body. This is not to say that anyone adopting materialism would be saddled with the view that there must be some sort of life after death. Far from it; this form of materialism doesn't even imply that people continue being people after they die. But what's important here is that, on this view, the property of being alive is an accidental property of the things that have it. Something can have it for a while, and then cease to have it, without thereby ceasing to exist. Thus, the things that are people can continue to exist after death.

Here Feldman is attempting to refute premises in his arguments A and B which are equivalent to premise 6 in the first argument above:

6. It is possible for Descartes' body to exist at some time  $t$  even though Descartes does not exist at  $t$ .

It is curious to see an advocate of materialism advocate a view

which many have argued is unintelligible, and which is usually associated with the strong form of dualism: the view that a person can survive his own death! I now consider the doctrine motivating this view: that a person is identical to his body.

The first point to make about Feldman's comments here is that he implicitly appeals to a distinction he does not explain: a distinction between being alive and existing. Being alive is the mode of existence which characterizes a certain type of thing, namely, an organism. Of course, some things can exist without being alive (e.g. tables and chairs), but if Feldman wishes to claim that I can exist without being alive, then he should provide an account of this distinction.

Secondly, Feldman claims that being alive is not an essential property of the things that have it. If Feldman wishes to make such a claim, then he should make it clear what his view of essential properties is. Kripke gives the following criterion for essential properties:<sup>17</sup>

If we can't imagine a possible world in which Nixon doesn't have a certain property, then it's a necessary condition of someone being Nixon. Or a necessary property of Nixon that he [has] that property.

Here Kripke advocates an "imaginability" criterion for essential properties: if we can't imagine something without a property, then that property is essential to it (here I am interpreting 'necessary condition' to refer to essential properties). Kripke goes on to say:<sup>18</sup>

For example, supposing Nixon is in fact a human being, we might not imagine that there could have been a possible world in which he was, say, an inanimate object; perhaps it is not even possible for him not to have been a human being. Then it will be a necessary fact about Nixon that in all possible worlds where he exists at all, he is human or anyway he is not an inanimate object.

Here Kripke explicitly rejects the possibility that a person could be an inanimate object (e.g. a corpse). I think he would, on the same basis, also reject the possibility that a person could exist without being alive. Hence, if Feldman wishes to take this view about essential properties, he owes us an explanation of what his view is.

Of course, some dualists do believe that a person can survive death. However, such persons often identify the person with an immaterial substance or soul which can survive the dissolution of the body. It is not clear whether such theorists would regard being alive as an essential or accidental property of a person; but clearly Feldman cannot appeal to this type of view to support his argument since this is the type of view he is arguing against.

Third, there are strong reasons for rejecting the identification of a person with his or her body. One reason is that we often attribute to the body functions we do not attribute to the self, or person. For example, my body circulates the blood, digests food, and fights infection. These are all things my body does on its own, not things I do. The fact that we distinguish in this way things our body does from things we do is a strong reason for not identifying a person with his body.

Likewise, we attribute properties to ourselves which we don't attribute to our bodies. For example, a person can be sick even though there is nothing wrong with his body. I can be tired even though my body is not, and I can be a cold person even though my body is normal temperature, or is feverish. These are all examples of properties we attribute to ourselves and not our bodies, and provide an additional reason for not identifying a person with his body.

Finally, if I am identical to my body, then my body is the subject of all my psychological states. Thus, even if we don't ordinarily say this, it should make sense to say "My body is angry" instead of "I am angry", or "My body is bored" instead of "I am bored". But these substitutions yield statements with no clear meaning. And, in some cases, this would result in quite different meanings; for example to say "You are boring" is quite different from "Your body is boring". And desiring another person's body is quite different from desiring the other person himself or herself.

Of course, it might be that a materialist could satisfactorily deal with all the above objections. But the general point here is that Feldman is no better off than he claims Kripke is, since all he has done is to state some very controversial consequences of materialist doctrine in reply to Kripke. Feldman claims that Kripke's argument begs the question, but it seems clear that his own comments, by making use of the unexplained distinction between being alive and existing, and by appealing to an unexplained notion



of "essential property", are as question-begging as Kripke's. Furthermore, there are strong reasons for rejecting Feldman's identification of person and body, which Feldman does not take into account. Hence, I conclude that these comments by Feldman fail to provide any reason for rejecting Kripke's argument.

Of course, it might be that a materialist could satisfactorily deal with all the above objections. But the general point here is that Feldman is no better off than he claims Kripke is, since all he has done is to state some very controversial consequences of materialist doctrine in reply to Kripke. Feldman claims that Kripke's argument begs the question, but it seems clear that his own comments, by making use of the unexplained distinction between being alive and existing, and by appealing to an unexplained notion of "essential property", are as question-begging as Kripke's. Furthermore, there are strong reasons for rejecting Feldman's identification of person and body, which Feldman does not take into account. Hence, I conclude that these comments by Feldman fail to provide any reason for rejecting Kripke's argument.

Feldman does say:<sup>19</sup>

The problem with Kripke's arguments, then, is that in each case the argument seems to presuppose that at death the person literally ceases to exist, whether the body does or not.

Feldman is right if he means that Kripke's argument does imply that the person ceases to exist after death, although I don't believe that Feldman has given us a single good reason to doubt this claim. However, I defer for now the question whether Kripke needs to argue

for this claim, or whether he has succeeded in establishing it. For now, I turn to Feldman's second line of reasoning.

Perhaps Feldman senses the implausibility of the position he ascribes to "any clear-headed proponent of [materialism]", for he now shifts ground as follows:<sup>20</sup>

Perhaps Kripke or someone wishing to defend these arguments would say that the property of being alive is an essential property of the things that have it. We can understand this to mean that everything that is alive is such that, necessarily, for any time  $t$ , if it exists at  $t$ , then it is alive at  $t$ :

(8)  $(x)(t)(Axt \supset \Box(t')(Ext' \supset Axt'))$

From this it would follow that, when Descartes dies, he literally ceased to exist. This might be thought to provide support for (8).

Feldman goes on to object:

Surely, Descartes's body was alive when Descartes was. If being alive is essential to the things that have it, then, just as death marked the end of Descartes, so must it have marked the end of his body as well. On this view the corpse must be construed as a new entity, brought into existence by "substantial change". Hence, far from providing support for it, the claim that everything is alive is essentially so actually lends overwhelming support to the denial of [premise 6].

The Kripkean thus has to say, if he wants to adopt this line of defense for [premise 6], that being alive is not an essential property of his body, even though it is a property his body has. Hence, the body can survive death, while Descartes cannot.

Feldman is right in saying that someone who wishes to defend a dualistic position might argue from the fact that a person is essentially alive and the body is not to the distinctness of person and body. But Feldman claims that such a defense makes Kripke's argument superfluous. He then shows that one can also construct an

argument to the conclusion that a person is distinct from his body by making use of Leibniz's Law:<sup>21</sup>

- 9. d has the property of being essentially alive.
- 10. b does not have the property of being essentially alive.
- 11.  $x = y$  iff  $(F)(Fx$  iff  $Fy)$ .
- 12.  $d \neq b$ .

Feldman is right that premises 9 and 10 can be used either in defense of premise 6 or as premises in an argument employing Leibniz's Law. The difference is that, in Kripke's argument, statements 9 and 10 are used in support of a premise which is conjoined with other statements in such a way as to prove the desired conclusion independent of Leibniz's Law. This does not render Kripke's argument superfluous; it simply means that there may be more than one way to prove that a person is distinct from his body.

Feldman also claims that "no argument has been given for 9"<sup>22</sup> (the statement that d has the property of being essentially alive). While Feldman may be right that Kripke has not given an argument for this claim, we saw above that Feldman has not given a single good reason to reject this claim. And no serious dualist would accept Feldman's "intuition" that "Surely, Descartes' body was alive"; he would say that Descartes' body was a living thing, but this is not to say that his body was alive. Hence, Feldman's comments up to this point represent little more than philosophical name-calling.

#### 2.4.2 Feldman's Criticisms of Kripke on Event Identity

Feldman also considers Kripke's argument against the identity thesis formulated in terms of the identity of mental events and brain events. Feldman quotes the following passage from Naming and Necessity:<sup>23</sup>

Prima facie, it would seem that it is at least logically possible that b should have existed (Jones's brain could have been in exactly that state at the time in question) without Jones feeling any pain at all, and thus without the presence of [a]. Once again, the identity theorist cannot admit the possibility cheerfully and proceed from there; consistency, and the principle of the necessity of [true] identity [i.e. using rigid designators] disallows any such course.

Feldman formulates Kripke's argument, which he calls argument H, as follows. Here 'Oxt' means "x occurs at t" and 'Pxt' means "x is a pain-event at t":

$$(28) \Diamond(\exists t)(Oat \ \& \ \sim(\exists x)(Pxt \ \& \ Oxt))$$

$$(29) (t)(Oat \supset Pat)$$

$$(30) \Diamond(\exists t)(Oat \ \& \ \sim Oat) \qquad (28, 29)$$

$$(31) \Diamond(\exists t)(Oxt \ \& \ \sim Oyt) \supset \Diamond(x \neq y)$$

$$(32) \Diamond(a \neq b) \qquad (30, 31)$$

Feldman objects that "(28) and (29) do not entail (30)". He continues:<sup>24</sup>

It does not follow, from the fact that there is a possible world in which b occurs without anything that is there a pain-event, and the fact that a is a pain-event in this world, that there is a world in which b occurs without a. For if a, which is a pain-event in this world, occurs in other worlds without being a pain-event, then perhaps a does occur wherever b does. So the argument turns on the question whether a, which is a pain-event in the actual

world, is a pain-event in every world in which it occurs. In other words, the argument turns on the question whether a is essentially a pain-event.

Feldman is correct in saying that Kripke's claim that a pain-event is essentially painful is crucial to his argument. He goes on to say that:<sup>25</sup>

Any serious materialist should recognize that his view entails that painfulness is never part of the essence of a pain-event. Pain-events are experienced as they are only as a result of contingent laws of nature...Certain brain-events are such that it is contingent that they are felt as pains.

He finally concludes that Kripke's argument "turns on an undefended, controversial premise that materialists have, and should have, rejected".<sup>26</sup>

Kripke's argument, according to Feldman, depends on the unsupported intuition that being a pain is an essential property of a feeling of pain. Feldman does not refute this claim; he contents himself, once again, with pointing out that this intuition is inconsistent with a materialist point of view. However, I believe that there is a serious difficulty in Kripke's argument. Can Kripke's argument be stated in a way which does not seem simply to appeal to our intuitions?

Kripke apparently believes that an appeal to intuitions is one of the most forceful ways of presenting a philosophical argument. However, I believe that, with respect to the mind-body problem, one's intuitions are influenced by his philosophical point of view. Hence it may be that, if the truth about the mind-body problem rests, in the final analysis, on intuition, that it cannot be solved.

Furthermore, I see some merit in Feldman's objection that it is not an essential property of a feeling of pain that it is painful. Pains certainly vary in intensity, and one characteristic of pain is that it often involves an "overload" of our nervous system. Thus, we can put pressure on a certain part of our body, and slowly increase the pressure until it becomes painful. In a situation like this, I don't think it is implausible to argue that we have the same feeling all along, but that it suddenly becomes painful. First, it might be a feeling of pressure on our hand, but as the pressure increases, the feeling becomes painful. On this analysis, the very same feeling at first is not painful, but then, as the pressure is increased, that very feeling becomes painful. Hence, a single feeling can at one point not be painful, and then become a feeling of pain. If this is right, being painful is merely a contingent property of some feelings, which a feeling might have at one time  $t$  and not have at another time  $t'$ . On this view, then, painfulness is a merely contingent property of feelings.

I believe, however, that Kripke's argument can be stated in a way that avoids this difficulty. For, even if pain is a merely contingent property of some feelings, a feeling which is at one time painful and another time is not is still a feeling of some kind. Hence, even if painfulness is not an essential property of feelings, at least being a feeling is an essential property, and Kripke's argument can be stated as before, except that instead of "pain" being an essential property, "being a feeling" is an essential



property. So even if we grant Feldman's claim that painfulness is a contingent property of some feelings, we can formulate Kripke's argument, substituting "being a feeling" for "painfulness" as the essential property.

Feldman grants that, if we change 29 to 29':

$$(29') \Box(t)(Oat \supset Pat)$$

that Kripke's argument is valid.<sup>27</sup> Let us reformulate Kripke's argument to accomodate the above point, substituting (Fa) ("a is a feeling") for (Pa) ("a is a pain-event"):

$$(28) \Diamond(\exists t)(Oat \ \& \ \sim(\exists x)(Fxt \ \& \ Oxt))$$

$$(29) \Box(t)(Oat \supset Fat)$$

$$(30) \Diamond(\exists t)(Oat \ \& \ \sim Oat) \quad (28), (29)$$

$$(31) \Diamond(\exists t)(Oxt \ \& \ \sim Oyt) \supset \Diamond(x \neq y)$$

$$(24) \Diamond(a \neq b) \quad (30), (31)$$

Perhaps Feldman would not be convinced by this reformulated argument. His general strategy seems to be, to argue for a candidate essential property  $p$  of  $x$ , that  $p$  is really an accidental property of  $x$ , since what appears to be  $x$  is really fundamentally  $y$ , and  $p$  is an accidental property of  $y$ . Thus, he argued that a person is only accidentally alive, and that someone might continue to exist after death, event though he or it was no longer a person. Feldman might be expected to reply there that a feeling is just a kind of event, and that some events are and others are not feelings. Furthermore, he might say, the very same event that is a feeling in one world is not a feeling in another world; hence, being a feeling

is a merely contingent property of some events. This argument would parallel his argument on pain.

## 2.5 Feldman's Three Conceptions of Events

Feldman does make an argument similar to the one just sketched elsewhere. In a recent paper, he distinguishes three conceptions of what an event is. While he concedes that, on the first two views he considers, the "propositional" view and the "structural" view, the identity thesis does not seem plausible, yet, he maintains, on the "concrete" events view, Kripke's argument does not succeed. Feldman tells us very little about concrete events, except to say that events are more "coarse-grained" on this conception than they are on other conceptions. The idea seems to be to treat each event as a concrete individual.

After introducing this concept, Feldman argues as follows:<sup>28</sup>

If a and b are concrete events, then, so far as I can see, Kripke's point can not be established. Event a is a case of someone being amused. But why should it be essentially so? It is as concrete an individual as Jones himself. Surely no one would say that Jones is essentially someone who is amused. Just as being amused is a contingent property of Jones, so being a case of someone being amused may be a contingent property of a.

If we are to have a coherent version of the contingent psycho-physical event identity theory, then it seems most reasonable to assume that the events in question are concrete events. And if we do this, then Kripke's argument loses its foothold. He has no way to show that a is essentially a case of someone being amused. For all we know, that very event could have occurred without anyone being amused. Hence, the fact that we can imagine b

occurring without anyone being amused does not substantiate the claim that we can imagine b occurring without a.

If we substitute "Jones' having a feeling" for "Jones' being amused", we get the claim that the very event of Jones' having a feeling could occur without anyone having a feeling. But surely this is not the case. It is easy to see how Jones can exist unamused; it is not easy to see how the event of Jones' having a feeling could exist if no one had a feeling. It might be somewhat plausible to argue that being amused is a merely contingent property of certain feelings; it is not at all plausible to argue that being a feeling is a merely contingent property of certain events.

In order to see why this is so, consider what a feeling is on the concrete events view. Of the structural view, Feldman says:<sup>29</sup>

According to the structural view, events are structures of property, individual, and time. They are individuated by reference to their components. We have the same event if and only if we have the same property, individual and time.

Concrete events are apparently like structural events, except that they are more "coarse-grained". I interpret this to mean that the criteria for what counts as the same event are less rigid. For example, Feldman says that the very same concrete event could make both the following statements true:

1. Jones is very amused.
2. Jones is amused about what happened to the chickens.

Concrete events still appear to be a structure of an individual, a property, and a time; the difference seems to be that the criteria

for "same property" (and perhaps also for "same individual" and "same time") are less rigid (looser).

Using this conception of an event, it is easy to see how Feldman might become convinced that being amused is a merely contingent property of certain events which consists in a person's having a feeling at some time  $t$ . For if the same event can make "being very amused" and "being amused about what happened to the chickens" true, then perhaps the very same event could make something similar true like "being entertained by the story about the chickens", which perhaps does not imply being amused, but is nonetheless similar. In such a case, it might seem plausible to argue that Jones' being amused is a contingent property of the event which makes both "Jones is entertained" and "Jones is amused" true. Hence, there is a world in which this event occurs without anyone's being amused.

However, if the property involved is not "being amused" (having a particular feeling) but "having a feeling", I do not see how Feldman's argument can be made. For Feldman has to argue that the very same event which constitutes Jones' having a feeling can occur without anyone's having a feeling. Perhaps Feldman would argue that Jones' having a feeling is an accidental property of some event more generally described, such as Jones' having a perception. But feeling, if it is a form of perception at all, is a very different form of perception from other forms of perception such as seeing and hearing, and no plausible criterion of event identity would be so

broad that seeing a chair makes "I have a feeling" true (unless 'feeling' has been given a greatly expanded sense). The problem with Feldman's argument is that it must be possible for the very same event which consists in Jones' having a feeling at  $t$  to occur even though Jones has no feeling at all. Perhaps he sees a chair at  $t$  instead. But Jones' seeing a chair is a different event altogether from Jones' having a feeling; if no one has a feeling at  $t$ , then certainly Jones doesn't either, and the event of Jones having a feeling doesn't occur. Hence, this is not an example of a possible world in which the very same event occurs but Jones does not have a feeling.

Feldman is perhaps right that for any candidate essential property of some individual  $a$ , it is possible to redescribe  $a$  in a more general way to that  $e$  appears to be a merely accidental property of  $b$ , the newly described  $a$ , and then claim that  $b$  is really  $a$ . It is this type of reasoning according to which being alive and even being a person is a merely contingent property of someone (or something), who might continue to exist after death even though he (or it) is no longer alive or a person. (Of course, Feldman does not say what I really am if I am not a person). While this type of reasoning might work at a fairly specific level of description, it does not continue to work indefinitely. For eventually, the redescription becomes so general that it is no longer plausible to maintain the identity between the old individual,  $a$ , and the newly described individual,  $b$ . Thus, in the

case of having a feeling, it is implausible to maintain that having a feeling is really just an accidental property of a more generally described event such as having a perception. For even if feeling is a form of perception, it is a very specific type of perception, and seeing a chair (a different form of perception) could not possibly count as the very same event as having a feeling of pain on any reasonable criterion of event identity.

I conclude, then, that Feldman's "de-essentialization by redescription" does not work in every instance, and hence that Kripke's premise 2, if construed to read "having a feeling" instead of "having a feeling of pain", is not susceptible to Feldman's refutation.

Furthermore, Feldman states that Kripke's argument rests on an undefended intuition about essential properties. One might just as well point out that Feldman's argument rests on a poorly defined and unclear conception of events, and that, if he is going to be successful in refuting Kripke's argument, he owes us a better explanation of his conception of "concrete events".

## 2.6 Other Issues Relating to Dualism

Even if I have shown that Feldman has failed to refute Kripke's premise 5, of course, I have not demonstrated that either argument is sound. My defense of premise 5 constitutes indirect support for it. However, premise 2 is also important: the claim that certain



terms which designate mental states are rigid designators. This question seems to turn on the question whether being a pain is an essential property of a feeling of pain. While we saw that this may not be the case, yet being a feeling does seem to be an essential property, so the argument works with 'feeling' substituted for 'pain'.

#### 2.6.1 The Problem of Interaction

Even if Kripke's argument B is sound, however, it at most constitutes an argument for some form of property dualism, a dualism of events, not the "full-blooded" substantial dualism which Descartes advocated. However, even if an argument for a theory cannot be refuted, if the theory has insuperable difficulties elsewhere, then the best possible argument will not get us very far. It is believed by many people that interaction is such an insuperable problem for dualism, so that even if one can give a sound and apparently non-trivial argument for dualism, yet the problem of interaction has sometimes frightened away even the most loyal of defenders.

There are several different arguments which purport to show that interaction between mind and body cannot be explained on a dualist account. One is the argument from conservation of energy. Michael Levin states one formulation of this argument in the following passage:<sup>30</sup>

Another traditional argument...does seem to show decisively that the causal relations between the mind as the dualist conceives it and the body are impossible. Originally my leg, which has mass  $w$ , is motionless. Then it moves distance  $d$ . Enough work was done to move a mass  $w$  a distance  $d$ . Where did the energy come from? It is an empirical fact, the law of the conservation of energy, that the energy had to come from somewhere, and not only does it come from some preceding physical event, no immaterial substance could possibly supply mechanical energy. The mind can apply no physical force to the leg, while the contraction of the hamstring muscle supplies just the right amount - and the bodily event just is the motion of an object of mass  $w$  through distance  $d$ . Thus the contraction of the muscle is the cause of  $e$ . If there were non-physical causes of physical events, the energy in the universe would increase.

The main point is that an immaterial object cannot impart motion to a material object. If a thought is an immaterial object with no material properties, then it seems impossible to conceive that a thought could impart physical motion to my arm, or to any physical body. Of course, interaction seems to be an undeniable fact of life. A thought of embarrassment can make my face turn red, while my poor physical condition can cause me to become depressed. Yet, it seems, this interaction cannot be accounted for on dualist principles.

Why is interaction seemingly unsolvable within a dualist framework? The chief reason for this is that dualists and materialists both accept a common conception of what mind or a mental event is and what a body or physical event is. They then proceed to debate whether, based on these conceptions, mind and body are identical or distinct. Often, these conceptions are Cartesian in nature: body is extended, material, and exists in space while

mind is unextended, immaterial, and does not exist in space. Given these assumptions, it is no wonder that the mind-body problem seems unsolvable, or that interaction presents a problem for the dualist. For, by accepting these conceptions of mind and body, the dualist is forced to argue that something which essentially has a property (e.g. a material object), can interact with something which essentially lacks that property (e.g. an immaterial object). This might be possible if either object had other properties through which they might interact, but, unfortunately, mental and physical events are often conceived as having only one essential property (in fact, that this is so was a cardinal doctrine of Descartes). This conception of mind and body makes it possible to use Leibniz's Law to quickly demonstrate distinctness of mind and body. However, the price of having such a conception of mind and body is that interaction is difficult if not impossible to account for.

Of course, a dualist is simply committed to the distinctness of mind and body, or of mental events and physical events. It is not inconsistent with the principles of dualism to draw the mind-body or mental-physical distinction in such a way that it marks out a dichotomy. Because I believe that it is fruitless to try to solve the problem of interaction until the mental-physical distinction itself is examined more closely, I defer further discussion of this issue until Chapter 6, after I discuss the mental-physical distinction in Chapter 5.

## 2.7 One Consequence of Kripke's Views on Origin

Unfortunately, Kripke does not, to my knowledge, comment on the problem of interaction. Furthermore, he does not say how he thinks the distinction between mental events and physical events should be drawn. However, it is worth pointing out one apparent consequence of his theory that the origin of something is an essential property of that thing. It seems fairly clear that some mental states originate in the body. For example, a headache might originate in a pinched nerve, and a toothache originates in a tooth. Furthermore, it seems that we distinguish between mental states according to their different physical origins. Thus, a headache which originates in a different pinched nerve is counted a different headache, even if it is roughly in the same phenomenal location as the first. Likewise, toothaches that originate in different teeth are different toothaches. If Kripke is right that the origin of something is one of its essential properties, then it may turn out that some mental states have at least one physical property essentially, viz. that they originate in some particular part of the body. Furthermore, this consequence seems consistent with the way we do in fact individuate mental events such as feelings. If this is correct, then Kripke may be committed to a more complicated view than straight-forward dualism.

The relation between a feeling and the bodily state in which it originates is typically described as being that of cause and

effect. For example, it might be said that a pinched nerve causes my headache, instead of saying that my headache originates in a pinched nerve. It is not clear that these claims are consistent, although 'origin' and 'cause' are not synonymous. One difference is that the origin of x is sometimes the very first of a chain of events which ultimately lead to x, while the cause of something usually immediately precedes x. Thus, questions about the origin of man sometimes focus on a "missing link" which would bridge the gulf between man and beast; if such a missing link were discovered, it would not be described as the cause of mankind, although it would be plausible to claim that mankind originated in this creature. From a Biblical point of view, of course, human life originated in the Garden of Eden.

If it is objected that the above consequence of Kripke's views about origin depend on an undefended and controversial distinction between the physical cause and the physical origin which Kripke might not accept, I reply that Kripke does explicitly argue that it is an essential property of a person that he came from the sperm and egg that he in fact came from. And a sperm and egg would ordinarily be taken to be physical things (at least they have physical properties: whether they are "purely" physical depends on how one analyzes the fact that they are living things). Hence, Kripke is certainly committed to the claim that a person has at least one essential property which is physical: viz. that he came into being in virtue of the union of two particular physical things. i.e. a

certain sperm and a certain egg. Furthermore, Kripke seems to accept a general similarity between the types of arguments which purport to show person-body distinctness and mental-physical distinctness. Hence, I believe he would accept this consequence for mental states as well (that they essentially originate in some particular bodily state).

Once this consequence has been pointed out, it begins to appear that "mental" and "physical" are more intimately intertwined than the formulations of the mind-body problem would lead us to believe. To use another example of this, let us consider premise 3 in Kripke's argument for dualism:

3. Let 'B' be a rigid designator of Descartes' body or of some bodily state.

The problem here is that, in an argument purporting to show the distinctness of Descartes from his body, Descartes' body is identified by reference to Descartes. While it might be thought that Descartes' body could be identified in some other way (e.g. ostensively), the problem is even greater when one considers the problem of saying which bodily state one wishes to designate by 'B'. If one says "the bodily state associated with some feeling of pain", then he has relied on the feeling of pain to individuate the bodily state. But this is to assume their distinctness. For if  $x$  is identical to  $y$ , then  $y$  cannot be the individuating principle of  $x$ . In other words, nothing can be its own individuating principle.

This problem is usually side-stepped in favor of some description such as "C-fiber firings". However, it is a generally



accepted fact that it is not possible to find a physiological state which correlates strictly with pain. Hence, the description of "C-fiber firings" to pick out the bodily state is based on a misconception. Furthermore, even if C-fiber firings did correlate with pain, it would be necessary to specify which C-fiber firings one wanted to pick out as being correlated with some particular feeling of pain. And one would also have to specify what other parts or states of the body he wished to include in the reference of 'bodily state B'. The only way to distinguish among bodily states is by picking out those that contribute in some way to the feeling of pain. But, once again, this is to use the mental state whose distinctness from the bodily state is being proved as a principle of individuation for the bodily state. Of course, one might say "the total bodily state", but then the issue would be formulated as being whether a particular feeling of pain is identical to or distinct from the entire bodily state of some body. But no clear-minded materialist would want to identify a feeling of pain with a total bodily state - for then if any other mental events were occurring simultaneously, there would be nothing to identity them with. And now one must still specify the state of which body he wishes to identify with pain. But this is the same type of circularity we were trying to avoid!

The problem of identifying some bodily state which "correlates" with pain may appear to be a problem for Kripke in stating premise 3, but it's really more a problem for the identity theorist. For

Kripke's argument can apply to any bodily state the identity theorist wishes to choose, while the latter must specify some bodily state to identify with any feeling which might be considered. This creates a serious problem for the identity theorist because there is no non-circular way of specifying which brain state is to be identified with (e.g.) pain. Either he uses some mental criterion (e.g., "the bodily state correlated with pain"), or he doesn't. If he uses a mental criterion, then he has already conceded the argument to dualism, for the individuating principle must be distinct from x, otherwise it could not serve as a rigid designator. On the other hand, if he doesn't rely on a mental criterion of some kind, he will never succeed in picking out the right brain state, for then the materialist must devise some purely physical criterion for picking out the brain state. Because there is no bodily state that correlates strictly with pain, I do not see how this can be done.

## FOOTNOTES

1. Saul Kripke, Naming and Necessity, in Semantics of Natural Languages, Gilbert Harman and Donald Davidson (eds.) (Dordrecht: Reidel, 1971). pp. 253 - 355.
2. Fred Feldman. "Kripke's Argument Against Materialism", Philosophical Studies, 24 (1973), pp. 416-19 and "Kripke on the Identity Theory", The Journal of Philosophy, November, 1974, pp. 665-6.
3. See, for example, Gottlieb Frege, "On Sense and Reference", in Translations from the Philosophical Writings of Gottlob Frege, Peter Geach and Max Black (eds.), (Oxford: Basil Blackwell, 1952), pp. 56-78.
4. Bertrand Russell, "On Denoting", in R. C. Marsh, Logic and Knowledge: Essays, 1901 - 1950 (London: G. Allen & Unwin, 1956) pp. 39-56.
5. John Stuart Mill, A System of Logic Ratiocinative and Inductive: Being a Connected View of the Principles of Evidence and the Methods of Scientific Investigation (London: Longman, 1970)
6. Kripke, p. 270.
7. Ibid., p. 322.
8. Ibid., p. 319.
9. Ibid., p. 304.
10. Ibid., p. 314.
11. Ibid.
12. Ibid, pp. 334-5.
13. Rene Descartes, The Philosophical Works of Descartes, trans. Elizabeth S. Haldane and G. R. T. Ross, Vol. I (New York, Cambridge University Press, 1972), p. 190.
14. See, for example, Margaret Wilson, Descartes (London: Routledge and Kegan Paul: 1978), pp. 185-201.
15. Kripke, p. 335.
16. Feldman (1974), p. 667.

17. Kripke, p. 268.
18. Ibid.
19. Feldman (1974), p. 667.
20. Ibid., pp. 667-8.
21. Refer to Feldman (1974), p. 668.
22. Ibid.
23. Feldman (1974), p. 673 from Saul Kripke, Naming and Necessity, p. 335.
24. Feldman (1974), pp. 673-4.
25. Ibid., p. 675.
26. Ibid., p. 676.
27. Ibid., p. 674.
28. Feldman, "APA Paper on Kripke"; this quotation is taken from an unpublished copy of the article, p. 13.
29. Ibid., p. 12.
30. Michael Levin, Metaphysics and the Mind-Body Problem (Clarendon Press, Oxford: 1979), pp. 84-5.

## C H A P T E R   I I I

### DONALD DAVIDSON'S ANOMALOUS MONISM

#### 3.1 Introduction

Some philosophers have rejected both dualism and materialism in favor of views which try to find a middle ground between these two positions. Donald Davidson is an example of such a philosopher. While Davidson is well-known for his work on action theory and on causation, he has also published several articles which relate directly to the mind-body issue. In this chapter, I discuss these articles. While I mainly discuss "Mental Events", I also refer to "Philosophy as Psychology" and "The Material Mind".<sup>1</sup>

Davidson's primary purpose in "Mental Events" is to show how three principles he believes are true can be rendered consistent with one another. In section 3.2, I state these three principles. Davidson believes that these principles can be reconciled by adopting what he calls anomalous monism. This view can roughly be stated as saying that no mental descriptions instantiate a strict, deterministic scientific law. I formulate this principle of anomalous monism in section 3.2 and show how it resolves the apparent inconsistency.

Most of this chapter is devoted to examining the merits of anomalous monism as Davidson formulates it. In section 3.3, I formulate what I take to be Davidson's main argument for anomalous monism. This argument has as its conclusion the claim that there are no psychophysical laws; from this conclusion one can derive anomalous monism with only a few additional assumptions.

One factor on which the plausibility of Davidson's defense of anomalous monism depends is the correctness of and way in which he draws the distinction between mental and physical events. I discuss this issue in section 3.4. I argue that Davidson's conception of mental events is too restricted and that his criterion that an event is mental just in case it falls under a mental description leads to unacceptable results.

Davidson's argument for anomalous monism also depends for its soundness on the claim that any scientific law must draw on a homonomic domain. This claim, together with Davidson's claim that the physical conceptual domain is homonomic, has important consequences for Davidson's theory and for scientific law. I examine this claim and these consequences in section 3.5. One consequence is that the physical conceptual domain includes only the languages of chemistry and physics and does not include the languages of biology and neurology which are heteronomic. It also means that Davidson must identify mental events with molecular events rather than with brain events. I discuss these and other consequences in this section.



In section 3.6 I place Davidson's claim that a scientific law must draw on a homonomic domain within the context of his argument that there are no psychophysical laws. The claim that there are no psychophysical laws is the main result he needs to justify anomalous monism. I conclude that since Davidson does not argue for the claim that scientific laws must draw on a homonomic domain, and since there are reasons for doubting that this is so, that he has failed to establish anomalous monism and that there are good reasons for doubting that anomalous monism is true. Furthermore, since it is by adopting the principle of anomalous monism that Davidson reconciles the apparent inconsistency in the three principles he believes to be true, I conclude that Davidson has failed to establish that his way of reconciling these principles is correct. This is not, however, to deny that he has provided a way of reading these principles which renders them logically consistent with each other.

Davidson also gives a demonstration of the identity of mental and physical events. In section 3.7 I formulate this demonstration as an explicit argument. I conclude that this demonstration fails because it depends on two claims which Davidson has not established and which seem very dubious. The first claim is the one discussed in section 3.4 which states that an event is physical just in case it has a physical description. The second claim is the principle of anomalous monism, which is discussed in section 3.5.

In section 3.8 I briefly summarize the chapter. I conclude that Davidson has failed to establish either the "anomalous" part or

the "monism" part of anomalous monism. Furthermore, there are good reasons for doubting anomalous monism.

### 3.2 Davidson's Views in "Mental Events"

In "Mental Events", Davidson formulates three principles which he believes to be true but which seem to be inconsistent with each other.<sup>2</sup> These principles are as follows:

- I. At least some mental events interact causally with physical events (the Principle of Causal Interaction).
- II. Where there is causality, there must be a law: events related as cause and effect fall under strict deterministic laws. (the Principle of the Nomological Character of Causality).
- III. There are no strict deterministic laws on the basis of which mental events can be predicted and explained. (the Anomalism of the Mental).

Although the inconsistency is not formal and explicit in the above three statements, it can be made explicit as follows:

1. There exists at least one mental event  $m$  and physical event  $p$  such that  $m$  causes  $p$ , and there exists at least one physical event  $p'$  and mental event  $m'$  such that  $p'$  causes  $m'$ .
2. For any events  $x$  and  $y$ , if  $x$  causes  $y$  or  $y$  causes  $x$ , then there exists some strict deterministic law  $L$  such that  $x$  and  $y$  fall under  $L$ .
3. There exists no strict deterministic law  $L$  such that any mental event falls under  $L$ .
4. Therefore, there exists some strict deterministic law  $L'$  such that mental event  $m$  and physical event  $p$  fall under  $L'$ . (1,2)

Statement 4, which contradicts statement 3, follows from statements 1 and 2.

Davidson attempts to resolve the apparent inconsistency in the above three principles by revising II, the principle of the Nomological Character of Causality, and III, the Anomalism of the Mental. Fundamental to his solution is his claim that events are mental only as described. Davidson believes that a single event can have both a physical description and a mental description, and that whether an event is mental or physical depends on how it is described. This allows him to say that principle III applies to events only described as mental. He then reads principle II as saying that when events are related as cause and effect, they have descriptions that instantiate a law.

I now reformulate Davidson's principles to show how the inconsistency is eliminated:

- I'. At least some mental events interact causally with physical events. (the Principle of Causal Interaction).
- II'. For any events  $x$  and  $y$ , if  $x$  causes  $y$  or  $y$  causes  $x$ , then, there exists a strict deterministic law  $L$  and a description  $d(x)$  such that  $d(x)$  and  $d(y)$  instantiate  $L$ .
- III'. There is no strict deterministic law  $L$  and description  $d(x)$  of any event  $x$  such that  $d(x)$  instantiates  $L$ , where  $d(x)$  is a mental description of  $x$ .

Now I' and II' together only imply that:

- IV'. There exists a description  $d(x)$  and a description  $d(y)$  and law  $L$  such that  $d(x)$  and  $d(y)$  instantiate  $L$ .

IV' does not contradict III', since IV' does not assert that  $d(x)$

and  $d(y)$  must be mental descriptions. In fact, as we shall see, Davidson maintains they must be physical.

### 3.3 Davidson's Argument for Anomalous Monism

In order to reconcile the three principles which he cites, Davidson adopts a position he calls anomalous monism. He characterizes this position as follows:<sup>3</sup>

Anomalous monism resembles materialism in its claim that all events are physical, but rejects the thesis, usually considered essential to materialism, that mental phenomena can be given purely physical explanations. Anomalous monism shows an ontological bias only in that it allows the possibility that not all events are mental, while insisting that all events are physical.

Davidson has characterized his position as a combination of conceptual dualism and ontological monism.<sup>4</sup> It is conceptual dualism in that he does not believe that mental descriptions can be reduced to physical descriptions, or vice versa. It is ontologically monistic in that he believes that mental events are identical to physical events. The form of monism is materialism; he believes that, ontologically speaking, mental and physical events are physical events, even if they have mental descriptions.

In addition to conceptual dualism and ontological materialism, Davidson's anomalous monism involves the view that there are no psychophysical laws. A psychophysical law for Davidson would be one which formulates a lawlike connection between a psychological event

and a physical event. In fact, principle III' as formulated above explicitly states this view:

III'. There is no strict deterministic law  $L$  and description  $d(x)$  of any event  $x$  such that  $d(x)$  instantiates  $L$ , where  $d(x)$  is a mental description of  $x$ .

In this section, I primarily discuss Davidson's argument for III'.

Davidson devotes much of "Mental Events" to defending III'.

While he does not formally state his argument, it is possible to construct a formal argument on the basis of several passages. These include:<sup>5</sup>

In our daily traffic with events and actions that must be foreseen or understood, we perforce make use of the sketchy summary generalization, for we do not know a more accurate law, or if we do, we lack a description of the particular events in which we are interested that would show the relevance of the law. But there is an important distinction to be made within the category of the rude rule of thumb. On the one hand, there are generalizations whose positive instances give us reason to believe the generalization itself could be improved by adding further provisos and conditions stated in the same general vocabulary as the original generalization. Such a generalization points to the form and vocabulary of the finished law: we may say that it is a homonomic generalization. On the other hand there are generalizations which when instantiated may give us reason to believe there is a precise law at work, but one that can be stated only by shifting to a different vocabulary. We may call such generalizations heteronomic.

I suppose most of our practical lore (and science) is heteronomic. This is because a law can hope to be precise, explicit, and as exceptionless as possible only if it draws its concepts from a comprehensive closed theory. This ideal theory may or may not be deterministic, but it is if any true theory is. Within the physical sciences we do find homonomic generalizations, generalizations such that if the evidence supports them, we then have reason to believe they may be sharpened indefinitely by drawing upon further physical concepts: there is a theoretical asymptote of perfect coherence with all the evidence, perfect predictability

(under the terms of the system), total explanation (again under the terms of the system). Or perhaps the ultimate theory is probabilistic, and the asymptote is less than perfection; but in that case there will be no better to be had.

Elsewhere, he says:<sup>6</sup>

There are no strict psychophysical laws because of the disparate commitments of the mental and the physical schemes. It is a feature of physical reality that physical change can be explained by laws that connect it with other changes and conditions physically described. It is a feature of the mental that the attribution of mental phenomena must be responsible to the background of reasons, beliefs, and intentions of the individual. There cannot be tight connections between the realms if each is to retain allegiance to its proper source of evidence.

Based on the above passages, the following argument can be constructed:

1. For any scientific law L which exists, there exists a set of terms and descriptions which forms a closed, homonomic domain whose elements are drawn on by L.
2. There exist scientific physical laws.
3. Therefore, there exists a set of terms and descriptions which forms a closed, homonomic (physical) domain which physical laws draw on. (1,2)
4. If there exists a scientific psychophysical law L', then there exists a set of terms and descriptions which forms a closed, homonomic domain which is drawn on by L'.
5. The set of terms and descriptions drawn on by a psychophysical law is not a closed, homonomic domain.
6. Therefore, no psychophysical laws exist. (4,5)

Does this argument, whose conclusion is that there are no psychophysical laws, provide a defense of III', the principle that no mental descriptions instantiate a law? Davidson says that the claim that there cannot be psychophysical laws "is not quite the



principle of the anomalism of the mental, but on reasonable assumptions entails it".<sup>7</sup> Such an argument can be constructed as follows:

7. There is a strict deterministic law  $L$  and description  $d(x)$  of any event  $x$  such that  $d(x)$  instantiates  $L$ , where  $d(x)$  is a mental description of  $x$ , only if  $L$  is a psychophysical law or a psychological law.
8. There are no psychophysical laws.
9. There are no psychological laws.
10. Therefore, there is no strict deterministic law  $L$  and description  $d(x)$  of any event  $x$  such that  $d(x)$  instantiates  $L$ , where  $d(x)$  is a mental description of  $x$ . (7,8,9)

Davidson would undoubtedly give the same argument for 9 that he gives for 8: that psychological laws do not draw on a homonomic domain. Hence, statements 1 - 10 together constitute an argument for III', the principle of anomalous monism. Clearly, however, the argument for the claim that there are no psychological laws (statement 6) is the most important part of the argument. Hence, I now consider the argument for this claim, statements 1 - 6 above.

There are several important issues and claims relating to this argument. One is the issue regarding how to draw the distinction between mental and physical events and between mental and physical descriptions. Another is Davidson's claim that a scientific law must draw on a set of concepts and descriptions which forms a closed, homonomic domain. He also claims that the set of physical concepts and descriptions does, and the set of mental concepts and

descriptions does not, form such a closed homonomic domain. I discuss these issues and claims in the following two sections.

### 3.4 Davidson's View of Mental and Physical Events

One fundamental assumption which seems implicit in most discussions of the mind-body problem, including Davidson's, is that there is a definite and definable distinction between what is mental and what is physical, and between the terms 'mental' and 'physical'. Davidson explicitly directs his discussion towards mental and physical events, but similar questions arise in relation to distinguishing between mental and physical descriptions, substances, processes, states, attributes, objects, subjects, phenomena, causes, effects, and properties. Even those who admit that they cannot provide a satisfactory way of drawing "the" mental-physical distinction nevertheless insist it can be done, at least "in principle".

How someone who discusses the mind-body problem draws the mental-physical distinction is very important in relation to the discussion. For there is a variety of mental phenomena, and it is easy for someone to make the mistake of thinking he or she has given an account of all mental phenomena, when what is said applies only to one type. For example, supporters of the identity thesis tend to use sensations as their "paradigm" mental phenomenon; if they considered emotions they would find that their arguments are less

compelling. Davidson restricts his discussion to mental events, indeed to ones having intentionality. Like many philosophers, he assumes the primary problem is to characterize "the mental", while characterizing the "physical" is a secondary problem which follows naturally from solving the primary problem.<sup>8</sup>

I now consider Davidson's attempt to distinguish between mental and physical events. He begins by suggesting that "an event is physical if it is describable in a purely physical vocabulary, mental if describable in mental terms".<sup>9</sup> He rejects this criterion if it means that an event is physical if some physical predicate is true of it. He does so on the grounds that:<sup>10</sup>

Every event is truly describable either by 'x took place at Noosa Heads' or 'x did not take place at Noosa Heads'.

Next Davidson offers the following criterion for distinguishing mental verbs from non-mental verbs:<sup>11</sup>

We may call those verbs mental that express propositional attitudes like believing, intending, desiring, hoping, knowing, perceiving, noticing, remembering, and so on. Such verbs are characterized by the fact that they sometimes figure in sentences with subjects that refer to persons, and are completed by embedded sentences in which the usual rules of substitution appear to break down.

He then offers an example to show that it includes many events we would intuitively accept as physical:<sup>12</sup>

However, the criterion actually covers not only the havings of pain and afterimages, but much more besides. Take some event one would intuitively accept as physical, let's say the collision of two stars in distant space. There must be a purely physical predicate 'Px' true of this collision, and of others, but true of only this one at the time it occurred. This particular time, though, may be pinpointed as the same time that Jones notices that a pencil starts to roll across his desk. The distant

stellar collision is thus the event  $x$  such that  $Px$  and  $x$  is simultaneous with Jones's noticing that a pencil starts to roll across this desk. The collision has now been picked out by a mental description and must be counted as a mental event.

Davidson correctly concludes that "this strategy will probably work to show every event to be mental".<sup>13</sup>

A similar strategy shows every event to be physical. Simply reverse the example and take as an intuitively mental event Jones' noticing that a pencil starts to roll across his desk. Suppose that there is a purely mental predicate ' $Mx$ ' true of this act of noticing, and of others, but true of only this one at the time it occurred. This particular time, though, may be pinpointed as the same time two stars collide in distant space. The act of noticing is thus the event  $x$  such that  $Mx$  and  $x$  is simultaneous with the collision of two distant stars. This act of noticing has now been picked out to be a physical description and must be counted as a physical event.

Davidson's criterion seems to leave us with the conclusion that every event is mental and every event is physical. This may not be an undesirable result, but it requires further examination. Ordinarily, " $x$  is mental" is thought to rule out " $x$  is physical"; they are regarded as being mutually exclusive categories. Davidson appears to accept this conception of the relation between "mental" and "physical". Commenting on the strategy which shows every event to be mental, he says "it would be instructive to try to mend this trouble".<sup>14</sup> I now consider how this could be done.

According to Davidson's criterion, an event is mental if and only if it can be picked out by a mental description or open sentence, where this is one which contains at least one mental verb essentially. Often a description picks out some property or feature which an event can have or relation it can stand in, and ascribes this property, feature, or relation to the event. One reason Davidson's criterion doesn't work is because mental events can stand in relations to physical events, and physical events can stand in relations to mental events. In Davidson's example, the distant stellar collision stands in the relation of simultaneity to Jones' act of noticing that a pencil started to roll across his desk. Likewise (though this reciprocity doesn't hold for all relations), Jones' noticing act stands in the relation of simultaneity to the distant stellar collision. Hence, the distant stellar collision can be picked out by a description which specifies this relation of simultaneity to the act of noticing, and vice versa.

These examples are not idiosyncratic; in fact, every mental event and every physical event stands in many types of relation. Some are more trivial than others, but this does not mean they do not obtain. A new Rolls Royce which is the grand prize in a drawing might be the object of the desires of hundreds, thousands, or even millions of people. An astronomical event such as a solar eclipse can be viewed by millions of people; this event is the object of perception by them. We can pick out a physical object by saying

"It's the one I'm looking at". In all these examples, mental and physical events stand in relations to each other.

It might be suggested that we rule out descriptions which ascribe relations to events in our criterion for mental and physical events. Besides being arbitrary, it would often be difficult to specify location or place. For example, suppose a group of identical bottles (identical in type) is being processed through a bottling plant. The most natural way, and perhaps the only way, to distinguish them is by their position or their relation(s) to some other thing. For example, one might say "The bottle in the front row, second from the left". This is to identify the bottle by appealing to its position in the row. Of course, sometimes we can identify objects without specifying a relation, as if only one bottle is cracked, we can say "the cracked bottle". But in cases in which a number of objects look the same, we rely on relational ascriptions.

The suggestion that we rule out relational ascriptions is even less attractive when applied to mental events. For Davidson makes intentionality his criterion of the mental, and intentional verbs carry a relational component. This is sometimes expressed by saying they have "aboutness", or they are "about" something. If I think that I might take the train and I think that I might take my car, the difference between these thoughts is that one is about the train and the other is about my car. The train and the car are the objects of my thought. Likewise, I might be angry at Jones, or have



love for my best friend. While philosophers have debated the nature of intentional objects, it seems clear that in many cases at least, identifying a mental event requires making a relational ascription. Often the relation ascribed is to the object of the mental event, which is physical in many cases.

It seems, then, that we cannot fix Davidson's criterion by ruling out descriptions of relations from the set of eligible descriptions. This still leaves us with Davidson's original criterion, which seems to have unacceptable consequences. Bernard Katz spells out another, though similar, consequence of Davidson's criterion in these remarks:<sup>15</sup>

Suppose *m*, a mental event, caused or was caused by *p*, a physical event. Then there will be a physical description *P* which designates *p*, and *m* will be describable as 'the cause of *P*' or 'the effect of *P*'. So if we accept the additional claim that every mental event either caused or was caused by a physical event, it will follow almost immediately that every mental event has a physical description. This procedure for providing physical descriptions to mental events is, no doubt, a trivial one. It does, however, accomplish the task Davidson sets out for a proponent of the identity theory. Of course, if this is what the identity theory comes to, one wants to know who ever denied it.

Here Katz demonstrates another way of showing that it follows from Davidson's criterion that events which we intuitively take to be mental are physical, and vice versa; indeed, that all events (or at least all those with physical or mental causes or effects) are both mental and physical.

Davidson is certainly right in saying about the problem with his criterion that "it would be instructive to try to mend this

trouble".<sup>16</sup> For what trying to mend this trouble show is that the very assumption that the ontological distinction between mental and physical events can be correctly drawn by a linguistic criterion is suspect. Indeed, Davidson's assumption that one can make the effortless inference from "x has a mental description" to "x is a mental event" has unacceptable consequences. Bernard Katz points out two consequences of this assumption, which he characterizes, following Davidson, as "an event is mental only under a description". One way of reading this, he says, is "an event is mental only if it is correctly described by means of a mental description". He says:<sup>17</sup>

Suppose M, a mental description, and P, a physical description, are singular terms which designate the same event. On this interpretation 'M is a mental event' will be true, while 'P is a mental event' will be false. A consequence of this would be that mental events do not form a well-defined set of events. (If there were such a set, the event designated by M would be a member of it, while the event designated by P would not; which is to say one and the same event would both be a member and not be a member of the set.) Another consequence would be that the central thesis of the identity theory - that every mental event is identical with some physical event - would not make clear sense. That claim, understood literally, could be put as follows:

$$(\forall x)(x \text{ is a mental event} \rightarrow (\exists y)(y \text{ is a physical event} \wedge x = y))$$

The truth conditions of this quantified sentence make reference to the set of events which satisfy the antecedent of the open sentence obtained by deleting the initial quantifier. That set, of course, is the set defined by the open sentence 'x is a mental event'. As we have just seen, however, there is no such set.

Katz goes on to consider a different reading:<sup>18</sup>

A different reading of the claim that events are mental only as described might be that mental events are just those events which have mental descriptions; that is, mental events form the subset of events which have mental descriptions. If this is what is intended, however, it seems odd to say that events are "mental only as described": surely a mental event will be mental (that is, a member of the relevant subset) regardless of how it is described. In any case, a problem with Davidson's use of this understanding of "mental events" is that it should come as a conclusion toward the end of an argument in favor the identity theory, and not as a premiss at the outset. The assumption that there really is not any ontological distinction to be made at all, or that [the] mental/physical distinction is only a linguistic one, begs the question. For it presupposes the truth of an important part of what is in dispute - namely, that interesting forms of dualism (versions which assert ontological as opposed to linguistic dualism) are mistaken.

As Katz says, by assuming in advance that the mental-physical distinction is only linguistic and not ontological, Davidson assumes the falsehood of ontological dualism. But this is precisely the question at issue.

I conclude, then, that Davidson has failed to provide an adequate criterion for distinguishing mental from physical events, and that attempts to mend his criterion only further undermine the assumption underlying it: that it is possible to provide a linguistic criterion for making the ontological distinction between mental and physical events. Not only does this assumption have unacceptable consequences, but it also seems to beg the question against the ontological dualist.

One fact which does emerge from the above discussion, however, is one already alluded to in chapter II; namely, that the concepts "mental" and "physical" are much more closely interconnected than

most discussions of the mind-body problem indicate. For these discussions generally presuppose, as Davidson's does, that there is a clearly definable set of mental events, objects, properties, etc. which is distinct from a clearly definable set of physical events, objects, properties, etc. In fact, it is this assumption which has, in large part, made the mind-body problem seem unsolvable. In chapter V I discuss several attempts to find a single criterion for identifying mental phenomena. In chapter VI, I distinguish three conceptions of what is physical. If there are three conceptions of what is physical, then there is not one clearly definable set of physical events, objects, properties, etc.

### 3.5 Davidson's Mental and Physical Domains

The first premise of Davidson's argument for premise III, which formulates anomalous monism, reads as follows:

1. For every scientific law L there exists a set of concepts and descriptions which forms a closed, homonomic domain whose elements are drawn on by L.

This is the key premise in Davidson's argument, since, once this premise is accepted, he goes on to argue that physical laws satisfy this condition while psychophysical laws do not. Davidson apparently believes that there is a physical domain drawn on by physical laws; this domain is homonomic. The mental domain is not homonomic, so any "laws" which must draw on this domain do not

satisfy a necessary condition of scientific laws. Hence, they are not scientific laws. In this section, I discuss premise 1.

I argued in section 3.4 that Davidson has not provided an adequate or accurate way of distinguishing mental from physical events. Arguments in section 3.4 also cast doubt on his assumption that the ontological mental-physical distinction can be drawn by a linguistic criterion. The task in this section is linguistic rather than ontological: to consider whether there is a homonomic physical domain made up of physical concepts and descriptions which is distinct from a mental domain made up of mental terms and descriptions. Then I will consider whether a scientific law must draw on a homonomic domain.

First I consider whether there is a homonomic conceptual domain of the physical made up of physical terms and descriptions. This question turns on whether there is a set of physical terms and descriptions which is homonomic in Davidson's sense. I begin by considering whether there is a set of physical terms and descriptions.

Unfortunately, Davidson does not make any positive attempt to characterize what is "physical". He says this problem is "recessive" in relation to characterizing what is "mental". Davidson characterizes mental events by way of mental descriptions, and mental descriptions are ones which contain at least one mental verb essentially.<sup>19</sup> He says that "we may call those verbs mental that express propositional attitudes like believing, intending,

desiring, hoping, knowing, perceiving, noticing, remembering, and so on".<sup>20</sup> He admits that his criterion does not include "events that have often been considered paradigmatic of the mental"<sup>21</sup> such as feeling a pain or having an afterimage. However, he does not choose to modify his criterion.

If Davidson is right that characterizing a physical vocabulary is "recessive" in relation to characterizing the mental vocabulary, then we might suppose he means that all verbs which do not exhibit intentionality are physical. This does not seem acceptable, however, even apart from mental verbs he admits do not fit his criterion such as 'to feel' as in "to feel a pain" and "to see", when what is seen is an afterimage. There are other verbs such as 'to be', 'lack', 'live' (as in "He lives"), 'function', 'grow', 'neuter', 'boom', 'crash' (both of which two can apply to the stock market), 'evolve', 'has', and 'pump' which are not intentional but do not seem "physical" either. Let us see whether we can give a positive characterization of a physical verb, or, more broadly, a physical vocabulary.

When some philosophers talk about a physical vocabulary, they mean "the language of physics". However, since Davidson's discussion is directed towards the human brain and body, I doubt whether he wants to restrict his description of neural events to descriptions of mechanics, heat, acoustics, optics, electricity, magnetism, and atomic structure. If he does, the physical events would be molecular in nature, and certainly no one has succeeded in



reducing neurology to physics. Similar considerations rule out saying "the language of chemistry and physics". Such a characterization would leave the statements of the neurologist in limbo between the mental and physical domains.

Perhaps Davidson means be a physical vocabulary "the language of the physical sciences". Let us suppose this, and suppose what is not obvious: that biology and neurology, as well as physics and chemistry, are included in this domain.

What about the "vocabulary" or "domain" part of our attempt to define a physical vocabulary or domain? Davidson originally draws the homonomic/heteronomic distinction as being between two types of generalizations, but his talk of a physical "vocabulary" makes it clear that a conceptual domain is a subset of a broader language. Hence, a conceptual domain must include certain logical and mathematical terms which are neither physical nor mental. It will also have a syntax.<sup>22</sup> This leaves the descriptive terms as being the ones which make a conceptual domain of a certain type. The physical domain, then, will contain at least the following elements:

1. Logical terms ('and', 'or', 'if', etc.)
2. Mathematical terms ('number', 'line', 'point', etc.). The mathematics chosen can be governed by pragmatic considerations.
3. Syntactic rules.
4. Physical terms which are descriptive. These terms can be nouns, verbs, adjectives, or adverbs.

Obviously, it is condition 4 which makes the above a description of a physical domain; it is possible to construct a description of a mental domain (and perhaps of other types as well) by replacing the term 'physical' with the term 'mental' (or some other term).

Given this conception of a physical domain, I now consider whether this domain is a closed, homonomic domain. For the moment, at least the languages of physics, chemistry, biology, and neurology are included within this domain (these are among the physical sciences); hence, the question turns on whether the descriptive terms and generalizations of these sciences are homonomic.

I first consider biology. A widely accepted conception of biology is that it is the study of living things or organisms. Biologists study human beings, animals, and plants. One descriptive term that occurs in biology is 'organ'. Examples of organs are sense organs such as the eye, ear, nose, and the touch and taste receptors. Terms for biological organs exhibit a feature that is common among biological terms: they are functionally defined. This means they apply by definition to things which perform a certain function. And in the case of a sense organ, the terms apply to particular types of tissue that perform a mental function. For example, the ear is by definition a tissue of a certain type whose function is to enable the organism to hear. Likewise, the eye is by definition an organ whose function is to enable the organism to see, etc. Hence, certain fundamental descriptive terms of biology

contain an implicit reference to the mental domain, and could not be allowed in a purely physical vocabulary.

Of course, the term 'eye' does refer to an organ which has a physical aspect. The eye is a mass of tissue, and perhaps it can also be considered as simply a piece of matter. But if biologists were only to study the eye as a piece of matter, independent of its mental function, they would really be treating it from the viewpoint of physics. And from this viewpoint they would be unable to account for most of its interesting properties. Hence, proposing to save the homonomic nature of biological terms and generalizations by stripping these terms of their implicit mental reference would be to strip biology of much of its explanatory and descriptive ability.

The following quotation from a discussion of human vision helps substantiate this point:<sup>23</sup>

Hubel and Wiesel suggest that the interaction between simple and complex cells may be important for the [following] issue:...the perception of patterns irrespective of where they fall on the retina. Take a vertical line--or an object with vertical edges--that is located in front of you. The vertical edge (or the line) will excite a population of simple cells and a population of complex ones, each with a vertical axis of orientation. A slight (saccadic) movement of your eye or of the object will call into play a new population of simple cells (because these cells are very sensitive to the exact position of the line in the receptive field). However, for a small movement the stimulus will still excite the original population of complex cells (because these cells are less sensitive to movement within the receptive field)....

The following discussion of color vision by Peter Gouras also shows how important the use of terms containing a mental reference are in describing and explaining how human and animal eyes work:<sup>24</sup>

Animals with color vision simply have great perceptual capability, and this presumably has great survival value. Think for a moment of a black and white version of a work by a colorist such as Turner, Monet, or Renoir; almost a million nuances of contrasting shapes are lost by an achromatic rendition of one of their works. This high degree of discrimination in color vision can actually be understood in quantitative terms. The subjective experience of color can be broken down into three semi-independent sensibilities: (1) hue, (2) saturation, and (3) brightness. All color experience is composed of these three psychological impressions.

Hue has the strongest effect on color and is the major determinant of principal colors such as red, yellow, green, and blue. It is what we ordinarily mean by "color". This impression is determined by the proportion in which the three cone mechanisms are activated by the object and its background. The brain must keep track of how much each of the three photoreceptor systems contributes to the detection of an object. Most of us have names for only a family of hues, but actually about 200 varieties can be distinguished. The second distinct quality of color is saturation, which reflects how much a hue has been diluted by grayness; this is determined by the degree to which all three cone mechanisms are stimulated in common by the object and by the background. There are about 20 distinguishable steps of saturation for each hue.

The third quality of color, brightness, is a sensation shared with achromatic visual systems. It is due to the total effect on all three cone mechanisms of an object relative to its background. We shall see later that one of the three cone mechanisms (the so-called blue mechanism) makes little or no contribution to brightness. It is the brightness factor that turns orange into brown and gray into black or white. There are about 500 distinguishable steps of brightness for every hue and grade of saturation. In contrast to achromatic vision with only 500 steps, color vision has more than 1 million gradations with which to detect the contours of shapes in the external world....It is no wonder that natural selection uncovered its power.

By the latter part of the eighteenth century it was realized that the ability to experience myriad different colors might not be due to an equivalent number of detectors in the retina but to a minimum number--three--with every color determined by the proportions in which each of these three detectors responds to light.

The issue raised here regarding the use of terms with a mental reference in biology is related to a more general issue discussed at some length in the literature of the philosophy of science. This issue is the one of functional explanation within science.<sup>25</sup>

There is also discussion of whether biology can be "reduced" to physics. One important consideration in relation to any such "reduction" is whether it is possible to give a physical explanation or definition of the concept of "life", a fundamental concept in biology. While I won't pursue these more general questions here, I believe the discussion does show, as did the discussion in chapter II, that 'mental' and 'physical' are more closely interconnected than they appear to be. This result is examined in more detail in chapter VII.

A more specific conclusion is that Davidson cannot include the language of biology within his physical domain and also maintain that this domain is homonomic. Perhaps he could provide a conception of "physical" which somehow disallows the above objection. But since he has not even provided a concept of what is "physical", I conclude that either Davidson's physical domain is not homonomic or else it does not include the language of biology.

Because the basis for Davidson's argument for anomalous monism is his claim that the physical domain is homonomic, I believe he would prefer to shrink his domain than to give up the homonomic nature of the physical domain. However, this leaves the language of biology

in a nebulous middle domain between the mental and the physical domains.

Similar considerations show that the descriptive terms of neurology cannot be included in Davidson's homonomic physical domain. The brain itself can be defined as the part of the central nervous system associated with thought, perception, and emotion. This is a definition in terms of a mental function. Also, specific parts of the brain are defined in relation to the mental functions they perform. The cerebrum is the part associated with conscious mental processes. Other parts of the central nervous system such as the touch and heat receptors are defined by their mental function. Hence, the language of neurology cannot be included in Davidson's homonomic physical domain either.

This problem is related to one raised at the end of chapter II. Here I pointed out that if a materialist wishes to specify some brain state to be identified with a mental state, he must either do so via a mental criterion or a physical criterion. If he uses a mental criterion (e.g. "the bodily state correlated with pain"), then he has already acknowledged the distinctness of pain from the bodily state (since the individuating principle of  $x$  must be distinct from  $x$ ). However, he cannot use a physical criterion since there is no bodily state which correlates strictly with pain. It seems clear that a neurologist, who must succeed in picking out neurological structures, does so in many cases via a mental criterion. Once he has picked them out, he can proceed to study how



they function in relation to the rest of the body, and in relation to other mental functions.

Perhaps Davidson would reply that many of the terms for the mental functions such as 'seeing', 'hearing', 'touching', and 'tasting' by reference to which certain organic structures are defined do not exhibit intentionality and hence do not fit his criterion for being mental terms. If this is so, then terms which implicitly refer to these supposedly mental functions will still be homonomic since the reference will still be to the physical domain. If Davidson wished to argue in this fashion, it would be surprising, since he lists 'perceiving' as a mental verb.<sup>26</sup> However, this reply mostly shows the inadequacy of his criterion since any criterion which makes terms such as 'seeing', 'hearing', and 'feeling' (which Davidson admits are paradigmatic of the mental) non-mental is almost certainly wrong. Finally, not all the examples I have given are of non-intentional verbs; for example, the cerebrum was defined as being the part of the brain associated with conscious mental processes. Actually, many of these terms have intentional and non-intentional uses (as Davidson acknowledges), so that Davidson's criterion makes it difficult to decide whether a given term is in an abstract conceptual domain or not.

Davidson might also reply that it is incorrect to view "the organ of hearing" as a definition of 'ear', and so on for the other examples; that these are at most true descriptions of these organs. While I don't accept this reply, it wouldn't help Davidson defend

the homonomic character of biology and neurology, since, on his own principles, any event with a mental description is a mental event. Hence, generalizations such as "The function of the ear is hearing" and "The function of the cerebrum is to enable the person to think" will not be a part of the physical domain because they are not homonomic. However, we saw earlier that a biology or neurology stripped of all reference to mental events would really be physics and would be so impoverished that it would not accomplish the purposes of its science.

If the languages of biology and neurology are not to be included in the physical domain, we seem to be left with the languages of chemistry and physics. (Of course, there are other physical sciences; I am picking representative examples). The question whether the conceptual domain consisting of the languages of chemistry and physics is homonomic depends on whether the descriptive terms of the fields are homonomic in the sense that they do not contain a reference to any non-physical domain. Examples of such terms are 'kilogram', 'mass', 'energy', 'calorie', 'proton', 'gas', 'wave', 'particle', 'heat', 'field', 'watt', 'mile', etc. The objection raised for biological terms does not apply here since these terms are not defined by reference to a mental function. I believe that one can raise important questions about these phenomena; for example, they don't all fit the standard Cartesian conception that physical things are extended. Force fields, waves, and rays do not seem to be extended in any obvious sense. However,

I will not discuss this issue here and will grant that the language of chemistry and physics forms a closed, homonomic domain.

Let us recall the premise under discussion:

1. For every scientific law L there exists a set of concepts and descriptions which forms a closed, homonomic domain whose elements are drawn on by L.

Now that I have discussed in some detail what a homonomic conceptual domain might look like, and have tried to determine what Davidson's homonomic physical domain is, I will look at Davidson's more general claim that every scientific law must draw on a homonomic domain.

### 3.5.1 Davidson's Claim that Every Scientific Law Must Draw on a Homonomic Domain

Davidson does not argue for the claim that every scientific law must draw on a homonomic domain, but it has some interesting consequences. One is that there are no scientific laws of biology, neurology, psychology, history, sociology, or economics since these disciplines all draw on heteronomic conceptual domains. Davidson acknowledges this consequence in "Psychology as Philosophy":<sup>27</sup>

The nomological irreducibility of the psychological means, if I am right, that the social sciences cannot be expected to develop in ways exactly parallel to the physical sciences, nor can we expect ever to be able to explain and predict human behaviour with the kind of precision that is possible in principle for physical phenomena.

Davidson also says:<sup>28</sup>

When the world impinges on a person, or he moves to modify his environment, the interactions can be recorded and codified in ways that have been refined by the social sciences and common sense. But what emerge are not the

strict quantitative laws embedded in sophisticated theory that we confidently expect in physics, but irreducibly statistical correlations that resist, and resist in principle, improvement without limit.

This quotation invites the observation that Davidson may be operating with an outmoded conception of physical laws. Since the advent of quantum theory, scientists have come to believe that many laws of physics are statistical in nature; if this is so, the deterministic model of scientific law may be incorrect or only partially correct. Davidson says elsewhere of this that "perhaps the ultimate theory is probabilistic and the asymptote is less than perfection; but in that case there will be no better to be had".<sup>29</sup> But if his basis for rejecting psychological laws is that they are only statistical and cannot be improved without limit, then perhaps probabilistic laws in physics should be rejected too. Perhaps the problem is not the concept of probabilistic law but Davidson's conception of scientific law.

The claim that scientific laws must draw on a homonomic domain (premise 1 of Davidson's argument for principle III, which formulates anomalous monism), is not only unargued for; it is not self-evident or obviously true, as Ted Honderich observes:<sup>30</sup>

There are then two descriptive domains or systems, depending upon disparate connections between the systems. There can be no psychophysical lawlike connections. There is no gainsaying the difference between the two domains, and the account given of the mental is perhaps more impressive than any predecessor.... Still, we are not given a reason for thinking that the given conclusion follows from the description of the two domains. As others have asked, what reason is there for thinking an item which falls in one domain, and whose description then

depends on X, cannot be in lawlike connection with an item in the other domain, whose description then depends on Y? There is no clear general truth to the effect that there cannot be lawlike connection between items whose descriptions have different necessary conditions. There appears to be no relevant sense in which two domains can be said to be 'closed' and which precludes lawlike connection without begging the question. What then is the relevant feature of this admittedly unique pair of domains? Davidson remarks that his argument is not proof. It must also be said, I think, that his argument is at least crucially incomplete.

Davidson's requirement that scientific laws draw on a homonomic domain also has an important consequence for his claim that mental events are identical to physical events. For according to Davidson, if some physical event F causes a mental event, then there is a physical description of the mental event which, together with F, instantiates a physical law that draws on a homonomic domain. However, we saw above that the biological and neurological domains are not homonomic, so that the physical event which the mental event is identical to will not be a brain event but a chemical event or an event involving concepts from physics such as mass, weight, or atomic structure. Furthermore, the physical event which causes the mental event must be of this type as well. This is not a happy consequence for someone who finds it difficult to defend the view that a mental event is identical to some "high-level" brain event. Identifying a mental event with some molecular event is even less plausible.

Bruce Goldberg comments on this tension within Davidson's use of 'physical' and concludes that he is using this term in two senses. He begins with a quote from Davidson:<sup>31</sup>

The demonstration of identity follows easily. Suppose *m*, a mental event, caused *p*, a physical event; then under some description *m* and *p* instantiate a law. This law can only be physical.... But if *m* falls under a physical law it has a physical description; which is to say it is a physical event.<sup>32</sup>

The difficulty with this argument is that it appears to equivocate between two different senses of the term 'physical'. In one sense, that in which every physical event falls under a law, it is not clear that mental events do cause physical events. In the other, that in which there are clear classes of mental events causing physical ones, it is not clear that the physical events fall under any law.

An example of the sense of 'physical' according to which mental events cause physical events would be Davidson's case of a physical event, namely a certain bodily movement, causing the sinking of a ship:

...and the fact that someone sank the Bismarck entails that he moved his body in a way that was caused by mental events of certain sorts, and that this bodily movement caused the Bismarck to sink.<sup>33</sup>

But a different sense of 'physical' that in which every physical event falls under a law, appears to be involved in the following passage, where Davidson is concerned with the notion of an ultimate physical theory:

Physical theory promises to provide a comprehensive closed system guaranteed to yield a standardized, unique description of every physical event couched in a vocabulary amenable to law.<sup>34</sup>

Davidson's requirement, then, that physical laws draw on a homonomic domain entails that he use 'physical' in the sense of "physical theory", while his attempt to solve problems of causation of mental events leads him to use 'physical' in the sense in which the bodily movement of a person is a physical event.

I conclude the following with respect to premise 1 (that a scientific law must draw on a homonomic domain):



1. Davidson has not established this premise since he does not argue for it.
2. The premise is not obviously true, and there are reasons for doubting it.
3. When this claim is coupled with Davidson's claim that the physical conceptual domain is homonomic, the result is that Davidson's physical domain shrinks to the languages of chemistry and physics. This is so because the conceptual domains of biology and neurology are not homonomic.
4. The premise is based on a deterministic conception of scientific law which does not take into account the probabilistic nature of some laws of physics and psychology.
5. The premise has the consequence that there are no scientific laws of biology, neurology, psychology, history, or economics. In order to defend this consequence, an anomalous monist would have to give a more convincing account of laws and explanation in these sciences than Davidson's account in "Psychology as Philosophy".
6. This premise has undesirable consequences for Davidson's identity claim, since Davidson is committed to identifying mental events not with brain events but with molecular events.
7. This claim generates a fundamental tension within Davidson's discussion between a restricted sense of 'physical' (as in the language of physics and chemistry) and a broader sense in which a bodily movement made by a person is a physical event.

### 3.6 Davidson's Argument for Anomalous Monism

Let us now put premise 1 in the context of Davidson's more general argument. This premise is the first in an argument whose conclusion is that there are no psychophysical laws:

1. For any scientific law  $L$  which exists, there exists a set of terms and descriptions which forms a closed, homonomic domain whose elements are drawn on by  $L$ .
2. Physical laws exist.
3. Therefore, there exists a set of terms and descriptions which forms a closed, homonomic (physical) domain which physical laws draw on. (1,2)
4. If there exists a psychophysical law  $L'$ , then there exists a set of terms and descriptions which forms a closed, homonomic domain that  $L'$  draws on.
5. The set of terms and descriptions that a psychophysical law draws on is not a closed, homonomic domain.
6. Therefore, no psychophysical law exists. (4,5)

I also formulated the following four statements to make complete an argument for principle III':

7. There is a strict deterministic law  $L$  and description  $d(x)$  of any event  $x$  such that  $d(x)$  instantiates  $L$ , where  $d(x)$  is a mental description of  $x$ , only if  $L$  is a psychophysical law or a psychological law.
8. There are no psychophysical laws.
9. There are no psychological laws.
10. Therefore, there is no strict deterministic law  $L$  and description  $d(x)$  of any event  $x$  such that  $d(x)$  instantiates  $L$ , where  $d(x)$  is a mental description of  $x$ . (7,8,9)

Principle III', together with principles I' and II', represent Davidson's way of reconciling three apparently incompatible principles:

- I'. At least some mental events interact causally with physical events. (the Principle of Causal Interaction).
- II'. For any events  $x$  and  $y$ , if  $x$  causes  $y$  or  $y$  causes  $x$ , then, there exists a strict deterministic law  $L$  and a description  $d(x)$  such that  $d(x)$  and  $d(y)$  instantiate  $L$ .

III'. There is no strict deterministic law  $L$  and description  $d(x)$  of any event  $x$  such that  $d(x)$  instantiates  $L$ , where  $d(x)$  is a mental description of  $x$ .

Without premise 1 to help justify III', I conclude that Davidson has failed to establish III'. This does not mean that I', II', and III' do not together form a possible resolution of an apparent contradiction. It does mean, however, that Davidson has failed to establish that III' is true; furthermore, there are reasons to doubt that it is true. These reasons to doubt it are primarily the reasons listed above to doubt premise 1 (that scientific laws must draw on a homonomic domain) on which it is based. These reason include the seemingly unacceptable consequences of premise 1. Since principle III' formulates Davidson's anomalous monism, I conclude that Davidson has failed to establish anomalous monism, and that there are good reasons for doubting that anomalous monism is true.

### 3.7 Davidson's Demonstration of Identity

The monistic part of anomalous monism is the claim that mental events are identical to physical events. Davidson states his demonstration as follows:<sup>35</sup>

The demonstration of identity follows easily. Suppose  $m$ , a mental event, caused  $p$ , a physical event; then, under some description  $m$  and  $p$  instantiate a strict law. This law can only be physical, according to the previous paragraph. But if  $m$  falls under a physical law, it has a physical description; which is to say it is a physical event. An analogous argument works when a physical event causes a mental event. So every mental event that is

causally related to a physical event is a physical event. In order to establish anomalous monism in full generality it would be sufficient to show that every mental event is cause or effect of some physical event; I shall not attempt this.

We can construct the following argument based on this passage:

1. Suppose  $m$ , a mental event, causes  $p$ , a physical event.
2. If  $m$  causes  $p$ , then under some descriptions  $d(m)$  and  $d(p)$ ,  $m$  and  $p$  instantiate a strict law  $L$ .
3. There is no strict deterministic law  $L$  and description  $d(x)$  of any event  $x$  such that  $d(x)$  instantiates  $L$ , where  $d(x)$  is a mental description of  $x$ .
4. The only strict laws are physical laws.
5.  $L$  is a physical law. (1,2,3,4)
6. If  $m$  falls under a physical law, it has a physical description.
7. Therefore  $m$ , a mental event, is identical to a physical event.

There are two main objections to this argument. The first has to do with premise 7. We saw above that the quick inference from " $x$  has a physical description" to " $x$  is a physical event" is unacceptable. It is unacceptable because it is possible to give a mental description of any physical event. Davidson sees this problem but thinks it can be mended; on the contrary, the attempt to mend it simply further exhibits the reasons why it is untenable. On this criterion, most and perhaps all events are both mental and physical. This conclusion might be acceptable to Davidson as a conceptual claim, but not as an ontological claim. A distant stellar collision should not be and is not a mental event. Yet it is, as Davidson admits, if one assumes that " $x$  has a mental

description" justifies the inference to "x is a mental event". This linguistic criterion should not be mended; it should be abandoned.

Secondly, we saw above the Davidson has failed to establish anomalous monism, which is stated as premise 3. Hence, Davidson has not demonstrated that one can infer from the fact that "m and p instantiate a strict law" to the claim that "that law can only be physical". Davidson has failed to justify this crucial step in his argument, and there are reasons for believing that this inference is fallacious. These reasons are the same as the reasons for doubting premise 1 above, that scientific laws must draw on a homonomic domain.

In conclusion, it seems clear that Davidson's "demonstration" of identity relies on two very dubious claims: that "x has a physical description" implies "x is a physical event" and the claim that there are no psychophysical laws. The first claim seems clearly unacceptable, while Davidson has not established the second claim. Furthermore, there are good reasons for believing that this claim is false. Hence, I conclude that Davidson has failed to demonstrate the identity of mental events and physical events, and that there are good reasons for doubting the correctness of his attempt.

### 3.8 Conclusion

Davidson formulates anomalous monism as a way of reconciling three apparently incompatible principles. Anomalous monism has an "anomalous" part and a "monistic" part, and can be stated as follows:

1. There is no strict deterministic law  $L$  and description  $d(x)$  of any event  $x$  such that  $d(x)$  instantiates  $L$ , where  $d(x)$  is a mental description of  $x$ .
2. Mental events are identical to physical events.

In this chapter, I have considered both these claims. I conclude that Davidson fails to establish either one of them, and that there are good reasons for doubting his arguments for both of them. Thus, unless further, different arguments are made, there are good reasons for doubting anomalous monism.



## FOOTNOTES

1. Donald Davidson, Essays on Actions and Events (Oxford: Clarendon Press, 1980). All three articles appear in this book, and all quotations from Davidson in this article refer to this book. "Mental Events" appears on pages 207-224, "Philosophy as Psychology" from pages 229-244, and "The Material Mind" from pages 245-260.
2. Ibid., p. 208.
3. Ibid, p. 214.
4. In conversation during a tutorial I took with Donald Davidson at The Rockefeller University in 1975.
5. Davidson, Ibid., p. 219.
6. Ibid., p. 222.
7. Ibid., p. 209.
8. What Davidson actually says is that "It is less important to characterize a physical vocabulary because relative to the mental it is, so to speak, recessive in determining whether a description is mental or physical". (Ibid., p. 211).
9. Ibid., p. 210.
10. Ibid.
11. Ibid.
12. Ibid., p. 211.
13. Ibid., p. 212.
14. Ibid.
15. Bernard Katz, "Davidson and the Identity Theory", Canadian Journal of Philosophy, Volume VII, Number 1, March, 1977.
16. Davidson, p. 212.
17. Katz, p. 89.
18. Katz, pp. 89-90.
19. Davidson, p. 211.

20. Ibid., p. 210.
21. Ibid., p. 211.
22. Davidson says "Think of the physical vocabulary as the entire vocabulary of some language L with resources adequate to express a certain amount of mathematics, and its own syntax". (Ibid., p. 214).
23. Eric R. Kandel, "Visual System III: Physiology of the Central Visual Pathways", in Principles of Neural Science, ed. Eric R. Kandel and James H. Schwartz (New York: Elsevier/North Holland, 1981), p. 245.
24. Peter Gouras, "Visual System IV: Color Vision", in Kandel and Schwartz, op. cit., p. 253.
25. See, for example, Carl Hempel, "The Logic of Functional Analysis", in Aspects of Scientific Explanation (New York: The Free Press, 1965), pp. 297-330. Also, see Jerry Fodor, Psychological Explanation (New York: Random House, 1968), esp. pages xvi-xxi, pp. 111-119, and pp. 136-145.
26. Davidson, Ibid., p. 210.
27. Donald Davidson, "Philosophy as Psychology" in "Essays on Actions and Events", p. 230.
28. Ibid.
29. Davidson, from "Mental Events", in Essays on Actions and Events, p. 219.
30. Ted Honderich, "Psychophysical Lawlike Connections and Their Problem" in Inquiry, 24, pp. 292-3.
31. Bruce Goldberg, "A Problem with Anomalous Monism" Philosophical Studies, 32, 1977, p. 178.
32. Davidson, "Mental Events", p. 224.
33. Ibid., p. 208.
34. Ibid., pp. 223-4.
35. Davidson, in "Mental Events", p. 224.

## CHAPTER IV

### DANIEL DENNETT'S DEFENSE OF FUNCTIONALISM

#### 4.1 Introduction

In Brainstorms, Daniel Dennett gives an account of mental phenomena that is functional or intentional in nature. While the book is a collection of essays, it nevertheless constitutes a sustained, if not completely systematic, attempt to defend a theory of mind that he at one point calls "type intentionalism".<sup>1</sup> Dennett distinguishes this view from several other theories which have been variously advocated, including type identity theory, Turing machine functionalism, and token functionalism. Dennett apparently regards his theory as a form of physicalism, and is concerned to distinguish his formulation of physicalism from others that have been offered.

Dennett characterizes type intentionalism as follows:<sup>2</sup>

If one insisted on giving a name to this theory, it could be called type intentionalism: every mental event is some functional, physical event or other, and the types are captured but by any reductionist language but by a regimentation of the terms we ordinarily use - we explain what beliefs are by systematizing the notion of a believing-system, for instance.

In later chapters he makes an attempt to give such a systematized account of a number of different mental phenomena, including

consciousness, mental images, beliefs, experience, and pain. His account of pain is one of the most interesting, since pain is what has been called a "program-resistant" feature of mentality.<sup>3</sup> That is, it does not seem implausible to suppose that computers can simulate or duplicate certain "intellectual" features of mind such as chess-playing or theorem-proving. It seems much less plausible, however, to ascribe other phenomena such as feelings and emotions to a computer.

#### 4.2 Dennett's Flowchart of Pain

What I primarily wish to do here is to examine Dennett's account of pain. He devotes chapter 11 to giving such an account, and it is given in terms of the language of cognitive psychology. Dennett begins by noting that, contrary to Turing, a computer simulation need not be indistinguishable from what it models. He considers the example of a computer simulation of a hurricane, which responds to descriptions of meteorological conditions with descriptions of subsequent hurricane behavior. He compares to this a computer program which responds to descriptions of painful occurrences which happen to a person with descriptions of the pain that person has. He goes on to say that what is needed is not a simulation of pain but the actual synthesis of pain in a robot. In other words, the robot must actually be in pain, not merely simulate

or describe pain behavior. Much of the rest of the paper is devoted to a discussion of this possibility.

Rather than claiming at the outset that he can show how to design a pain-feeling robot, Dennett says he will attempt to describe such a robot in terms of current physiological theory relating to pain, and then see how close this description is to a description of a pain-feeling robot. He proceeds to draw a flowchart making use of notions from cognitive psychology, and then discusses whether any system realizing this flowchart would be in pain. While I am not so much concerned with the physiological details of his account, I do wish to examine to what extent Dennett has succeeded in giving an account of pain.

Dennett's flowchart looks like this:<sup>4</sup>

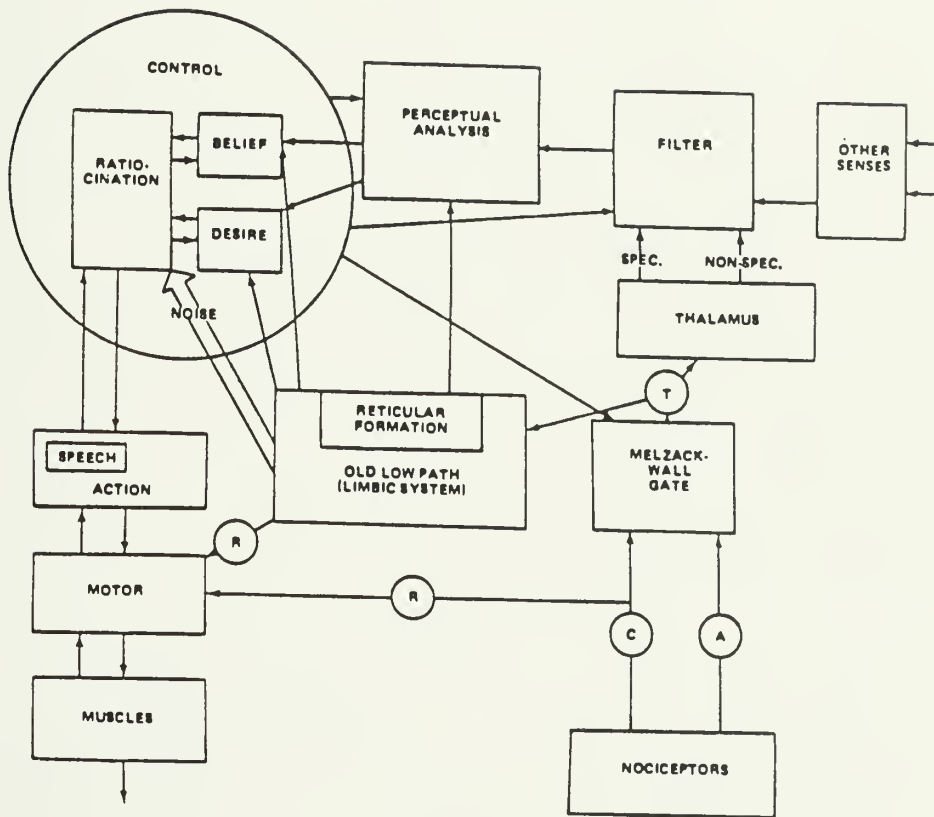


Figure 4-1. Dennett's Flowchart of Pain

Here the journey of physiological response begins at the skin with the nociceptors. The outputs from these travel through either the A-fibers or the C-fibers. At the Melzack-Wall gate, according to the Melzack-Wall theory,<sup>5</sup> the A-fibers act to inhibit the effect of the C-fibers. The T-cell transmissions which are the output of



this gate go either to the older portion of the brain, the limbic system, or through the thalamus to the neocortex, the phylogenetically newer part of the brain.

At this point, Dennett says that "we must abandon the pretence that our boxes represent anatomically salient structures, and proceed with a purely functional, abstract breakdown of the system".<sup>6</sup> He then draws in a filtering system in the "new high" path to account for the fact that distracting one's attention diminishes pain. The purpose of the belief box is to account for the fact that "when we have a pain we believe we have a pain (at least normally)".<sup>7</sup> The belief box is really a memory and belief box, since "pains can be remembered, usually, for some time".<sup>8</sup> The "goals and desires" box, labeled 'desire', is there in virtue of the fact that pains are a reliable form of punishment. The ratiocination box reflects the "oft-used myth...that higher controls in human beings is accomplished by something like logical processing of the material in the belief and desire boxes".<sup>9</sup> Finally, the purpose of the boxes underneath the control center is to reflect the control center's effects on behavior.

After discussing ways in which the above model can account for various facts about pain, Dennett makes the following remarks:<sup>10</sup>

A related, but somewhat different question is this: can we locate pain, as distinct from its typical causes and effects, on the flow chart? The flow chart gives us a functional description at what I have called the sub-personal level. I have labelled the various boxes "belief", "desire", "action" and so forth, but that was taking a liberty. The flow chart deals directly not with a person's acts, beliefs, thoughts, feelings, but with the

behind-the-scenes machinery that governs speech and the like. It has been convenient to talk as if the objects of our attention, what we pay attention to, were impulse trains in the nervous system, to talk as if the muffled outputs from the filter were the diminished pains, to talk as if we recognize or fail to recognize a neural signal as a pain, but this is loose talk, and the conceptual confusions it invites are not inconsequential. When we retell the subpersonal story without taking these liberties we seem to be leaving something out.

Since Dennett has given us what is supposed to be a flowchart description of pain, it is important to know either where the place for pain is on the flowchart or what the relation is between pain and the different phenomena on the flowchart.

Dennett considers the possibility of adding a "pain-box"<sup>11</sup>:

The silence of the sub-personal account here is due simply to the fact that pain itself does not appear on our flowchart, which seems to concern itself solely with the causes and effects of pain. As we trace through the chart we find the causal contributions include nociceptor and C-fiber stimulation, T-cell activity, the process of perceptual analysis and the contributions thereto of old low path activity; and among the effects we find muscle contraction, avoidance reactions and...powerful goal modifications. The absence of a "pain" box might seem to be a simple omission, easily corrected. The most plausible place to insert a pain box is between the perceptual analysis box and the higher control centers. Isn't pain the result of perceptual analysis and the cause of our reactions to discomfort? Let us call the inserted box the pain center. Now what does it do? If one claims its function is simply to serve as the locus for the transmissions just mentioned, the go-between, then contrary to our suspicion, pain was already represented in our model; we simply had not drawn a line around it. If the point is rather that there is a separable and terrible something we had hitherto left out how could possibly add it with this box?

Here Dennett suggests that it would be pointless to add a "pain-box"

- either the pain was already there or the addition of such a box would be pointless.

Dennett concludes that:<sup>12</sup>

There is no way of adding a pain center to the sub-personal level without committing flagrant category mistakes, by confusing the personal and sub-personal levels of explanation.

He then suggests:<sup>13</sup>

Let us consider instead the hypothesis suggested by it, viz., that we have simply not sent the woods for the trees, that pain is not to be found in any one box of our flowchart, but is a function or combination somehow of the elements already present.

This appears to be the position Dennett wishes to adopt, although this statement needs qualification. For he goes on to argue that our intuitions about pain are inconsistent and hence that no living creature or robot could instantiate "the true theory" of pain.

However, then he says this:<sup>14</sup>

If and when a good physiological sub-personal theory of pain is developed, a robot could in principle be constructed to instantiate it. Such advances in science would probably bring in their train wide-scale changes in what we found intuitive about pain, so that the charge that our robot only suffered what we artificially called pain would lose its persuasiveness. In the meantime, (if there were a cultural lag) thoughtful people would refrain from kicking such a robot.

Here Dennett endorses the view that a robot could instantiate a sub-personal theory of pain - in other words, that a robot could feel pain. The plausibility of Dennett's conclusion depends on whether he has succeeded in showing that our intuitions about pain are inconsistent, and in whether it is plausible to maintain that pain is a function of a sub-personal flowchart. I now discuss these two claims.

#### 4.2.1 The Personal/Sub-personal Distinction

Because the distinction between the personal and sub-personal levels of explanation is important to understanding Dennett's analysis of pain, I begin by discussing this distinction. In his account of consciousness, Dennett describes the sub-personal level:<sup>15</sup>

Functionalism theories are theories of what I have called the sub-personal level. Sub-personal theories proceed by analyzing a person into an organization of subsystems (organs, routines, nerves, faculties, components--even atoms) and attempting to explain the behavior of the whole person as the outcome of the interaction of these subsystems. Thus in the present instance the shortcoming emerged because the two access notions introduced, computational access simpliciter and the computational access of a print-out faculty, were defined at the sub-personal level; if introduced into a psychological theory they would characterize relations not between a person and a body, or a person and a state of affairs or a person and anything at all, but rather, at best, relations between parts of persons (or their bodies) and other things. So far as I can see, however, every cognitivist theory currently defended or envisaged, functionalist or not, is a theory of the sub-personal level. It is not at all clear to me, indeed, how a psychological theory--as distinct from a philosophical theory--could fail to be a sub-personal theory. So the functionalists' problem of capturing the person as subject of experience must arise as well for these cognitivist theories. At best a sub-personal theory will see to give us no grounds for believing its instantiations would be subjects of experience, and at worst (as we have seen) a sub-personal theory will seem to permit instantiations that obviously are not subjects of experience. Take your favorite inchoate cognitivist theory and imagine it completed and improved along the lines of its infancy; is it not always easy to imagine the completed theory instantiated or "realized" by an entity--an engineer's contraption, for instance, or some kind of inner zombie--to which we have no inclination to grant an inner, conscious life?

Intuition then, proclaims that any sub-personal theory must leave out something vital, something unobtainable moreover with subpersonal resources. Intuitions can

sometimes be appeased or made to go away, however, and that is the task I set myself here. I propose to construct a full-fledged "I" out of sub-personal parts by exploiting the sub-personal notions of access already introduced.

The personal/sub-personal distinction, then, is a distinction between two levels of description, as well as between two levels of explanation. We can describe some event at the sub-personal level, e.g. we can say "Neuron N fired at time t" independent of giving an explanation of anything. Or, this statement could also serve as an explanation of why some other brain-event occurred. The personal level involves descriptions and explanations of a person while the sub-personal level involves descriptions and explanations of parts of person and their bodies.

How can we characterize this distinction more precisely? Is every description or explanation of a person at the personal level? It seems reasonable to say that psychological ascriptions such as "I feel tired" or "He is angry" are at the personal level, but how about non-psychological predicates such as "I am 5 feet 5 inches tall" or "He weighs 148 pounds"? There are also uses of the possessive pronouns which imply personhood--thus, "My chair is red" is equivalent to "I have a red chair". Even statements about seemingly sub-personal elements can carry such a personal implication. Thus "My C-fibers are conducting impulses" is equivalent to "I have C-fibers which are conducting impulses". Presumably, an equivalent statement without the personal reference, e.g. "C-fibers A, B, and C are transmitting impulses", would be at



the sub-personal level. Finally, Dennett writes as if talk about pains is at the personal level, while talk about the mechanisms of pain is sub-personal. But he also writes as if the sub-personal level is the level of components and parts. And a person can talk about pain as an element or component of his mind or of himself without referring to himself. For example, someone might say "This feeling of pain is intense". How, then, shall we draw Dennett's distinction?

One difficulty with drawing Dennett's distinction is that it seems to cut across the distinctions between the person/body and the person/elements of a person distinction and the mental/physical distinction. In Content and Consciousness, Dennett contrasts "the explanatory level of people and their sensations with the sub-personal level of brains and events in the nervous system":<sup>16</sup>

But when we abandon the personal level in a very real sense we abandon the subject matter of pains as well. When we abandon mental process talk for physical process talk we cannot say that the mental process analysis of pain is wrong, for our alternative analysis cannot be an analysis of pain at all, but rather of something else--the motions of human bodies or the organization of the the nervous system.

Here he seems to equate the personal/sub-personal distinction with the mental-physical distinction.

In Brainstorms, he says:<sup>17</sup>

Sub-personal theories proceed by analyzing a person into an organization of subsystems (organs, routines, nerves, faculties, components--even atoms) and attempting to explain the behavior of the whose person as the outcome of the interactions of these subsystems.



He then says that a theory at the sub-personal level "would characterize relations not between a person and a body, or a person and a state of affairs or a person and anything at all, but rather, at best, relations between parts of persons (or their bodies) and other things".<sup>18</sup> A reference to the faculty of sight would ordinarily be taken to refer to something mental, not something physical, but it also refers to an element of a person, or subsystem. Hence it would seem that even statements about certain mental phenomena can be at the sub-personal level, provided they are about subsystems which are or purport to be elements of a person.

Dennett seems to be primarily interested in distinguishing statements about persons from statements about their bodies, parts of their bodies, or parts or elements of persons. Hence we can say that a statement is personal, or at the personal level, if it contains a reference, implicit or explicit, to a person. An example of an explicit reference is "I am tired" or "Jones is tired", where 'Jones' refers to some person. An example of an implicit reference is "My arms are tired", or "Jones' face is white". A statement is sub-personal if it contains an explicit or implicit reference to some person's body, to a part or state of his body, or to some element of a set of elements of a person. Statements about physical objects which are not bodies of persons will neither be personal nor sub-personal--we might call them a-personal. Examples of sub-personal statements are "Neuron N is firing", "That finger is white", and "The heart is beating". "Jones' neurons are firing" and

"His heart is beating" are both personal and sub-personal in virtue of the implicit reference to a person and the explicit reference to a bodily part or subsystem. Hence, these terms cannot be regarded as mutually exclusive as applied to statements when viewed in this way.

#### 4.2.2 Dennett's Account of Pain

Having briefly discussed Dennett's personal/subpersonal distinction, I turn to a consideration of Dennett's account of pain. Before criticizing this account at a more general level, I wish to make some specific comments about Dennett's flowchart. One comment is that if Dennett's flowchart does give an account of any kind of pain, it does so only of physical pain, and not of psychological and emotional pain. Psychological and emotional pain are closer to being emotions than sensations and are not so directly connected to specific neuronal events. The pain of an insult, the pain of rejection, and the pain of knowing that you have been betrayed by a friend are all instances of psychological and emotional pain which are not (physical) sensations but rather feelings of emotional hurt. Hence, this "cognitive" kind of pain should be analyzed similarly to the way emotions are analyzed. While I do not wish here to defend the claim that one could not give an analysis of emotional states like the one Dennett has given of physical pain, I believe that such an analysis would look quite different. Hence, I believe I am at least justified in claiming

that Dennett's account will not do for emotional or psychological pain. Perhaps this would not bother him, but it is significant that the same term 'pain' is used in both contexts.

Secondly, the stimulation of the A and C fibers is not correlated with pain. Peter Milner reports that:<sup>19</sup>

During innocuous stimulation of the skin, both A and C afferents are stimulated and produce a net excitatory effect in cells of the substantia gelatinosa. These cells depolarize the terminals of both A and C fibers on the transmission cells of the dorsal horn, producing presynaptic inhibition and preventing those cells from being fired too vigorously. If a noxious stimulus is presented, however, though few extra A fibers will be recruited, many extra C fibers will fire, thus increasing the inhibition of substantia-gelatinosa cells and reducing the inhibition of the input to the dorsal horn cells, which will therefore fire at a much higher rate.

While the occurrence of pain is correlated with increased C and A fiber activity, the stimulation of these fibers is correlated with touch and warmth, not just pain. Furthermore, Dennett's "nociceptors" are not exclusively pain receptors. As Kenneth Casey says:<sup>20</sup>

The contention of von Frey in 1895, that free nerve endings are "pain receptors" and that excitation of these endings or receptors resulted in pain sensation gained widespread acceptance. Subsequent physiological and psychophysiological experiments implicated finely myelinated and unmyelinated fibers as subserving "epicritic" and "protopathic" pain mechanisms. However, current anatomical and psychological evidence indicates that free nerve endings must function as receptors for many other forms of somatic sensation and are not exclusively "pain receptors". Furthermore, recent physiological studies have shown that the small diameter afferents do not, as a group, respond only to intense stimuli. Indeed, the majority of those studied are quite sensitive to thermal or mechanical stimuli, and many are sensitive to both forms of energy.

Finally, Donald Becker tells us that:<sup>21</sup>

Since pain is such a ubiquitous phenomenon, this relative lack of very high threshold units suggests that, although some receptors may have significantly higher thresholds than others, the concept of specific nociceptor receptors and fibers can be misleading. Pain is not really a modality like touch, temperature, vision or audition. Rather, pain should be defined as an intensity phenomenon. Excesses of any of the specific modalities, i.e. temperature extremes or excessive pressure, may result in pain perception. As Noordenbos has suggested, pain is "too much".

Hence, stimulation of the "nociceptors" is not exclusively correlated with pain, since receptors which are responsive to pain stimuli respond to other stimuli as well.

Dennett does not explain what he means by the "instantiation" of the flowchart, but presumably, in order for a pain to occur, a system must not merely contain the "structures" represented by the flowchart; rather, the C-fibers must be conducting impulses, etc. But if the nociceptors and the A- and C-fibers are correlated with phenomena other than pain, such as touch, then the flowchart can be instantiated even though not pain, but some other phenomenon is occurring. Hence, Dennett's flowchart does not distinguish pain from certain other mental phenomena such as touch and temperature sensations.

Dennett's flowchart also does not account for all cases of pain. Much physical pain originates elsewhere than in the receptors in the skin. People have headaches, and also pain which is internal to the body. Hence, Dennett's flowchart represents only a restricted number of cases of pain. Dennett would probably

acknowledge this, but reply that his flowchart is merely a sketch, and may need to be revised and supplemented. However, if Dennett is seriously trying to give an account of pain, it is significant to observe how close he comes to achieving this general goal, rather than merely capturing a restricted set of cases.

Finally, Dennett uses terms such as 'thalamus' and 'nociceptor' in his flowchart which refer to organic structures. While Dennett may feel that these entities can be functionally defined, independent of their structures, nevertheless these terms do refer to biochemical parts of the body. And it is not clear that functionally analogous entities in a mechanical, non-organic creature would have the same names. For example, something in the center of a robot which serves as the power source of the robot but does not pump blood but rather sends electrical impulses throughout the robot would not be called a heart. Likewise, no fluid in a non-living thing would be called blood, even if it were red (or blue) and it circulated throughout the body. Georges Rey makes a similar point when he says:<sup>22</sup>

One wouldn't suppose, that there might be machines that engaged in digestion, but of a form quite unlike any such processes that take place in humans and animals. If, for example, the process they underwent for production of their energy was something more like the processes of an internal combustion engine, then surely that wouldn't count as digestion (cars don't digest gasoline). What gets called digestion depends on its resemblances, in natural respects, to what gets called digestion around here.

The general point is that certain bodily parts and functions are essentially parts and functions of living, organic beings, and hence

that non-living mechanical robots could not have the same parts and functions, even if their corresponding parts and functions were analogous in some way to those of a living, organic being.

Dennett would perhaps deny that it is essential to any part of the human body that it be part of the body of a living thing; probably he would claim that all these parts can be functionally defined independent of origin and structure. And it does not seem necessarily wrong to refer to an artificial arm, or even the mechanical arm of a robot, as an arm. Perhaps this is not implausible in virtue of the fact that those functions in terms of which 'arm' is defined, such as grasping and picking up objects, can be performed by a non-living machine. But terms such as 'heart' and 'blood' are essentially connected with life in virtue of their functions of helping to carry food to the organism, and in virtue of other life-sustaining functions. And some of the terms on Dennett's flowchart such as 'nociceptor', are essentially connected with an organism's ability to feel sensation and pain. If these receptors did not play a role in feeling and sensation, they would not be nociceptors. Hence, it seems that the issue of whether an organism can feel pain and sensation needs to be settled before we draw our flowchart, and cannot be settled in advance by stipulation.

It is doubtful, then, that a robot could be said to instantiate Dennett's flowchart, containing as it does terms such as 'nociceptor', 'thalamus', and 'speech'. Perhaps Dennett could construct a more general flowchart which could apply both to men and



machines by abstracting features which are common to both. He might use as a model, Roger Schank's attempt to construct a "language-free" representation of meaning by attempting to reduce the verbs of English down to fourteen primitives.<sup>23</sup> The difficulty with this is in preserving common elements without losing any elements necessary to insure that pain is really occurring.

#### 4.2.3 Perceptual Analysis

Let us consider another aspect of Dennett's flowchart. One box which he says very little about is the box labelled "perceptual analysis". This box occurs after the "filter" box which represents our ability to diminish or banish pain by distracting our attention and before the "control" circle containing the "belief", "desire", and "ratiocination" boxes. The following comments by Hubert Dreyfus suggest that we evaluate the language of the flowchart carefully. Having distinguished between the neurological and phenomenological levels of description, he comments on an effort by Ulric Neisser to define an information-processing level of description:<sup>24</sup>

As Neisser puts it, "the sensory input is not the page itself; it is a pattern of light rays..." So far so good, but then, Neisser goes on to bring the physical and the phenomenological levels together: "suitably focussed by the lens...the rays fall on the sensitive retina, where they initiate the neural processes that eventually lead to seeing and reading and remembering. Here, however, things are by no means obvious. There are two senses of "lead to". Light waves falling on the retina eventually lead to physical and chemical processes in the brain, but in this sequential sense, the light rays and neural processes can never eventually lead to seeing. Seeing is not a chemical process; thus it is not a final step in a series of such processes. If, on the other hand, "lead to" means

"necessary and sufficient condition for", then, either seeing is the whole chain or something totally different from the chain or any link of it. In either case it is no longer clear why Neisser says we have no immediate access to the perceptual world.

Dreyfus' point here is that seeing, and hence perception in general, is not the final step in a series of chemical and neural processes -- rather, it occurs on the phenomenological level, which corresponds to Dennett's personal level. The relevant point here is that the "perceptual analysis" box cannot serve as a connecting link that begins at the neural level with the stimulation of the nociceptors. Dennett seems to acknowledge this in the following remarks:<sup>25</sup>

Can we locate pain, as distinct from its typical causes and effects, on the flowchart? The flowchart gives us a functional description at what I have called the sub-personal level. I have labelled the various boxes "belief", "desire", "action" and so forth, but that was taking a liberty. The flowchart deals directly not with a person's acts, beliefs, thoughts, feelings, but with the behind-the-scenes machinery that governs speech dispositions, motor subroutines, information storage and retrieval, and the like.

Hence the perceptual analysis box really refers not to the process of perceptual analysis but to the physiological mechanisms which govern perceptual analysis. For this reason, the flowchart is at the physiological level, not the personal or phenomenological level.

The "perceptual analysis" box is significant because, in order for input to be experienced as pain it must be interpreted or analyzed as pain. This does not refer to a cognitive process in the case of pain; rather, it refers to the way in which the sensation is perceived by the person. In order for a sensation to be perceived

as pain, rather than, for example, mere touch, it must be perceived as hurting. A sensation becomes painful just at that point when it begins to hurt. Dennett does not deny this; he merely denies that it is useful. He concludes that:<sup>26</sup>

one does not distinguish the sensations that hurt or are painful by applying some criterion; one simply distinguishes them. Their only distinguishing characteristic is painfulness, an unanalysable quality that can only be defined circularly.

However, in my view it is useful to point out that pains are sensations which hurt since then we have at least performed some analysis; now it is easier to understand why we need a "perceptual analysis" box in the flowchart. For some sensations feel good, some pleasant, and some neither good nor bad, pleasant nor unpleasant. Those and only those sensations which hurt are correctly called pain.

Unfortunately, the preceding discussion presents a difficulty for Dennett's flowchart. For if his flowchart, as he says it does, represents the physiological mechanisms of perceptual analysis, and not that analysis itself, then only the physiological mechanisms accompanying an analysis of a sensation as hurting, in other words, as pain, can properly be said to instantiate the flowchart. But it is unlikely that we can identify any physiological mechanisms which always and only accompany pain perception -- we have already seen that the trusted philosophers' friend "C-fiber stimulation" will not do.

At this point, one might wonder to what extent Dennett has succeeded in giving an account of pain, in telling us what pain is,

as opposed to merely giving us the conditions under which it occurs. Dennett concedes that the flowchart "seems to concern itself solely with the causes and and effects of pain".<sup>27</sup> He seems to want to reject this claim, but it is not clear what his basis for the rejection is. After rejecting a homuncular hypothesis, he suggests approvingly that we do the following:<sup>28</sup>

Let us consider instead the hypothesis suggested by it, viz., that we have simply not seen the woods for the trees, that pain is not to be found in any one box of our flowchart, but is a function or combination somehow of the elements already present.

To say that pain is function or combination of the elements in the flowchart can, I believe, mean only one of two things:

1. To be in pain just is to instantiate the flowchart.
2. Instantiating the flowchart is causally or conditionally relevant to feeling pain.

I now consider these two alternatives in turn.

What Dennett has really done is to give us a flowchart of the physiological and neurological mechanisms of pain. But is this the same thing as telling us what pain is? Someone could give a detailed flowchart of the mechanisms of sound production, say by diagramming a stereo system, but would being a sound just be equivalent to instantiating this flowchart? What one wants to say is that a stereo system produces sound, so that instantiating Dennett's flowchart is being what produces pain, but is not being sound itself. Likewise, one might say that instantiating Dennett's flowchart is being what produces pain, but is not being pain

itself. In general, being a mechanism of  $x$  is not being  $x$ , but rather being what produces  $x$ .

This leads to the second alternative. It would perhaps not be correct to call Dennett's diagram a flowchart of the cases of pain, since these are usually events in the world like finger-pricks and instances of being burned. If we are asked for the cause of our pain we generally point to some external event such as being nicked while shaving or having our finger crushed. We don't generally cite the physiological causes, which might better be called physiological conditions of pain. Nevertheless, these physiological conditions are causally relevant to feeling pain.

If Dennett's flowchart is a diagram of the causally relevant conditions of pain, we might compare pain to the hum of a well-tuned engine -- it is a noise produced by the functioning car as a whole, it isn't identical to or produced by any part of the car, and it is not a mental epiphenomenon. Dennett's flowchart would then be like a diagram of a car which, when running, produces a hum. Likewise, when neural impulses are conducted in the brain, pain is produced. The difficulty with this solution is that is the flowchart is merely conceived as giving the causally relevant conditions of pain, then ontologically we are back to the old issue of physicalism vs. dualism -- dualists don't deny that pain may have physical causes; what they deny is that it is essentially a physical phenomenon. Pain, they argue, unlike the hum of an engine (depending on what ontological status assigns to sounds), is essentially mental, and

mental phenomena cannot be "reduced" to or shown to be identical to physical phenomena. Hence it is difficult to see how Dennett's account cuts any ice ontologically, if construed in this way.

#### 4.3 Are Our Intuitions About Pain Inconsistent?

In the preceding I argued that Dennett's functionalist account of pain really tells us very little about the ontological status of pain. Furthermore, I suggested that the possibility of a robot instantiation of a flow-chart of pain might at least run counter to intuitions we have about the application of terms that refer essentially to living organisms, rather than to mechanical devices. However, Dennett does make some explicit comments about the ontological issue, and he acknowledges that the idea of robot pain is counter to some of our intuitions. He attributes this to an incoherency in the ordinary concept of pain:<sup>29</sup>

What then is this conclusion? It is that any robot instantiation of any theory of pain will be vulnerable to powerful objections that appeal to well-entrenched intuitions about the nature of pain, but reliance on such skeptical arguments would be short-sighted, for the inability of a robot model to satisfy all our intuitive demands may be due not to any irredeemable mysteriousness about the phenomenon of pain, but to irredeemable incoherency in our ordinary concept of pain.

Dennett says that our intuitions about pain are inconsistent:<sup>30</sup>

If, as I have claimed, the intuitions we would have to honor were we to honor them all do not form a consistent set, there can be no true theory of pain, and so no computer or robot could instantiate the true theory of pain, which it would have to do to feel real pain. Human



beings and animals could no more instantiate the true theory of pain (there being none), which lands us with the outrageous conclusion that no one ever feels pain. But of course we do.

These remarks seem consistent with his position in Content and Consciousness that 'pain' does not refer.<sup>31</sup> If our intuitions about pain are inconsistent, then there is no such thing as pain, at least in the sense of something which satisfies all our intuitions about pain. I now examine the basis for Dennett's claim that our intuitions about pain are inconsistent.

Dennett gives several examples of what he considers to be inconsistent intuitions. One example is the following:<sup>32</sup>

1. Pains are essentially items of immediate experiences of consciousness: the subject's access to pain is privileged or infallible or incorrigible.
2. Pains are essentially abhorrent or awful--"Pain is perfect misery, the worst of evils..."

Efforts to capture both of these "essential" features in a theory of pain are bound to fail; theories that contrive to maintain both of these claims do so only at the expense of equally well-entrenched claims from other quarters.

Dennett proposes capturing "at least part of what is compelling about 1" by this condition:<sup>33</sup>

3. It is a necessary condition of pain that we are "incorrigible" about pain; i.e. if you believe you are in pain, your belief is true; you are in pain.

He then says:

from (3) and

4. It is a necessary condition of pain that pains are awful

it follows that believing one is in pain is a sufficient condition for really experiencing or undergoing (and not merely believing one is experiencing or undergoing)

something awful. But the belief itself is not the pain, and it is not awful. Surely it is logically possible to be in a dispositional state bearing all the usual earmarks of belief that one is in pain, and yet not be experiencing or undergoing something awful. Not only is this logically possible, it is instanced routinely by morphine subjects and others.

Dennett need not cite examples in which drugs are used to provide examples in which people believe they are in pain but are not experiencing something awful. Ordinary cases of pain experienced during intense physical exercise, or even of minor pains such as pin-pricks or minor headaches are examples of such cases. A jogger might feel pain in his legs from running but not mind it because he realizes it is necessary to achieve his goal of improved health. And minor pains may sometimes be unpleasant, but they are not awful. But not all persons find pain either unpleasant or awful, as the example of the masochist shows.

Unfortunately for Dennett's argument, however, these examples are equally counterexamples to Dennett's condition 2 - that pain is essentially awful. It is curious that Dennett should offer this as an essential condition after he has discussed the example of morphine. Of course, Dennett may claim that according to "our" intuitions, pain is essentially awful, but it is no argument for the incoherence of a concept to cite a view that someone might hold unreflectively. Someone who has never seen a Manx cat might hold that it is essential to cats that they have tails, but we would not for that reason conclude that the concept of cat is confused, or that there are no cats. Only if they refused to apply the term

'cat' when faced with this possibility would we have a basis for believing that tails are essential. But descriptions of morphine-accompanied pain unhesitatingly ascribe pain to the subject:<sup>34</sup>

It's not so much the common idea that the pain becomes resistant to the drug. But the drug produces a variety of different kinds of effects than what we think of as analgesia. For instance, when you talk about morphine to people, they really don't have any concept of what morphine does, except that it blocks pain. And the closest experience that most people have to blockaded pain is that they go to the dentist, he shoots some gunk down around the fifth nerve and blocks it and they're anesthetized, and that to them is what analgesia is - they don't feel pain.

Morphine doesn't do that at all, because under the influence of an opiate/analgesic drug, people feel their pain. They just interpret it in a different way. It doesn't bother them, it doesn't worry them. They can go to sleep in pain. They're comfortable about pain.

Dennett has not been sufficiently discriminating in his search for an essential characteristic of pain; awfulness is not a likely candidate.

The above quotation about the effects of morphine suggests a distinction between the intrinsic quality of a painful sensation and a person's interpretation of or feeling about this sensation or quality. If I worry about pain, or find it unpleasant, I have certain feelings about the pain. These feelings are distinct from the qualities of the sensations itself. Because people's reactions to and feelings about pain vary so widely, given cases in which people don't mind pain or even find that they like pain, it is unlikely that we will find an essential feature of pain among people's reactions to or feelings about it. And awfulness is a

feeling about pain - if I characterize pain, or someone's behavior, as awful, I am referring not to some intrinsic property of what is described, but to how I feel about it. If we want to look for essential features, we should look to the intrinsic features of the sensation itself.

Dennett uses the example of awfulness because he is searching for a criterion to distinguish pain from other mental phenomena we may be incorrigible about. If we use as our criterion the intrinsic property that sensations must hurt in order to be painful, then his counterexample does not work. For it is no contradiction to say, "It hurts, but I don't mind it", or even "It hurts, and I like it". If morphine subjects feel pain, then what they feel hurts, even if they don't mind it. And what masochists feel hurts, even if they don't mind it. We can still maintain incorrigibility - if someone believes he is in pain, then something hurts, even if he doesn't mind it. Hence Dennett fails to show here that our intuitions are inconsistent because his example of an intuition is a poor choice and if we replace it with a good choice we cannot generate the counterexample.

Dennett claims to find several other cases of inconsistency in our intuitions.<sup>35</sup> Rather than discuss these cases in detail, I now discuss what the implications would be if he were right. It should not be surprising if not everyone's intuitions are the same about a phenomenon which occurs as often and in as many forms as pain. However, even if a "good physiological sub-personal theory of

pain" were developed, I doubt that if such a theory called for revisions in our concept of pain that the revisions would make the difference between robot pain and no robot pain. For even if we had to drop some intuitions, such as an intuition about whether pains can exist unfelt, or even infallibility about pains, I do not see that this change would be so drastic as to make robot pain possible. One reason robots cannot feel pain is that they don't have the kind of organic structure necessary for this experience, and until we are able to create living, organic robots I do not see the possibility of overcoming this obstacle to robot pain. As Georges Rey says, "this, then, would seem to be the grain of truth in the common reaction that machines can't be persons: they don't have our feelings because they don't possess our relevant physiology".<sup>36</sup> Creating a good sub-personal theory will not make this possible, and neither will changing our intuitions about the existence of unfelt pains or even about the awfulness of pain.

#### 4.4 Robot Emotions

I conclude with a few comments about the possibility of robot emotions. Materialists sometimes prefer talking about pain to talking about more complex mental states because it is easier to give a physiological account of pain than of emotions such as anger or righteous indignation. Not only is it impossible to find physical states which are always correlated with such complex mental

states, but these states often involve making moral judgments. For example, if I get angry at Jones for taking my pen, then my anger involves a judgment that his act was unjustified, and presupposes a complex circumstance involving the existence of moral rules and property rights. Hence it is difficult to see how anger could be reduced to the instantiation of a physiological flowchart.

Dennett acknowledges that even the concept of pain "is inextricably bound up with our ethical intuitions, our senses of suffering, obligation, and evil".<sup>37</sup> However, he concludes that these considerations, even if they make some difference, "do not make enough of a difference".<sup>38</sup> Even if he is right about pain, I think that these considerations make more of a difference with respect to anger and other emotions. The reason is that these emotions are not simply introspectible sensations but involve cognitive judgments that presuppose an understanding of morality, and presuppose that moral considerations apply to the subject of emotions. A full treatment of this issue would involve a discussion of the extent to which robots can be regarded as rational agents, and the extent to which moral rules would apply to such creatures. While I will not engage in that discussion here, it is significant to note that results about pain are not automatically generalizable to emotional states.



#### 4.5 Summary

I now summarize the main points in the preceding sections. After a brief introduction, I presented Dennett's flowchart of pain in section 4.2. In section 4.2.1, I discussed his personal/sub-personal distinction. In section 4.2.2, I considered Dennett's account of pain. In section 4.2.3, I argued that Dennett has merely presented an account of the causes and conditions of pain, and not an account of pain itself. In section 4.3, I discussed Dennett's claim that our intuitions about pain are inconsistent, and talked about what the implications of this would be. Finally, in section 4.4 I suggested that an account of pain might not be generalizable to emotional states such as anger.

## FOOTNOTES

1. See Daniel Dennett, Brainstorms (Cambridge, Mass.: Bradford Books, 1978), p. xix.
2. Ibid.
3. Ibid., pp. 29-30.
4. Ibid., p. 205.
5. See R. Melzack and P. D. Wall, "Psychophysiology of Pain", in H. Yamamura, ed., Anesthesia and Neurophysiology, International Anesthesiology Clinics, VIII, 1 (1970) for a presentation of this theory. The theory was first introduced in 1965.
6. Dennett, p. 202.
7. Ibid., pp. 203-4.
8. Ibid., p. 204.
9. Ibid.
10. Ibid., pp. 216-7.
11. Ibid., pp. 218-9.
12. Ibid., p. 219.
13. Ibid., p. 220.
14. Ibid., pp. 228-9.
15. Ibid., pp. 153-4.
16. Daniel Dennett, Content and Consciousness, (New York, The Humanities Press, 1969), p. 93-4.
17. Dennett, Brainstorms, p. 153.
18. Ibid.
19. Peter M. Milner, Physiological Psychology (Holt, Rinehart, and Winston, 1970), p. 165.

20. Kenneth Casey, "Some Current Views on the Neurophysiology of Pain", in Pain and Suffering-Selected Aspects, Benjamin L. Crue, Jr., ed. (Banner House, 1070), p. 168.
21. Donald Becker, "A Neurophysiological Examination of Central Pain Mechanisms", in Pain and Suffering-Selected Aspects, p. 139.
22. Georges Rey, "Functionalism and the Emotions", in Explaining Emotions, Amelie O. Rorty, ed. (Los Angeles, University of California Press, 1980), p. 192.
23. See Roger C. Schank, "Language and Memory", *Cognitive Science* 4, 1980, pp. 243-84 for an introduction to this project.
24. Hubert Dreyfus, What Computers Can't Do, (New York, Harper and Row, 1979), p. 180.
25. Dennett, Brainstorms, p. 216.
26. Dennett, Content and Consciousness, pp. 91-2.
27. Dennett, Brainstorms, p. 218.
28. Ibid., p. 220.
29. Ibid., p. 228.
30. Ibid.
31. See Content and Consciousness, p. 96, where he says "there are no pains; 'pain' does not refer".
32. Dennett, Brainstorms, p. 226.
33. Ibid., p. 227.
34. Samuel Mines, The Conquest of Pain, (New York, Grosset and Dunlap, 1974), pp. 66-7.
35. Including whether or not pains can exist unfelt; see Brainstorms, pp. 222-26.
36. Georges Rey, "Functionalism and the Emotions", in Explaining Emotions, p. 192.
37. Dennett, Brainstorms, p. 197.
38. Ibid., p. 198.

## C H A P T E R V

### INCORRIGIBILITY AND INTENTIONALITY: TWO CRITERIA FOR MENTAL PHENOMENA

#### 5.1 Introduction

Chapters 2, 3, and 4 discuss dualism, anomalous monism, and functionalism respectively. While each theory has some merit, each also has difficulties which may require that the theory be revised or abandoned. The chief difficulty for dualism is in giving an account of mind-body interaction. The chief difficulty for anomalous monism is that internal tensions within the theory relating to Davidson's conception of physical events and scientific laws have the consequence that anomalous monism applies to the entities of physics and chemistry but not to events in the brain. The most serious difficulty for functionalism is in accounting for the qualities of subjective experience.

The mind-body theories I have discussed so far contain implicit or explicit conceptions about the nature of mental and physical phenomena. What conceptions a mind-body theory contains about the mind-body distinction and about the nature of mental and physical phenomena has important consequences for the the theory. For example, materialists tend to view sensations as paradigmatic of mental phenomena, while Davidson chooses intentionality. Dualists,

on the other hand, emphasize the private, subjective aspects of mental phenomena. And, as became apparent when discussing Davidson, there may be more than one correct conception of what is physical.

In this chapter, I take a look at two criteria for mental phenomena that have been discussed at some length in this century: incorrigibility and intentionality. In Section 5.2 I discuss Richard Rorty's defense of incorrigibility as a criterion for mental phenomena. I conclude that while this criterion may identify a small subset of mental phenomena, it is not satisfactory as a criterion for all mental phenomena. In Section 5.3 I discuss Roderick Chisholm's three criteria for identifying intentional sentences. I conclude that no formulation of a criterion for intentional sentences will succeed in identifying all and only mental phenomena. In Section 5.4 I draw some conclusions from the earlier discussion.

## 5.2 Incorrigibility

The members of one group of criteria for distinguishing mental from physical phenomena that many philosophers have discussed since the time of Descartes have in common the claim that conscious beings have a special way of knowing mental phenomena. This special way of knowing is sometimes characterized as incorrigibility, infallibility, indubitability, privileged access, or certainty.

Rather than surveying all the members of this group, I have selected incorrigibility as a representative member. The reasons for my choice are that incorrigibility is at least as strong a candidate as any of the others and because there are several valuable discussions of this criterion in recent philosophical literature.

One of the articles containing such a discussion was written by Richard Rorty in 1970.<sup>1</sup> In this article, he considers incorrigibility as the mark of the mental. In this section, I will examine Rorty's discussion in this article and also his discussion in his later book entitled Philosophy and the Mirror of Nature.<sup>2</sup>

#### 5.2.1 Mental Events and Mental Features

Rorty begins by distinguishing two types of mental phenomena. He calls these types mental events and mental features. Rorty draws this distinction in the following paragraph:<sup>3</sup>

In settling upon a mark of the mental, it is important to begin by distinguishing between two different notions of what counts as mental. The distinction I have in mind is that between the sort of mental entity that is an event and the sort that is not. In the first class fall, paradigmatically and perhaps solely, thoughts and sensations. By "thoughts" here I mean not beliefs, but occurrent, datable thoughts--e.g., the entity referred to when one says "The thought that p suddenly struck me." By "sensations" here I mean not perceivings--not acquisitions of beliefs--but simply the entities that are reported in such ways as "Then I had a sensation of red" or "Then I had a painful sensation in my leg." These two sorts of entities make up the content of the stream of consciousness--what one finds when one asks "What's going on in me now?" In the second class fall all those mental entities which are not events and which are only dubiously "entities" at all--beliefs, moods, emotions, desires, purposes, intentions, motives, etc., etc. These might better be called "mental features" than "mental



entities." Not only are they not events, but it strikes one as an odd, peculiarly philosophical hypostatization to think of them as particulars of any sort.

Rorty is certainly correct in saying that it is important to draw distinctions within the class of mental phenomena. Furthermore, his selection of thoughts and sensations as paradigmatic of one class and of beliefs, emotions, and intentions (among others) as paradigmatic of the other class is a useful starting point. Rorty characterizes the class of mental events, of which thoughts and sensations are paradigm, as the "sorts of entities which make up the content of the stream of consciousness".<sup>4</sup> I will first examine this class of mental phenomena, beginning with thoughts.

Rorty describes occurrent, datable thoughts as "e.g., the entity referred to when one says 'The thought that p suddenly struck me'".<sup>5</sup> Perhaps he means this type of example. Yesterday I was sitting at my desk when the thought that I am missing a meeting I am scheduled to attend suddenly struck me. What is the entity which I have referred to? My thought has a certain content, which is that I am missing a scheduled meeting. This content is a proposition which has a truth value. Presumably, Rorty is not just referring to the content of my thought but to a sound image which is the thought and which occurs over a brief period of time. While a sound image which is a thought is a phenomenal occurrence, the aspect of this sound image Rorty is pointing to is that it is occurrent rather than dispositional.

Rorty describes sensations in a similar way. If I say "Yesterday I had a headache", I am reporting the past existence of a feeling that hurts viz. a feeling of pain. This is also a phenomenal mental event in that the presence of pain is the appearance of a hurtful sensation within my mind or consciousness. Again, however, it is the fact that a sensation is an occurrence that takes place over time that makes it fall in Rorty's first class of mental entities.

It is the occurrent character of thoughts and sensations that Rorty is apparently trying to characterize when he says these entities are "the contents of the stream of consciousness".<sup>6</sup> He cites "beliefs, moods, emotions, desires, purposes, intentions, motives, etc."<sup>7</sup> as members of another class he calls mental features. What apparently unites mental features is that they are dispositional, not occurrent, mental entities. This has some plausibility since, for example, it can be true of me that I love someone at time t even though I am not thinking about that person at t. However, if I love someone, I will at times have a feeling of love for them. At those times, I have a feeling of love which is an occurrent, datable experience. Hence it appears that just as one can have occurrent thoughts, so one can have occurrent emotional feelings. It seems, then, that even emotions can be occurrent mental entities at certain times.

Desires exhibit a similar feature. It is true of me that I have a desire to teach philosophy even though I am not experiencing

this desire at a particular moment. However, if I am sitting at my desk and I suddenly have a desire for Chinese food, this is an occurrent, datable desire. In this case, I think of Chinese food and have a felt inclination to have some. Occurrent desires, like occurrent emotional feelings, include occurrent thoughts, and also have a phenomenal aspect. However, other conditions which are present make them desires, not merely thoughts.

One way to analyze dispositional mental states is that having a dispositional mental state is having a disposition to have some occurrent mental state or states. For example, being dispositionally angry at someone is having a disposition to have occurrent angry feelings for that person. Having a dispositional belief that p is having a disposition to occasionally have an occurrent thought that p. If this view of dispositional mental states is correct, then Rorty is wrong to use broad categories such as "emotion", "belief", "emotion", and "desire" to characterize mental features. Since many of these mental entities can be either dispositional or occurrent, Rorty should instead try to distinguish when these mental entities are dispositional as opposed to when they are occurrent. Unless he can do this, Rorty has not drawn a meaningful distinction between mental events and mental features.

Rorty seems aware that his distinction between mental events and mental features is not clear-cut:<sup>8</sup>

As such mental features as beliefs and desires become more particular and limited and, thus, approach the status of episodes rather than dispositions, they become more incorrigible....Beliefs and desires about momentary

matters tend to collapse into thoughts, and momentary emotions tend to collapse into sensations. Short-run beliefs, desires, emotions, and intentions are less like predictions of future behavior than like avowals of contemporaneous thoughts or sensations. That is why they are more like episodes than like dispositions.

Here Rorty appeals to a mathematical analogy of something drawing near to a limit to explain how dispositions approach occurrences. The problem with this analogy is that some desires and emotions reach the limit and are occurrent events. In such cases, they must be classified as mental events, not mental features. Hence, while Rorty is pointing to an apparent and non-controversial distinction within the class of mental entities, viz. the distinction between occurrent and dispositional mental entities, he has not succeeded in characterizing this distinction accurately. To do so, he must draw the distinction between occurrent and dispositional mental entities within broad categories such as "belief", "desire", and "emotion".

### 5.2.2 Rorty's Definition of Incorrighibility

Rorty draws his distinction between mental events and mental features because he believes that only what he calls mental events are incorrigible. However, he believes that mental features have near-incorrighibility, so he can account for them in terms of incorrigibility as well. I will now consider his definition of incorrigibility.

Rorty gives the following definition of incorrigibility:<sup>9</sup>

S believes incorrigibly that p at t if and only if

- (i) S believes that p at t
- (ii) There are no accepted procedures by applying which it would be rational to come to believe that not-p, given S's belief that p at t.

Before giving this definition, Rorty makes the following remark:<sup>10</sup>

It is customary to define incorrigibility in terms of the notions of entailment or logical possibility....I wish, however, to eschew reference to logical modalities, both because of general Quinean doubts about the existence of necessities other than "natural" ones and because of a particular difficulty that arises when we try to spell out 'logically possible' in this context.

This particular difficulty involves his interpretation of a myth he attributes to Sellars.<sup>11</sup> According to this myth, thoughts and sensations were originally theoretical entities, postulated as inner states that explained certain sorts of behavior. They were originally proposed as theoretical entities by a man named Jones who, in this myth, invented the concept of mind. However, Jones proposed these sensations and thoughts not as objects of immediate experience but as inferred entities; entities inferred from the behavior they caused. Only after a long training process can others make noninferential reports of their own inner states.

Rorty comments on this myth that the sensations and thoughts have "no common intrinsic features save 'innerness'"<sup>12</sup> and that all the term 'inner' can mean is "beneath the skin".<sup>13</sup> He then says that not until reports of the existence or occurrence of these thoughts and sensations are incorrigible are these entities mental.

He relates his interpretation of Sellars' myth to defining incorrigibility in terms of logical possibility as follows:<sup>14</sup>

Suppose that, in the familiar manner, we try to spell out the force of this term in such sentences as "It is logically impossible that I believe that I am thinking that p, and not be" by the notion "impossible by virtue of the meaning of terms." We shall then arrive at the conclusion that the meaning of the terms 'thinking' and 'thought' is such that it is impossible to have incorrect contemporaneous beliefs about what one is thinking. But now let us recur to Jones, who uses the word 'thought' before people have learned to make introspective reports of their thoughts, much less come to view such reports as incorrigible. Must we say that when Jones first invented the notion of "thought," meaning by it "inner state that can be about X, be true or false..., etc.," he did not mean by the word what we do? Did the meaning of 'thought' change when people came to make noninferential reports of their own thoughts? Did it change when these reports came to be regarded as the last word? Would it change if cerebroscopes came to be regarded as offering better evidence for what someone was thinking than his own introspective reports?

I regard these questions as unanswerable.

The preceding statement by Rorty not only explains why he chose a definition of incorrigibility that does not invoke logical possibility, it also explains his attachment to incorrigibility as a criterion of the mental. Rorty argues in his later book Philosophy and the Mirror of Nature that if a society existed in which neurology is so advanced that the members of the society report their "mental" states by making reports of their neurological states that no descriptive or explanatory power would be lost.<sup>15</sup> In this society, instruments such as cerebroscopes have come to be regarded as offering better evidence for what someone is thinking than his own introspective reports. It is Rorty's view that if these neurological statements are not incorrigible, they are not mental.



Hence, incorrigibility as a mark of the mental is Rorty's ticket to eliminative materialism, a view he has defended elsewhere.<sup>16</sup> In Philosophy and the Mirror of Nature, his statement of eliminative materialism may appear to be more sophisticated, in that it is more cautiously stated and is materialism without identity of mental and physical events. Nevertheless, eliminative materialism is the view Rorty defends.

I will defer commenting on Rorty's discussion in Philosophy and the Mirror of Nature until later in this section. For the moment, I return to his definition of incorrigibility, which I repeat here:<sup>17</sup>

- S believes incorrigibly that p at t if and only if
- (i) S believes that p at t
  - (ii) There are no accepted procedures by applying which it would be rational to come to believe that not-p, given S's belief that p at t

Since Rorty's definition is intended mainly for thoughts and sensation, he must intend that 'p' be replaced by a third-person form of a statement that S makes or at least believes about his thoughts or sensations. Thus, (i) might read:

- (i)' S believes that he is in pain at t.
- (ii)' There are no accepted procedures by applying which it would be rational to come to believe that S is not in pain at t, given S's belief that he is in pain at t.

It does seem that if someone believes that he is in pain, that there are no accepted procedures to show that he is not in pain. It is of course important that he understand what pain is and be sincere in his report, if he has made a report of his feeling of

pain. However, some sensation reports that include more than a report on a felt quality are not incorrigible. Even Rorty's example, "Then I had a painful sensation in my leg", is shown not to be incorrigible by Descartes' example of the person who had an amputated arm but still felt pain where the arm would be. Such a person might say "Then I felt pain in my arm" before being aware that his arm had been amputated, or if he had forgotten this fact. Here, his report that he feels pain is incorrigible, but he has located that pain in the wrong place.

The preceding example shows that even if we are incorrigible about whether we have a sensation, there are other features of sensations such as their location that we are not incorrigible about. Where a sensation is located in a person's body is something he has to learn from experience, just as an infant has to learn that the object in front of him is his hand. And like many other things in life that we must learn, no one gets it right all the time. I recently tried touching another person's toes one at a time with a pencil tip to see if she could tell me which toe the sensation was in. When she identified the wrong toe, she was easily convinced of her mistake by the location of my pencil. She could correctly identify the toe the sensation was in only 50% of the time. Hence the statement "I have a pain in my middle toe" is not incorrigible for the person uttering it. Other similar examples are having a feeling that seems like hunger but is really a result of being sick

(as is discovered several minutes later) and having a feeling that seems on the borderline between being a pain and being an itch.

If Rorty wants to provide a criterion that shows that a person's beliefs about his own sensations are incorrigible, he must write a more restrictive criterion that only allows beliefs about felt qualities of the sensation. Beliefs about the location or cause of a sensation or about what the sensation is a sign of are not incorrigible. Such beliefs go beyond a report of a felt quality and involve making an inference. And inferences are often mistaken.

To summarize my first criticism of Rorty's criterion, I am saying that he may be right that a belief by a person about whether or not he has a particular sensation or that this sensation has a particular felt quality is incorrigible, but that beliefs which require an inference to the cause, location, or significance of the sensation are not incorrigible. I also stated that Rorty must add the conditions that the person is sincere and understands the language in which his belief is formulated. Rorty must, then, add these three conditions to his definition:

- (iii) S understands the language in which his belief that p is formulated.
- (iv) S is sincere in his belief or statement that p.
- (v) p is a non-inferential first-person belief about whether or not S is having a particular occurrent thought or sensation or that the thought or sensation has a particular felt quality.

Rorty acknowledges that two classes of non-mental statements may be believed incorrigibly in the sense he defines:<sup>18</sup>

- a) Statements knowable a priori
- b) Statements about how something looks, seems, or appears to someone

Rorty dismisses a priori statements on the ground that they are not reports -- they are not descriptions of particular states of affairs.<sup>19</sup> Even if one grants this way of dealing with a priori statements, there are non-mental beliefs that p which are descriptions of particular states of affairs such that there are no accepted procedures by applying which it would be rational to believe that not-p. My belief that I have a hand, that the earth is round, and that the earth has existed for at least 100 years are examples of such beliefs. In On Certainty, Wittgenstein says that "about certain empirical propositions no doubt can exist if making judgments is to be possible at all".<sup>20</sup> Wittgenstein believes that the truth of some empirical propositions such as the ones I have just listed is so fundamental to our system of beliefs that no doubt about them is conceivable. And while Rorty is talking about incorrigibility rather than indubitability, if no doubt is possible about an empirical proposition p then there is no accepted procedure by applying which it would be rational to believe that not-p.

Perhaps Rorty would dismiss these examples on the ground that if S asserts "The world is round", his asserting this fact is not part of the evidence for the truth of this proposition. This would mean that the phrase "given S's belief that p at t" in Rorty's definition asserts that S's belief is part of the evidence for the proposition. However, Rorty says this phrase "can be ignored" for

statements knowable a priori, choosing to rule out these statements on the ground that they are not reports. Yet for statements reporting mental events it is essential, since "It serves, roughly, to summarize the fact that present procedures for adjudicating belief claims are such that the fact of S's belief at t that p is at least as strong evidence for p as any imaginable state of affairs could be for not-p".<sup>21</sup> Apparently Rorty believes he does not need to appeal to the condition that someone believes that p be part of the evidence for the truth of p to rule out a priori statements since he believes these statements can be ruled out on the ground that they are not reports.

Rorty also discusses statements about how something appears, looks, or seems to someone. He says of these statements that:<sup>22</sup>

in their primary meaning...they simply mark refusals to commit oneself to making a report of a certain sort. To say that "X looks brown" is, at the least, to express hesitation about saying that X is brown.

Rorty is certainly correct that "seems" and "appears" statements sometimes function as expressions of hesitation. He says they also can be descriptions of particular states of affairs, in which case they are mental. Rorty rules out "seems", "appears", and "looks" statements that are expressions of hesitation on the ground that they are expressions and not reports. While this may deal effectively with one type of "appearance" statement, it has the consequence that certain mental statements are ruled out as well. There are many uses of mental statements which are expressions rather than descriptions. "It hurts", "I love you", and "I feel

angry" are all sometimes used as expressions of feeling, rather than as descriptions. If Rorty rules out expressions and allows only reports, his criterion will not capture mental events which the subject expresses but does not report.

To see more clearly why Rorty's criterion does not capture mental events for which the subject does not make a report, consider his revised statement of his criterion:<sup>23</sup>

We can sum up the results of this strategy by noting that we now have a set of necessary and sufficient conditions for something being a mental event, namely

If there is some person who can have an incorrigible belief in some statement P which is a report on X, then X is a mental event.

According to this criterion, a necessary condition for X to be a mental event is that some person make statement P which is a report on X. It follows that if a person who has pain does not make a report on his pain, but rather suffers in silence, the condition is not met and the event is not a mental event.

Perhaps Rorty would respond by requiring that some person have an incorrigible belief in a possible statement, not an actual statement. He might then phrase his condition as follows:

B. If there is some person who can have an incorrigible belief in some actual or possible statement P which is a report on X, then X is a mental event.

One problem with this reformulated condition involves the phrase "can have an incorrigible belief". Suppose I am driving my car and I feel a momentary twinge of pain in my finger. Because I am driving in downtown Boston in rush hour, I do not think about the



pain but continue to concentrate on driving. Can I have an incorrigible belief that I have a pain in my finger? Perhaps I can if I stop to think about it. Since B is a conditional, the correct phrasing would be "I can have an incorrigible belief about a momentary twinge of pain if I stop to think about it". But then by the time I stopped to think about the pain, the pain is over and I must rely on my memory. And Rorty does not claim that memories of mental events and thoughts are incorrigible. We are often mistaken in our memories of what we have thought and felt in the past. A memory of a very recent event can be inaccurate just as can a memory of a less recent event. Hence in cases of momentary twinges of pain, we must rely on memory to generate an incorrigible belief. And the reliance on memory denies the status of incorrigibility to the actual or possible belief.

Perhaps Rorty would reply that a person's belief that he feels pain is contemporaneous with his experience of pain. If we asked this person whether he believed that he was in pain at the time he felt the twinge of pain, he would probably reply that he did. At issue here is what it is for a person to believe he is in pain; is mere awareness of the pain sufficient or must a statement about his pain occur to him? Perhaps he has an implicit belief that he is in pain even if no statement about the pain occurs to him.

Even if Rorty could make out a conception of belief that meets the objection advanced in the preceding paragraph, his criterion, even if modified to allow possible statements (statement B), will

not account for all instances of pain. It does not, for example, account for the pains of infants who are incapable of making reports on their feelings and can only express them. A new-born infant is not capable of understanding language and hence cannot make a report on his pain or believe a report about it. Since Rorty apparently believes that only the person having a mental event can have incorrigible beliefs about the event, no one can have incorrigible beliefs about an infant's pain. Yet surely the pain an infant has is a mental event.

Rorty's criterion also does not capture the pains that animals have. Like infants animals cannot either report on their pain or believe reports on their pain. Yet it is believed by almost everyone that animals have feelings and can experience pain. Hence there appears to be no reason to deny that animals have mental events. If Rorty wants to deny this, he should certainly give an explanation. Since animals cannot give reports of their mental events, the ability to give and believe such reports as a criterion for mental events does not make the pain of animals a mental event.

### 5.2.3 Mental Features

Rorty concedes that we are not incorrigible about what he calls "mental features": beliefs, moods, emotions, desires, purposes, intentions, motives, etc. However, he says we are almost incorrigible:<sup>24</sup>

But the fact that we are not incorrigible in our reports of mental features as we are about mental events should not blind us to the fact that we are almost incorrigible. The possibility of overriding reports about such features is real, but it is actualized only rarely and with trepidation. We are far less likely to have a report about a mental state, even one that is not an event, overridden than to have a report about something physical overridden. Further, as such mental features as beliefs and desires become more particular and limited and, thus, approach the status of episodes rather than dispositions, they become more incorrigible.

He goes on to defend his criterion of incorrigibility:

The two factors we have just mentioned--the near-incorrigibility of reports of mental features, and their tendency to become strictly incorrigible as they become more particular and limited--account, I believe, for the term 'mental' having been stretched from the paradigm cases of the nonphysical--thoughts and sensations--to such things as beliefs, desires, emotions, and intentions. If I am right in saying that strict incorrigibility is the mark of mental events and if I was right in saying above that it was mental events, as opposed to mental features, which engendered the Cartesian notion of the mental and the physical as separate realms, then it is appropriate that near-incorrigibility should be the basis for widening the realm of the mental. The likeness of near-incorrigibility to strict incorrigibility is the family resemblance that ties the various things called "mental" together and makes it possible to contrast them all with the physical. But the distinctness of near-from strict incorrigibility is what makes it impossible to find any interesting set of necessary and sufficient conditions for mentality.

Here Rorty makes the general statement that we are far less likely to have a report about a mental state than a physical statement overridden. It is true that some physical statements are more easily overridden than some mental statements. However, not all mental statements are less easily overridden than some physical statements. When I stand in my office, I almost always feel more confident that there is a chair in my office than I do about what I

want to do that evening. I often feel uncertain about how I feel towards another person, particularly when I have conflicting feelings. Many people are uncertain of their religious beliefs, or are even uncertain if they have any. Sometimes when I do a good deed, I am unsure whether my motive was generosity or simply a desire to improve my image. All these examples show that what mental features a person has is often far from clear to that person. Often physical statements are more incorrigible than first-person reports of these mental features.

The advent of Freudian psychology and the accompanying acceptance into common vocabulary of terms such as "unconscious belief", "unconscious motive", and "unconscious feeling" provides an additional reason to doubt the correctness of incorrigibility as a criterion for mental features. If I have an unconscious motive for giving someone a gift, I am not incorrigible or nearly incorrigible in my belief that I have it; I do not even believe that I have this motive. Rorty does not say how he would account for unconscious mental features, but near-incorrigibility does not explain their mental status.

Another example of mental events that persons do not have either corrigible or near-correctible beliefs about is dreams. While a dream is occurring, the events in the dream usually seem real to the person having the dream. If I can be said to have beliefs during a dream, my beliefs are that the dream events are actually occurring. Yet when a person wakes up he realizes that "It's only a

dream" and that his beliefs were in error. The beliefs we have about the events that occur in our dreams are among the most corrigible beliefs we have since they are usually recognized to be false when we awaken.

Perhaps Rorty would reply that even if beliefs a person has during a dream are corrigible, beliefs a person has about his dreams when he awakens are incorrigible. This reply will not work for two reasons. It is possible to remember a dream incorrectly. Suppose I have a dream in which I am playing cards with three other people. I initially identify one of my opponents as Jones. As I recollect more detail about my dream, I realize that it was really Smith, and I correct myself. Secondly, most people do not remember all their dreams. Most people remember only a few dreams. I sometimes awaken with the clear impression that I had a dream, but I cannot remember it. On Rorty's criterion, unremembered dreams are not mental events since no one has incorrigible beliefs about them. It will not work to say "If the person remembered his dream, the resulting beliefs would be incorrigible", since we have already seen that it is possible to remember a dream incorrectly.

#### 5.2.4 Conclusion about Rorty's Defense of Incorrigibility

In conclusion, Rorty's attempt to establish incorrigibility as a criterion for mental events and to extend his criterion for mental events to cover all mental entities by using near-incorrigibility as a criterion for mental features fails for these five reasons:

1. It may be that the statement "I am in pain" uttered sincerely by someone who understands English is incorrigible. Even if this is so, only a small subset of the total number of statements about mental entities is incorrigible. Many other statements about mental entities are not even nearly incorrigible since many persons are often uncertain about their feelings and emotions. Dreams also provide an example of mental events about which our beliefs are not nearly incorrigible.
2. Rorty's criterion fails to capture the pains of infants, any pains or other mental entities animals may have, or unconscious mental entities such as unconscious motives and unconscious feelings.
3. At best, only some very cautious non-inferential first-person statements about mental events are incorrigible. For example, a first-person statement that locates pain at some place in the body can be mistaken.
4. Rorty's criterion fails to provide us with a clear contrast with physical entities. Some physical statements are more incorrigible than some mental statements.
5. Rorty's criterion is vague and difficult to apply, since Rorty does not explain how near to being incorrigible a statement must be to qualify as nearly incorrigible.

#### 5.2.5 Rorty's Argument for Materialism

Rorty's primary interest in his article "Incorrigibility as the Mark of the Mental" apparently is to use his conception of incorrigibility as a mark of the mental to state an argument for materialism. He makes the following statement:<sup>25</sup>

It might turn out that there are no entities about which we are incorrigible, nearly or strictly. This discovery would be made if the use of cerebrosopes (or some similar mechanism) led to a practice of overriding reports about mental entities on the basis of knowledge of brain states. If we should, as a result of correlations between neurological and mental states, begin taking a discovery



of a neurological state as better evidence about a subject's mental state than his own report, mental states would lose their incorrigible status and, thus, their status as mental. This possibility is a result of the way in which we defined 'incorrigible belief'. By phrasing our definition in terms of accepted procedures, rather than in terms of the logical impossibility of error, we leave room for the sort of change that would confirm "eliminative" materialism.

Rorty's claim here is that if some mechanism such as a cerebroscope came to be regarded as better evidence for the existence of mental states than a subject's own reports, mental states would lose their incorrigible status and hence their mental status. Clearly the correctness of this argument depends on whether Rorty has succeeded in establishing that incorrigibility is the mark of the mental. I have argued that Rorty has not established this claim. Hence, his argument for materialism, which depends on the claim that incorrigibility is the mark of the mental, fails.

In a recent book,<sup>26</sup> Rorty advances a similar argument. Here he imagines that explorers from earth discover on the other side of the galaxy a group of people who do not have a conception of mental events. While they are physiologically very similar to us, they report brain states instead of mental events such as pain. Instead of reporting pain, for example, the members of this group say "My C-fibers are firing". The explorers call the members of this group the Antipodeans, making a reference to a group of Australian materialists.

Rorty's argument in this example is more difficult to determine than in the earlier article I just discussed since he states his

argument less directly. However, he does say that "in corrigibility is all that is at issue in puzzles about the Antipodeans".<sup>27</sup>

Hence, he appears to be giving a more elaborate version of the argument for materialism than he gives in his first article: the argument I just discussed. Robert Schwartz also advances this interpretation:<sup>28</sup>

[Rorty's] position is that none of our mental-state reports are intrinsically incorrigible, although, at a given time, there may not be settled community standards for challenging such first-person claims. Were neurophysiology to develop sufficiently, no state would be incorrigibly reportable; hence, no state would be on the nonphysical side of the divide. Materialism wins out, but it is materialism without mind/body identity. The realm of the mental is eliminated or ontologically mental.

If Rorty's argument in this book is as Schwartz presents it, it is no better than his defense of incorrigibility as a mark of the mental. I have just shown that Rorty does not succeed in establishing incorrigibility as a mark of the mental in his earlier article. I have also given reasons for believing that this claim cannot be established. Furthermore, Rorty's defense of incorrigibility as the mark of the mental in Philosophy and the Mirror of Nature is mainly historical and is not as systematic or explicit as his defense of this criterion in the earlier article. His discussion in the book does not add additional substantive arguments for this criterion that go beyond those advanced in the earlier article. Hence, to the extent that Schwartz's interpretation is correct, Rorty's argument involving the Antipodeans is susceptible to the same objection as his argument for

materialism in the article. This objection is that Rorty's argument depends on a claim he has not established: the claim that incorrigibility is the mark of the mental.

Rorty does draw a more cautious conclusion in his book:<sup>29</sup>

The materialist should stop reacting to stories such as that about the Antipodeans by saying metaphysical things, and confine himself to such claims as "No predictive or explanatory or descriptive power would be lost if we had spoken Antipodean all our lives."

This claim is reminiscent of claims Rorty makes more explicitly in another article in which he defends the "disappearance" version of the identity theory.<sup>30</sup> In Philosophy and the Mirror of Nature, Rorty does not want to describe his view as being a version of the identity theory, though he regards his position as a form of materialism. In spite of Rorty's attempt to divorce himself from the identity theory, his claim in the above quotation which involves his interpretation of the Antipodean example is susceptible to the same objections as his earlier "disappearance" version of the identity theory.

The main objection to Rorty's claim that no predictive or explanatory power would be lost if we had spoken Antipodean all our lives is that in Antipodean there is no way to describe the subjective qualities of our experiences. Even if one conceded that Antipodeans succeed in reporting the existence of pain they cannot report the subjective qualities of pain. To see why this is so, suppose that we place ourselves in the position of the Antipodean. We can even suppose that science has discovered neural correlates

for qualities of pain such as its burning and intense qualities. Let us call these neural correlates B and I. Imagine a person, call him Matt, connected to a machine which gives him reliable readings of his brain processes. Matt can thus "read off" his state of mind from the machine. Now, let us suppose that such a person makes a report based on the sensation. If he reports "My C-fibers are firing in ways B and I", what is he reporting? He is reporting a brain process which has been empirically discovered to be "correlated" with what we (on earth) call "a painful, intense, burning sensation". But let us suppose we ask him to describe what he is reporting. What can he say? He cannot use terms like 'intense', 'painful', and 'burning' since these are terms used to report mental events. All he can do is to give further details about his brain process, such as its cause, location, etc. In other words, Matt will have no way of describing his experience--he can only report its cause.

Rorty's Antipodean example is an attempt to show that the objective point of view from which we view another person's feelings, thoughts, desires, etc., could replace the subjective point of view from which each person experiences his own feelings, thoughts, desires etc. with no loss of explanatory or descriptive power. Such an attempt is inherently confused since the objective point of view depends for its definition on the subjective point of view. It is also confused in that objective criteria such as "C-fibers firing" are supposed to be criteria for something, and

this something is not another brain state but a subjective experience such as pain. Antipodean is in fact a very impoverished language in which people are unable to describe what their experiences are like. The suggestion that Antipodean is just as descriptive and explanatory as our current language has no more merit than the suggestion that we abandon all talk of our bodies and brains in favor of strictly mentalistic language. In such language, one cannot say "I have a pain in my finger". Likewise, in Antipodean, one cannot say "I have an intense, throbbing pain."

### 5.3 Intentionality

Contemporary discussion of intentionality as a criterion of the mental begins with Franz Brentano.<sup>31</sup>

Every mental phenomenon is characterized by what the scholastics of the Middle Ages called the intentional (and also mental) inexistence (Inexistenz) of an object (Gegenstand), and what we could call, although in not entirely unambiguous terms, the reference to a content, a direction upon an object (by which we are not to understand a reality in this case), or an immanent objectivity. Each one includes something as object within itself, although not always in the same way. In presentation something is presented, in judgment something is affirmed or denied, in love [something is] loved, in hate [something] hated, in desire [something] desired, etc.

This passage on intentionality is sometimes cited by commentators on intentionality as the classic statement of the thesis of intentionality, although Brentano uses the phrase 'intentional inexistence' rather than 'intentionality'.<sup>32</sup>

Brentano states his thesis as an ontological claim about mental phenomena. Brentano is not making a conceptual or linguistic claim about the language for describing mental phenomena; he is saying that mental phenomena such as thoughts and desires have the property of being directed upon an object. Recent discussion, by contrast, centers on criteria by which intentional sentences or intentional terms within language can be identified.

Much recent discussion of intentionality centers around articles by Roderick Chisholm, responses to his articles, and replies to these responses.<sup>33</sup> In particular, Chisholm has proposed three criteria of intentionality. In Section 5.3.1, I state a recent formulation of these criteria. In Section 5.3.2 I examine recent discussion of these criteria by Albert Flores<sup>34</sup> and Ausonio Marras.<sup>35</sup> In Section 5.3.3, I discuss a reformulation by Chisholm of his criteria in response to criticisms by Flores and others. In Section 5.3.4 I examine a reformulation of Chisholm's criteria by Ansonio Marras. In Section 5.3.5 I state three criticisms of intentionality as a criterion of mental phenomena.

### 5.3.1 Chisholm's Criteria for Intentional Sentences

Chisholm originally states his three criteria for intentionality in Sentences About Believing<sup>36</sup> and in Perceiving: A Philosophical Study<sup>37</sup>. He restates these three criteria in a recent discussion:<sup>38</sup>



I had suggested a kind of criterion of the psychological. A phenomenon is psychological, I said, if we cannot adequately describe it in our ordinary language without using sentences that are intentional. And then I formulated three criteria of intentionality. (1) A sentence is intentional if it uses a substantival expression in such a way that neither the sentence nor its negation imply that there is, or that there isn't anything, to which that expression truly applies. (2) A sentence is intentional if it has a subordinate clause such that neither the sentence nor its negation imply that the subordinate clause is true or that it is false. And (3) a sentence is intentional if it has a context that is referentially opaque. Then I added that a compound sentence is intentional provided one of its components is intentional.

Here Chisholm restates his three original criteria for a sentence's being intentional. These three criteria can be formulated as follows, together with a criterion for compound sentences.

A sentence is intentional if:

1. It uses a substantival expression in such a way that neither the sentence nor its negation implies that this expression does or does not apply to any thing.
2. It has a subordinate clause such that neither the sentence nor its negation imply that the subordinate clause is true or false.
3. It has a referentially opaque context.
4. It is a compound sentence and one of its components is intentional.

According to Chisholm, "a phenomena is psychological...if we cannot adequately describe it without using sentences that are intentional".<sup>39</sup> Apparently Chisholm's view is that if we must use an intentional sentence to describe a phenomenon, the phenomenon we are describing is psychological. For example, suppose I say:

5. I believe that Reagan is from California.

Here I am stating a belief that I have. In this example, my belief that Reagan is from California is the psychological phenomenon. It is the presence of the verb 'believe' in sentence 5 that accounts for the intentionality of this sentence, not the presence of the sentence "Reagan is from California". However this sentence expresses the content of my belief, so its presence is essential to the intentionality of the sentence. Yet the psychological phenomenon I am describing when I use this sentence is my belief, not the fact that Reagan is from California.

Because the intentionality of a sentence depends in many cases on the presence of a verb that expresses a propositional attitude, some philosophers such as Davidson prefer to talk about mental verbs:<sup>40</sup>

We may call those verbs mental that express propositional attitudes like believing, intending, desiring, hoping, knowing, perceiving, noticing, remembering, and so on. Such verbs are characterized by the fact that they sometimes feature in sentences with subjects that refer to persons, and are completed by embedded sentences in which the usual rules of substitution appear to break down. This criterion is not precise, since I do not want to include these verbs when they occur in contexts that are fully extensional ('He knows Paris,' 'He perceives the moon' may be cases), nor exclude them whenever they are not followed by embedded sentences. An alternative characterization of the desired class of mental verbs might be that they are psychological verbs as used when they create apparently nonextensional contexts....On the proposed test of the mental, the distinguishing feature of the mental is not that it is private, subjective, or immaterial, but that it exhibits what Brentano called intentionality. Thus intentional actions are clearly included in the realm of the mental along with thoughts, hopes, and regrets (or the events tied to these).

Davidson talks here about mental verbs, e.g., believing, intending, hoping, desiring, knowing, etc. Whether a verb is mental depends on its use; these verbs may not always be mental. According to Davidson, a verb is mental when it occurs in an intentional context. Chisholm's criteria for intentional sentences should enable us to identify intentional contexts.

Apparently Chisholm intends each criterion to be a sufficient criterion for identifying psychological phenomena. The criteria cannot be individually necessary since they apply to different types of sentences. Albert Flores cites these three sentences as examples of intentional sentences by each of Chisholm's three criteria:<sup>41</sup>

Criterion 1: 6. Quine looked for the proper  
characterization of analyticity.

Criterion 2: 7. Othello believes that Desdemona loves  
Casio.

Sentence 7 is also intentional by Criterion 3 in virtue of the truth of the following identity statement:

8. Desdemona is the woman framed by Iago.

Sentences 6 and 7 are each intentional according to the first two of Chisholm's criteria respectively. Sentence 7 is also intentional according to Chisholm's Criterion 3. But sentence 6 is intentional only by Criterion 1. We can add a fourth example of a compound sentence as an example of Criterion 4:

9. Othello believes that Desdemona loves Casio or snow is  
white.

The most correct interpretation of Chisholm's criteria, then, is that the disjunction of the four criteria is a necessary condition

for identifying psychological phenomena, while each condition is individually sufficient. This allows for the possibility that a sentence can be intentional by more than one criterion, but need not be intentional by more than one.

### 5.3.2 Criticisms of Chisholm's Criteria

One criticism that has been raised against Chisholm's intentional criteria for psychological phenomena is that sentences that contain modal terms such as 'possible' and 'necessary' satisfy one or more of Chisholm's criteria yet are not used to describe psychological phenomena. Albert Flores cites the following example:<sup>42</sup>

10. It is necessary that Alaska is Alaska.

We can assume the truth of the following identity statement:

11. Alaska is the 49th state of the union.

Substitution in sentence 10 results in the following statement:

12. It is necessary that Alaska is the 49th state in the union.

Sentence 10 is intentional by Criterion 3, yet it is not used to describe a psychological phenomenon.

Flores cites other sentences which have been cited as counter examples to Chisholm's first two criteria:<sup>43</sup>

13. The fire needs some coal.

14. The frost may bring it about that the cliff will fall.

Sentence 13 is intentional by Criterion 1 since neither 13 nor

15. The fire does not need some coal.

implies that any coal is available, or even exists. Sentence 14 is intentional by Criterion 2 since neither 14 nor its negation:

16. The frost may not bring it about that the cliff will fall.  
implies that the truth or falsity of the subordinate clause.

Chisholm's response to examples like these and others is to claim that the alleged counterexamples can be transformed into equivalent non-intentional sentences:<sup>44</sup>

These sentences are not examples counter to our thesis. Anyone who understands the language can readily transform them into conditionals which are not intentional...Instead of using intentional sentences, we could have said, 'If there should be any new epidemics, the patient would not be affected by them'...I believe that any other ostensibly non-psychological sentence which is intentional can be transformed, in an equally obvious way, into a sentence conforming to our version of Brentano's thesis. That is to say, it will become a sentence of one of two possible types: either (a) it will no longer be intentional or (b) it will be explicitly psychological. Sentences about probability may be intentional, but, depending upon one's conception of probability, they may be transformed either into the first or into the second type.

Following Chisholm's proposal, D.H. Sanford proposes these transformations for sentences 10 and 14:<sup>45</sup>

- 17. It is necessary that Alaska is self-identical.
- 18. It is possible that the frost will bring it about that the cliff will fall.

Sentence 17 has the same truth value as 10, while the negation of 18 implies the falsity of the subordinate clause. Hence sentences 10 and 14 are not intentional after all.

Flores correctly attributes to Chisholm the following principle:<sup>46</sup>

- A. All non-transformable intentional sentences are psychological.

Sentences containing modal terms and other apparently non-psychological yet intentional sentences do not count as intentional if they can be transformed into non-intentional sentences. Or, if they can be transformed into explicitly intentional sentences, they count as intentional. For example, one might transform:

19. It is probable that there is life on Venus.  
into

20. It is reasonable for us to believe that there is life on Venus.

Flores takes 'transforms' to mean "transform into a logically equivalent sentence". He counters that if one sentence is intentional and the other is not, these two sentences have different implicatory relations and hence are not logically equivalent.<sup>47</sup>

### 5.3.3. Chisholm's Reformulated Criterion

In his response to Flores, Chisholm concedes that his original criteria need to be reformulated. He gives the following definitions which are intended to provide this reformulation:<sup>48</sup>

- (D14) T contains S autonomously in L = Df (i) S is a proper part of T; and (ii) for every U, if U results from replacing S in T by some sentence in L, then there is a P such that P is a proper part of U and the result of replacing S by P in T is a sentence in L.



- (D15) T contains S in a propositional context in L = Df  
 (i) S is a proper part of T; (ii) T does not  
 contain S autonymously in L; and (iii) for every U,  
 if U is a sentence in L, the result of replacing S  
 by U in T is a sentence in L.
- (D16) T contains S in an intentional context in L = Df  
 (i) T contains S in a propositional context in L;  
 (ii) there is an S' which is inconsistent with S in  
 L but which is such that the result of replacing S  
 by S' is consistent with T in L.
- (D17) T contains S in an exclusively intentional context  
 in L = Df (i) T contains S in an intentional  
 context in L; and (ii) there is a U such that (a) U  
 is logically equivalent to S in L and (b) the  
 result of replacing S by U in T is not logically  
 implied by T in L.

In his discussion of Chisholm's reply, Ausonio Marras correctly  
 attributes to Chisholm the following criterion:<sup>49</sup>

- (P) A sentence containing a propositional clause S is  
 psychological if and only if it contains S in an  
exclusively intentional context.

Marras goes on to discuss the problem of modal contexts:<sup>50</sup>

According to (P), 'John believes that Socrates is  
 mortal' is psychological, whereas 'It is possible  
 that Socrates is mortal' is not; the latter  
 contains S in an intentional but not in an  
 exclusively intentional context. And 'It is  
 necessary that Socrates is mortal' is also  
 non-psychological since S does not even occur in an  
 intentional context.

As Marras says, then, Chisholm's revised criterion appears to answer  
 objections to Chisholm's earlier criteria based on sentences  
 containing modal terms.

Marras points out, however, that some sentences used to  
 describe psychological phenomena are not intentional on Chisholm's  
 revised criterion:<sup>51</sup>

Let T be the obviously psychological sentence 'John knows that Socrates is mortal' and let L be 'Socrates is mortal'. T contains S in a propositional context but not in an intentional context, and thus not in an exclusively intentional context, since there is an S' (Socrates is not mortal') which is inconsistent with S but which is such that the result of replacing S by S' is not consistent with T. Hence T (and many other sentence containing verbs like 'see', 'notice', 'remember', 'realize', etc.) turn out to be non-psychological according to (P).

Other sentences that are not intentional by Chisholm's criteria are:

20. I see that my desk is brown.
21. I remember that you attended the meeting.
22. I am aware that she is wearing lipstick.

In sentences 20, 21, and 22, the result of replacing the embedded sentence with its negation yields a sentence which is inconsistent with the original sentence. Chisholm's revised criterion, then, while dealing with the problems of modal contexts, appears to rule out sentences which clearly describe psychological phenomena.

#### 5.3.4. Marras' Revision of Chisholm's Criteria

Marras proposes a different revision of Chisholm's original criteria:<sup>52</sup>

- (i) Drop Chisholm's third criterion of intentionality in SAB [Sentences About Believing], which fails to keep the (de dicto) modalities out of the intentional domain.
- (ii) Retain the first two criteria in SAB as (singly) sufficient conditions of intentionality.
- (iii) Add to those two criteria the following recursive condition:

- (C) Any (noncontradictory) sentence which entails a sentence which is intentional by either the first or the second criterion is itself intentional.
- (iv) Let Chisholm's first two criteria in SAB supplemented by the recursive condition (C) constitute jointly a necessary and sufficient condition of intentionality. (It will be noticed that modal sentences turn out to be non-intentional by this condition, whereas 'John knows that p', which entails 'John believes that p', turns out to be intentional by meeting conditions (C).)
- (v) Reformulate Chisholm's thesis about the psychological as follows:
  - (D) A sentence is psychological if and only if it is intentional and is not transformable into a non-intentional sentence.

In a footnote, Marras replies to four sentences offered by John Bacon as counterexamples to his proposal. One sentence is:

23. It is a contingent matter whether A.

In reply, Marras refers to an article published in Intentionality, Mind and Language. In a footnote in this article, Marras makes these remarks:<sup>53</sup>

Contingency sentences seem to be intentional by Chisholm's second criterion. For neither "It is contingent that p" (where 'contingent' means 'neither necessary nor impossible') nor its contradictory implies either that p is true or that p is false. However, since such contingency sentences are really abbreviations for compound sentences (i.e., sentences containing two or more sentences conjoined by means of such sentence connectors as 'and', 'or', 'because,' etc.), they are as such beyond the scope of Chisholm's second criterion, which applies only to noncompound sentences. Chisholm also provides a criterion of intentionality for compound sentences, according to which a compound (declarative) sentence is intentional if and only if at least one of its component sentences is intentional (Perceiving, p. 172). By this criterion, the compound sentence "It is neither necessary

that p nor impossible that p" (which is the unabbreviated form of "It is contingent that p") is not intentional since neither of its component sentences is intentional.

Similar considerations apply to sentences about probability dispositions, tendencies, abilities, etc., which may be analyzed into nonintentional sentences about frequencies and/or nonintentional counterfactual sentences.

In the above quotation, Marras deals deftly with many modal sentences. Included in his list of statements he believes he can show to be non-intentional are sentences of probability. Marras apparently believes that probability sentences can be analyzed into sentences about frequencies or into nonintentional counterfactual sentences. The frequency interpretation is one of several accepted interpretations of probability.<sup>54</sup> However, this interpretation of probability has difficulty accounting for the probability of unique events. Consider this sentence:

24. It is probable that the sun will burn out in 23 billion years.

Sentence 24 is intentional by Chisholm's criterion 2. Perhaps Marras would analyze 24 as:

25. The frequency with which suns burn out in 23 billion years is high.

However, 25 will not do as an analysis of sentence 24 since 24 refers to a particular sun and since this claim is based on evidence about the sun; not on a statistical or frequency analysis of probabilities.

Chisholm addresses this issue in Sentences about Believing:<sup>55</sup> Here Chisholm says that the following sentence

26. It is probable that there is life on Venus.

can be transformed into sentences about frequencies. But unique events occur only once so that it is a mistake to suppose that their probability is a matter of the frequency with which they occur. Chisholm also appears to be confusing giving an analysis or equivalent reading of a sentence with citing the evidence for it. The evidence for sentence 26, even if it is a matter of the frequency with which life is found on certain places, is not part of the meaning of 26. Hence, even if the evidential sentences for sentence 26 are non-intentional, this does not make 26 non-intentional.

Probably Chisholm and Marras would prefer an intentional reading of sentences 25 and 26. Chisholm's other interpretation is as follows:<sup>56</sup>

27. It is reasonable for us to believe that there is life on Venus.

Sentence 27 is intentional by Criterion 2. However, what is reasonable to believe is determined by weighing the reasons for and against a proposition, while probability is a matter of judging the likelihood of an event's occurring. Hence it can be reasonable to believe something to which the concept of probability does not apply, as it is reasonable for me to believe that  $2 + 2 = 4$ . Yet it is not merely probable but certain that  $2 + 2 = 4$ . Furthermore, what is reasonable for a person to believe is relative to the background of his beliefs and purposes, so it can be reasonable to believe something improbable. For example, it is reasonable for me

to believe that I can leap over a chasm if this will help me accomplish this task even if it is improbable that I will succeed.

Perhaps Chisholm or Marras could offer another "analysis" or "transformation" that would not be susceptible to the objections just advanced. Unless this is done, Marras' revised criterion does not fare better than either Chisholm's original criteria or Chisholm's revised criteria.

Marras' criterion is also notable for the psychological sentences that are not intentional according to his criterion. Consider once again Marras' criterion, this time with Chisholm's first two criteria spelled out:

1. It uses a substantival expression in such a way that neither the sentence nor its negation implies that this expression does or does not apply to any thing.
2. It has a subordinate clause such that neither the sentence nor its negation imply that the subordinate clause is true or false.
3. Any (noncontradictory) sentence which entails a sentence which is intentional by either the first or the second criterion is itself intentional.

Condition 2 applies only to sentences that contain a subordinate clause, so any sentence without such a clause, if it is intentional, has to be intentional by either Condition 1 or Condition 3. The sentences which Condition 1 are intended to meet are sentences such as:

- (6) Quine looked for the proper characterization of analyticity.

These sentences contain a "substantival expression" which is used in such a way that neither the sentence nor its negation implies that



there is or is not anything to which the substantival expression applies. By 'substantival expression', Chisholm apparently means "name or definite or indefinite description" to allow for examples such as:

28. Diogenes looked for an honest man.

In Perceiving Chisholm cites sentence 28, which originates with Brentano, as an example of an intentional sentence according to his first criterion.<sup>57</sup> Hence the substantival expression can be either a definite or indefinite description, as well as a name.

In the examples Chisholm gives of sentences that are intentional by Condition 1, the substantival expression is the object of a verb. Chisholm does not to my knowledge address the possibility that the substantival expression referred to in Condition 1 might be the subject of the sentence. Hence, sentences which neither contain a subordinate clause nor contain a substantival expression which is the object of a verb are apparently not covered by either Condition 1 or Condition 2. Examples of such sentences are:

29. I am thinking.

30. He is dreaming.

31. I am in pain.

32. I am angry.

33. She is depressed.

Sentences 29 to 33 are all mental or psychological sentences. None of these sentences contains a subordinate clause, and the only

substantival expression in any of these sentences is the subject of the sentence. I do not believe that Chisholm intended the substantival expression mentioned in Condition 1 to be the subject of the sentence. However, to the extent that any of these sentences implies anything about their subjects, the implication is that the subject exists, viz. that there is something to which the substantival expression truly applies. This is especially true of the first-person sentences, but it applies to the third person sentences as well. But if Chisholm or Marras were to deny that there is such an implication, and hence that these sentences are intentional, it is difficult to see why these sentences would not be intentional as well:

34. I am tall.

35. He is rich.

36. She is tanned.

These three sentences have subjects that function like the subjects of sentences 29 to 33. The difference is that they do not contain intentional terms. Hence Chisholm and Marras cannot account for the intentionality of sentences 29 to 33 by appealing to Condition 1.

Perhaps Marras' Condition 3 will account for the intentionality of sentences 29 to 33. Marras may wish to argue that:

31. I am in pain.

entails

37. I believe I am in pain.

In his Encyclopedia of Philosophy article on "Intentionality", Chisholm gives this very argument in an attempt to account for the intentionality of pain.<sup>58</sup> Chisholm does not incorporate this consideration into his criteria as Marras has. However, it is not clear that sentence 37 is intentional by Marras' criterion. Just as sentence 31 plausibly can be held to imply sentence 37, so 37 plausibly can be held to imply 31. This violates Condition 2, according to which neither a sentence nor its contradictory implies the truth or falsity of the subordinate clause. Hence even if sentence 31 does imply 37, sentence 31 is not intentional and therefore mental by Condition 3.

In the case of sentence 31 implying sentence 37, which seemed at first as if it might provide a way to argue that sentence 31 is intentional, this implication relies on the incorrigibility of first-person pain sentences. Hence, someone who relies on this implication to explain the intentionality of pain sentences is importing a different criterion from intentionality, viz. incorrigibility, and using it to support intentionality as a criterion of mental phenomena. However, I argued in Section 5.2 that incorrigibility does not extend to all our mental states, and that in particular it does not extend to dream states and emotional states. Hence, even if Marras could find a way to show that sentence 31 is intentional, or even that sentence 29 is intentional by Condition 3, his argument would not extend to the other three sentences listed:

30. He is dreaming.

31. I am angry.

32. She is depressed.

In summary, according to Marras' revised criterion of intentionality, some sentences which are not mental such as sentences of probability are intentional by Marras' criterion. Other sentences which are mental, such as sentences 29 to 33, are not intentional by his criterion. Whether or not Marras' criterion is an improvement on Chisholm's, it is not significantly more successful than Chisholm's in capturing those sentences which are generally recognized to be psychological or mental.

#### 5.3.5 Conclusion about Intentionality

I will now make some general comments about intentionality as a criterion of the mental. First, the Chisholm-Marras criteria attempt to identify certain sentences as being intentional and then claim that what these sentences describe or refer to is mental. On the other hand, it is clear from reading Brentano that he regarded intentionality as an ontological feature of mental phenomena, not merely a conceptual or linguistic one. Based on the attempts examined in this section, it is unlikely that any conceptual or linguistic criterion of intentionality will capture all and only those sentences we antecedently call mental. It may be, then, that future research should be directed upon formulating an ontological criterion.

Secondly, it seems unlikely that any criterion, either ontological or conceptual, could succeed in capturing all cases of mental phenomena. If the unique feature of intentional phenomena is that they are directed towards an object, it seems clear that not all mental phenomena share this feature. Both sensations such as pain and tiredness and objectless emotions such as certain cases of depression and fear do not share this feature of intentionality. Also, visual images which do not represent any thing, such as spots before someone's eyes, are not intentional. It appears, then, that many mental phenomena do not have the feature Brentano originally identified, that of being directed towards an object, and which has come to be known as intentionality.

Third, it is not clear that intentionality is exclusively a feature of mental phenomena. As Marion Knowles says:<sup>59</sup>

It has been suggested that there are physical phenomena which, in some sense, are intentional.

For instance, Scruton, following Goodman, points out that one can conceive of a painting as being 'about its object in the way that a sentence is about its subject.' And, clearly, certain paintings are paintings of something (e.g., portraits, landscape paintings), while others are, in a sense, 'about' something (e.g. Picasso's 'Guernica' is, *inter alia*, about war).

Another example is a road sign which points in a particular direction. In these cases, the intentionality of the physical entity is dependent on the intention of a person. Nevertheless, a sign still points in a particular direction even if one has to explain this fact by reference to human intentions. Any complete

account of intentionality will take into account the intentionality that attaches to some physical entities.

#### 5.4 Conclusion

In Section 5.2 I examined Richard Rorty's defense of incorrigibility as a criterion for identifying mental phenomena. In Section 5.3 I examined Roderick Chisholm's criterion of intentionality as a criterion for identifying mental phenomena. I argued that while a small subset of first-person statements may be incorrigible, this criterion fails to capture many other first person and third person statements about mental phenomena. In particular, it fails to capture many statements about what Rorty calls mental features, including statement about feelings, emotions, and dreams. It also fails to capture statements about infant and animal pains. Furthermore, some physical statements are more incorrigible than some mental statements. Hence, incorrigibility does not succeed in capturing all and only mental phenomena.

With respect to intentionality, I argued that the criteria proposed by Chisholm and Marras make non-mental sentences intentional and fail to capture sentences which clearly are mental. It seems unlikely that any linguistic criterion will be successful in capturing all and only mental phenomena. Ontologically speaking, it does not seem that all mental phenomena share the feature of intentionality originally identified by Brentano. In particular,



objectless emotions such as certain cases of depression and fear and certain sensations such as those of pain and tiredness do not have this quality of "aboutness" or intentionality. Finally, certain physical objects appear to have intentionality, or a feature very much like intentionality. Hence, the criterion of intentionality does not succeed in identifying all and only mental phenomena.

There are other criteria for mental phenomena that I have not discussed here. It is sometimes claimed that what is unique to mental phenomena is that they are subjective, nonspatial, qualitative, purposive, phenomenal, mnemic, holistic, or emergent. It is beyond the scope of this chapter to discuss all these proposed criteria for identifying mental phenomena. The most comprehensive discussion of this subject is still Herbert Feigl's classic article, "The Mental and the Physical".<sup>60</sup> Even though this chapter does not attempt the exhaustive survey that Feigl makes, the fact remains that incorrigibility and intentionality are two of the strongest candidates for a criterion of mental phenomena. And neither of these two criteria is close to being a satisfactory criterion for mental phenomena.

Finally, it may be that the most promising course for someone who wishes to defend incorrigibility or intentionality as criteria of the mental is to concede that neither one is a necessary condition for something's being a mental phenomenon and that the criterion is one but not the only criterion for identifying mental phenomena. Just as was the case with incorrigibility,

intentionality does characterize a subset of mental phenomena, but it does not characterize all mental phenomena. A logical conclusion to draw, then, is that there may be no single criterion of what is mental and that if there can be a successful criterion which states a necessary condition, it will be a disjunction of criteria each of which applies to some but not all mental phenomena.

## FOOTNOTES

1. Richard Rorty, "Incorrigibility as the Mark of the Mental", The Journal of Philosophy, LXVII, 12 (June 25, 1970): pp. 399-424.
2. Richard Rorty, Philosophy and the Mirror of Nature (Princeton: Princeton University Press, 1979).
3. Rorty, 1970: pp. 406-7.
4. Ibid.
5. Ibid.
6. Ibid.
7. Ibid.
8. Ibid., p. 420.
9. Ibid., p. 417.
10. Ibid., p. 414-15.
11. Ibid., p. 412.
12. Ibid.
13. Ibid.
14. Ibid., p. 415.
15. Rorty, 1970, p. 120.
16. See Richard Rorty, "Mind-Body Identity, Privacy, and Categories" in The Mind/Brain Identity Theory, ed. C. V. Borst (New York: St. Martin's Press, 1975), pp. 187-213.
17. Rorty, 1970: p. 417.
18. Ibid., p. 418.
19. Ibid.
20. Ludwig Wittgenstein, On Certainty (New York: Harper & Row, 1972), p. 39e.
21. Rorty, 1970: p. 418.

22. Ibid., p. 419.
23. Ibid.
24. Ibid., p. 420.
25. Ibid., p. 421.
26. Rorty, 1979.
27. Ibid., p. 121.
28. Robert Schwartz, Book Review, Philosophy and the Mirror of Nature in The Journal of Philosophy, 1983.
29. Rorty, 1979, p. 120.
30. See Rorty, 1975.
31. Franz Brentano, "The Distinction Between Mental and Physical Phenomena", trans. by D. B. Terrell, in R. M. Chisholm, ed. Realism and the Background of Phenomenology, p. 50.
32. See, for example, R. M. Chisholm, Perceiving: A Philosophical Study (Ithaca, N.Y.: Cornell University Press, 1957), pp. 168-9. Also see Linda L. McAlister, "Chisholm and Brentano on Intentionality" Review of Metaphysics 28 (1974), pp. 328-38.
33. The most discussed Chisholm sources are R. M. Chisholm, 1957, pp. 168-88 and R. M. Chisholm, "Sentences About Believing", in Minnesota Studies in the Philosophy of Science, ed. Herbert Feigl, Michael Scriven, and Grover Maxwell (Minneapolis: University of Minnesota Press, 1957), pp. 510-20. Ausonio Marras, Intentionality, Mind, and Language (Chicago: University of Illinois Press, 1972) is an excellent collection of articles on intentionality, including an important article by William G. Lycan and a valuable commentary by Robert Sleight on Chisholm's "Notes on the Logic of Believing". Also see McAlister, 1974 for a discussion of how faithful Chisholm is to Brentano's thesis. More recent articles include Albert Flores, "The Thesis of Intentionality", Philosophia 7 (1978), pp. 501-14; Chisholm's reply, *ibid.*, pp. 612-15. Further discussion occurs in Ausonio Marras, "Intentionality Revisited", Philosophia 12 (1982), pp. 21-35. Also, Marion Knowles, "Some Remarks on the Intentionality of Thought", Philosophy and Phenomenological Research 43 (1983), pp. 267-79.
34. Flores, 1978.
35. Marras, 1982.

36. R. M. Chisholm, "Sentences About Believing", 1957.
37. R. M. Chisholm, Perceiving, A Philosophical Study, 1957.
38. R. M. Chisholm in A. Flores, 1978, p. 612.
39. Ibid.
40. Donald Davidson, "Mental Events", in Essays on Actions and Events (New York: Oxford University Press, 1980), pp. 210-11.
41. Flores, 1978, p. 503.
42. Ibid., p. 505.
43. Ibid.
44. R. M. Chisholm, "Sentences About Believing", p. 512.
45. David H. Sanford, "On Defining Intentionality", Proceedings of the XIVth. International Congress of Philosophy, Vol. II (1968), p. 220.
46. Flores, 1978, p. 507.
47. Ibid., p. 511.
48. R. M. Chisholm in A. Flores, 1978, p. 614.
49. Marras, 1982, p. 30.
50. Ibid.
51. Ibid.
52. Ibid., pp. 31-2.
53. Austio Marras, "Intentionality and Cognitive Sentences", in Intentionality, Mind, and Language. op. cit., p. 69.
54. Other interpretations of probability are the logical interpretation and the subjective interpretation. See Brian Skyrms, Choice and Chance, (Belmont, Cal.: Dickenson Publishing Co., 1966) for a discussion of these three interpretations of probability.
55. "Sentences About Believing", op. cit., p. 512.
56. Ibid.

57. R. M. Chisholm, Perceiving, p. 169.
58. R. M. Chisholm, "Intentionality", in The Encyclopedia of Philosophy, ed. Paul Edwards (New York: MacMillan Publishing Co., 1967), Vol IV, p. 204.
59. M. Knowles, 1983, p. 273.
60. Herbert Feigl, Minnesota Studies in the Philosophy of Science, op. cit., pp. 370-497.



## C H A P T E R   V I

### MIND-BODY INTERACTION AND THREE CONCEPTIONS OF THE PHYSICAL

#### 6.1   Introduction

In Chapter 2, I considered a form of dualism as presented by Saul Kripke. I also considered identity theory in this chapter in that much of the chapter was devoted to defending Kripke against the criticism of an identity theorist, Fred Feldman. However, I noted that the problem of interaction is a serious problem for dualism. I devoted Chapter 3 to considering anomalous monism, a position advocated by Donald Davidson. This position has its roots in Spinoza's dual aspect view.<sup>1</sup> I argued that Davidson has failed to establish anomalous monism and that there are good reasons for doubting that anomalous monism is true. Chapter 4 considered functionalism as presented by Daniel Dennett. This is really a very contemporary form of materialism, perhaps a more sophisticated form than the identity theory. I argued in this chapter that Dennett's functionalist account deals with the mechanisms of pain and not pain itself, and that functionalism cannot account for the subjective qualities of pain and other experiences. Finally, Chapter 5 considered two criteria for mental phenomena: incorrigibility and intentionality.

The following statements formulate the results from the preceding which are most important for future discussion:

- I. The dualistic division of the world into two ontologically distinct classes makes the distinctness of mind from body fairly easy to argue for. However, dualism will not be a viable theory until a plausible account of mind-body interaction is given which is consistent with dualist principles.
- II. In justifying his claim that there are no psychophysical laws, Davidson utilizes the claim that the physical conceptual domain is homonomic. The homonomic quality of the physical conceptual domain has the result that this domain includes only the language(s) of physics and chemistry, and not biology or neurology.
- III. Like other forms of materialism, functionalism describes the mechanisms and causes of pain and other mental phenomena, but does not account for mental phenomena. In particular, functionalism, like other materialist theories, is unable to account for the subjective qualities of experience.
- IV. It is unlikely that any single criterion can be found which uniquely identifies all and only mental phenomena. It is especially unlikely that a necessary condition for mental phenomena can be found. If such a condition is found, it is likely to be a disjunction of properties rather than one single property.

In this chapter I return to issues relating mainly to Statements I, II, and IV. In Section 6.2, I discuss mind-body interaction as a problem for dualism. In Section 6.3, I discuss the other half of the mental-physical distinction: the physical. I provide three different definitions of the term 'physical' which correspond to three distinct conceptions of what is physical. The belief that it is possible to provide a single definition of 'physical' that covers all physical entities and phenomena is one

reason the mind-body problem has proved so intractable. Section 6.4 summarizes the chapter. I develop the implications of the different definitions of 'physical' and the implications for the discussion of mind-body interaction in Chapter 7.

## 6.2 Dualism and Mind-Body Interaction

One formulation of dualism is the view that mental events, states, and phenomena are distinct from physical events, states, and phenomena. Other formulations hold that mind is distinct from body, or that a particular mind is distinct from any particular body. These formulations are all versions of ontological dualism, according to which dualism is a theory about the relation between phenomena or substances in the world or universe. Conceptual dualism, on the other hand, is a theory about terms which refer to these phenomena or substances. According to conceptual dualism, terms which refer to mental phenomena or substances cannot be reduced to or identified with terms that refer to physical phenomena or substances. In terms of the philosophers considered in earlier chapters, Saul Kripke (following Descartes) is both an ontological dualist and a conceptual dualist, while Donald Davidson is a conceptual dualist but an ontological monist.

Dualism is a view supported by what is sometimes called "common sense"<sup>2</sup>, and is suggested by many everyday locutions. Thus, we say that a person has a mind and a body, implying that there is a

difference, and we say "My mind is playing tricks on me". It certainly seems as if mental phenomena such as thoughts and feelings are distinct from brain events. When I am aware of another person's feelings, I do not seem to be aware of a state of his brain. Many people also hold that every human being has a soul, and some hold that this soul survives death and passes into another mode of existence or being. Dualism, then, is almost certainly the predominant view among those who have not reflected on the mind-body problem. This should perhaps not be considered evidence for dualism, in and of itself, but it is a fact about dualism.

At this same "common sense" level, interaction between mind and body, or between mental and physical phenomena, seems to be an undeniable fact. If I am stuck with a pin, it certainly seems as if the pin-prick causes my pain. Seeing a bear in the road will cause almost anyone to shiver with fright. If I commit a social faux pax, my face might turn red with embarrassment. Drugs such as marijuana, alcohol, and LSD can cause modified perceptions or entirely new perceptions. Voluntary action is often explained as follows: my decision to perform an act such as raising my arm is the cause of my arm's being raised. The world is also full of artefacts such as bridges, books, desks, and cars which are products of human minds. There are also common sense examples of mental events interacting with other mental events, as when a person's fear of flying causes him to decide to take alternative transportation. In all these

cases, mental and physical phenomena appear to interact in some way with each other.

In spite of the *prima facie* evidence for mind-body interaction, interaction has always been difficult to explain within a dualist framework. Explaining interaction has, in fact, been the principal difficulty for dualism, and this difficulty has provided materialists with several distinct arguments against dualism.

Michael Levin formulates one such argument as follows:<sup>3</sup>

Armed with this notion [that a given event e has only one cause which is the cause of e], the anti-dualist presses what appears to be the empirical fact that for every physical event e involving a human body, there is some preceding physical event which is the cause of e. As far as anyone knows there are no gaps in the sequence of bodily events to be bridged by a mental event, or into which a mental event might slip. There is no physical event whose cause is mental, in the dualist's intended sense.

Levin suggests that a dualist might give the following reply:<sup>4</sup>

The dualist can...reply that the causes of some bodily events are not purely physical. His argument here is reminiscent of a standard defence of epiphenomenalism, the thesis that some physical events cause mental events but mental events have no causal efficacy of their own. Epiphenomenalists reply to the charge that their theory makes mentality causally superfluous by observing that if mental event m had not occurred the physical event b that caused m would not have occurred, nor any further physical events b was the cause of. Similarly, replies the dualist, suppose that when I form the intention (event m) to move my leg (event e), and thereafter move it, neural event c causes e. Assuming that the formation of an intention is an event, both c and m are necessary for e. Had I not intended to move my leg, it would not have moved. Consequently, c does not contain all the necessary conditions for my leg's moving and is not the cause of my leg's moving.

According to this reply, both neural event c and mental event m are individually necessary for my leg's moving, and if c occurs without m or m occurs without c, my leg will not move. Thus, m and c are individually necessary and jointly sufficient for my leg to move. This makes the intention to move my leg a partial cause of my leg's moving. Levin does not say what the relation is between neural event c and my intention, m, or whether they occur simultaneously. If they occur independently of each other such that either could occur without the other, then it is difficult to see why the intention is also necessary. However, if they are related in such a way that they either occur together or not at all (c cannot occur without m or m without c), then this reply seems plausible. The only difficulty for a dualist, however, is that mental events are conceived as being conceptually and ontologically distinct from physical events, and hence a consistent dualist may not wish to admit a relation between mental and physical events which seems to imply that they together form a kind of mental-physical unity. However, it may be that this type of relation must be admitted if the problem of interaction is to be solved.

I discuss in more detail a view according to which mental and physical events form a kind of unity in Chapter 7. For the moment, consider another argument which Levin advances against interaction:<sup>5</sup>



Another traditional argument, however, does seem to show decisively that the causal relations between the mind as the dualist conceives it and the body are impossible. Originally my leg, which has mass w, is motionless. Then it moves distance d. Where did the energy come from? It is an empirical fact, the law of the conservation of energy, that the energy had to come from somewhere, and not only does it come from some preceding physical event, no immaterial substance could possibly supply mechanical energy. The mind can apply no physical force to the leg, while the contraction of the hamstring muscle supplies just the right amount - and the bodily event just is the motion of an object of mass w through distance d. Thus the contraction of the muscle is the cause of e. If there were non-physical causes of physical events, the energy in the universe would increase.

[C.D.] Broad cites, on the dualist's behalf, systems whose parts are causally related but which, allegedly, exchange no energy. The state of the bob of a pendulum is constantly changing, yet the shaft applies no force to the bob. This analogy is misguided in several ways. When my leg moves a force is being applied to a physical object; energy is being transferred. Since all the requisite force is supplied by a change in another physical object, any change in any other (especially immaterial) object is superfluous. The pendulum case is at best a precedent for a very limited number of mind-body interactions. More damaging is the fact that energy is actually being exchanged in the pendulum case. The shaft is exerting centripetal force on the bob, and the bob is exerting centrifugal force on the shaft; and at different positions the bob is exerting and bearing different amounts of both. The potential and kinetic energy of the bob are constantly changing.

Levin is right that energy is being exchanged in the pendulum case cited by Broad.<sup>6</sup> He goes on to say:<sup>7</sup>

The dualist may complain that this appeal to the conservation of energy reifies energy as a kind of stuff passed from object to object, whereas legitimate talk of energy (and mass) is shorthand for facts about acceleration. The conservation of energy in particular is shorthand for an empirical fact about constancies in such ratios. But even if energy is analytically eliminable, the present objection stands. Suppose we 'operationalize' energy and mass, explaining the mass of a body as its

(constant, additive) tendency to induce acceleration. My leg, which has a constant additive tendency to induce acceleration  $\underline{w}$ , moves with average velocity  $\underline{v}$  during time-interval  $\Delta t$ . It is a hard empirical fact that some object with a constant additive tendency  $\underline{w}'$  to induce acceleration must have moved with such an average velocity  $\underline{v}'$  during  $\Delta t$  that  $\int_{\Delta t}^{4t} 1/2 \underline{w} \underline{v}^2 dt = - \int_{\Delta t}^{4t} 1/2 \underline{w}' \underline{v}'^2 dt$ . Since...immaterial objects have no tendency to induce acceleration, it cannot have been an immaterial object that slowed down. But the cause of my leg's moving was the slowing down of the object with mass  $\underline{w}'$ .

Levin has formulated quite forcefully an argument which is often taken to be fatal to dualism: the argument from conservation of energy. In doing so, he cites one formulation of Newton's Third Law:

$\int_{\Delta t}^{4t} 1/2 \underline{w} \underline{v}^2 dt = - \int_{\Delta t}^{4t} 1/2 \underline{w}' \underline{v}'^2 dt$ . This law can also be formulated as follows: "The force that one object exerts on a second object is equal and opposite to the force that the second exerts on the first".<sup>8</sup> In this example, the leg is one object, while the "object with mass  $\underline{w}'$ " is the other object which is the cause of my leg's moving. Here this object with mass  $\underline{w}'$  is presumably my hamstring muscle. Since immaterial objects cannot induce acceleration or exert force on a physical object, the object with mass  $\underline{w}'$  cannot be a mental object or a mental phenomenon.

Let us examine this argument to see what, if anything, it shows. Levin says that the cause of his leg's moving was the slowing down of an object with mass  $\underline{w}'$ . Let us look at the physiology of the movement of limbs, to identify this object and see how the causation occurs:<sup>9</sup>

Prop your elbow up on a table with your fingers pointing toward the ceiling. Turn your hand so that you are looking at the palm. Now make a fist and alternately bend your fist first toward you and then away from you. When you bend your fist toward you, the muscles running from the inner surface of your wrist to the inner surface of your elbow are contracting. When you bend your fist away from you, the same muscles are being stretched...When the muscles contract, they pull on the tendons. Because the tendons are attached to the bones, those bones that are free to move (in this case, the bones of your hand) will be pulled into a new position. In those places where the bones cannot be moved (in this case, the bones of your elbow), there will simply be increased tension on the tendons.

When the muscles are stretched (in our example, the muscles are stretched when the muscles on the other side of the forearm contract), movable bones are pulled into a new position. Because of that new position, they, in turn, pull on the tendons and muscles.

When the muscles contract, they pull on the tendons which are attached to bones. Because the bones are in the hand, the entire hand moves. Even here we have a chain (series) of causal events, not a single cause. But the term 'cause' is often used to designate some event which occurs and which, together with certain preexisting or standing conditions, is sufficient for the event to occur. Thus the cause of the collapse of a weak bridge might be the passing over it of an especially heavy truck even though the bridge would not have collapsed if it had not already been weak. In our example, it is the contraction of the muscle that causes the hand movements, and it is muscle contractions that cause Levin's leg to move as well.

If Levin's point is simply that muscle contractions and not thoughts or volitions cause leg movements, then his argument does not cast doubt on dualism in relation to interaction. For no

responsible dualist would deny this; he would maintain that there is a series of events which originate in a thought or act of will, and the final event in this series is the movement of the leg. The muscle contracts in response to commands from the nervous system, so that it is in the brain and/or mind that we should look for the origin of the series of events which culminate with leg movement. Insisting that there must be only one cause and that it must be the immediate physical cause of the leg's moving ignores the role of the nervous system in voluntary motion.

Levin's argument is also oversimplified in that he requires that there be a single object with constant additive tendency  $w'$  to induce acceleration that must have moved with an average velocity  $\underline{v}'$  during  $t$ . The movement of a leg in most cases is the result of the contractions of a number of muscles, not just one. Perhaps Levin would say that  $\underline{v}'$  represents the average velocity of the relevant muscle contractions, taking into account the fact that some muscular action is inhibitory.

To provide an argument against dualism, Levin's argument must apply to the nervous system as well, since muscles contract or stretch in response to neural commands. Neural commands which originate in the higher portions of the brain and pass through the spinal cord give the muscles the signal to contract. A neuron consists of many dendrites, which carry information in to the soma, or cell body, and an axon, which carries information away from the soma. Figure 6-1 contains an illustration of the neuron. An action

potential is formed at the axon hillock, where the axon protrudes from the soma.

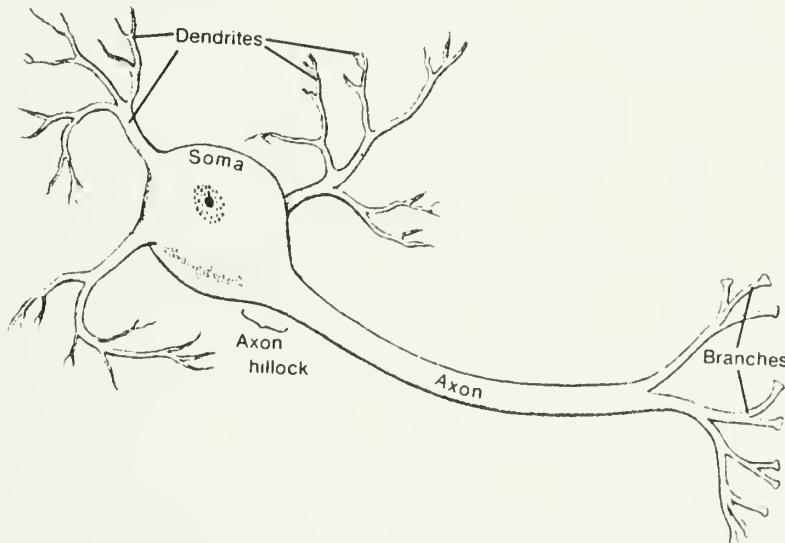


Figure 6-1. An Illustration of a Neuron

It is a well-known fact that the action potential produced by a neuron is not always equal to the sum total of incoming potentials. An axon has a threshold and it either produces a full-sized potential or it does nothing. This principle, known as the All-or-None Law, is explained as follows:<sup>10</sup>

If an extremely mild electric shock is delivered to the dendrites or soma, the axon does not produce any action potential. It does nothing. If the intensity of the shock is gradually increased, a point will be reached when the axon does produce an action potential. Therefore, we can say that the axon has a threshold; that is, the axon requires a certain level of disturbance before it reacts. And an axon either reacts completely by producing

a full-sized action potential, or else it does nothing. An axon cannot produce a half-sized or quarter-sized action potential. Either the sodium pump shuts off or it stays on; either an action potential is produced or the resting potential is maintained. This characteristic is called the All-or-None Principle: either the axon reacts by doing all it can (producing an action potential) or it does nothing. If the axon is to react, the disturbance that arrives at the axon hillock must reach or exceed the axon's threshold. Anything less intense will not cause an action potential.

William Uttal also comments on the All-or-None Law in the context of discussing sensory coding:<sup>11</sup>

The discovery of the all-or-none law effectively removed the amplitude of the response of a single axonal spike potential from the list of candidate coding dimensions. It now seems certain that the all-or-none law is valid for axonal spikes and that the amplitude of the individual nerve impulse is related only to the metabolic state of the axon and not to any characteristic of the stimulus once the spike threshold is exceeded. However, it should be remembered that in other parts of the neuron, it has been equally clearly established that slow potentials of graded amplitude and prolonged duration are the significant information symbols.

In the last sentence, Uttal is referring to graded potentials which also exist and whose amplitude does vary with incoming impulses.

But the important point here is that action potentials do obey the All-or-None Law, and that the amplitude of an action potential is not a function of incoming impulses.

One reason for the All-or-None Law is the existence of synapses in the brain. A synapse is a gap which serves as a point of communication between neurons, between neurons and effector organs, and between two muscle fibers. A single neuron may have many single synaptic connections, as we see from the following account:<sup>12</sup>



The axon of a neuron may terminate in only a few synapses or in up to 200,000 synapses (Purkinje cells of cerebellum). On the other hand, the dendrites and cell body of a single neuron may receive synaptic contacts from many neurons - from several hundred to as many as 200,000 separate axon contacts. In some neurons, as much as 40 percent of the surface of the cell body is covered with synapses. An average neuron is said to make from 1000 to 10,000 synaptic contacts with other neurons and may receive synaptic input from about 1000 other neurons.

Transmission of information across synapses occurs by both electrical and chemical means. In the chemical synapse, the arrival of an action potential generates the release of chemical substances known as neurotransmitters. These are released in packets or quanta which are the minimal unit of release. These neurotransmitters can evoke a response in the postsynaptic cell. In the electrical synapse, the currents generated by the action potential spread directly into the postsynaptic neuron across the synapse.

Neurotransmitters are not always released; it is the neurotransmitters which determine whether the action potential is continued.

Of course, a neuron has many inputs at once. We learn more about how it "decides" what to do from the following description:<sup>13</sup>

Each neuron may be influenced by hundreds or thousands of stimuli on its excitatory and inhibitory receptive sites. The complex interplay of the subliminal excitatory postsynaptic potentials (EXSP) and subliminal inhibitory postsynaptic potentials (IPSP) on each neuron endows the neuron with a variety of inputs; this provides the basis for the great plasticity of activity. Neurons are under continual synaptic bombardment. In this battleground of activity, the neuron reacts and responds. If the algebraic summation of the polarizing EPSPs and hyper-polarizing IPSPs results in a depolarizing event at the low-resistance, low threshold axon hillock, an action potential is generated and conducted along the axon. On

the other hand, if the algebraic summation of the graded subliminal potentials is not enough to stimulate the initial segment sufficiently, an action potential is not generated in the axon.

In summary, each neuron is a miniature "integration" center where the confluences and algebraic summation of all the decremental spreads of EPSPs and IPSPs take place on the dendrites and cell body in relation to their geometric distribution and locations on the neuron...We are dealing, in these central nervous system cells, with a brain in miniature. The resulting action potential is, in effect, a code message conveying information about graded events in the dendrites and cell body of one neuron to a distant location where a graded event can be initiated on another neuron.

Obviously, then, whether a neuron does fire on any given occasion is a complex matter; perhaps there is an internal rule built in to the neuron which it operates by. In any case, whether it does fire and the amplitude of the impulse is not simply a function of the sum total of the inputs.

Due to the All-or-None Law, it is not the case that when an action potential is generated, some body with mass  $w$  must have slowed down by the amplitude or speed of the action potential. In fact, Newton's third law does not even apply here since it applies to bodies with mass and an action potential is not a body with mass. An action potential is an electrical signal which passes through an axon, a body which does have mass. Neither graded potentials nor action potentials are bodies with mass, and both can occur in the chain of events beginning usually in the cerebral cortex and ending with the movement of a leg. What Levin needs to show is that mental events such as thoughts and volitions cannot trigger neural activity, not that mental events cannot move legs.

Levin has taken a mechanical principle of motion applying to bodies with mass and applied it to the human body which contains structures whose motion is not just governed by force and acceleration, but by codes, rules, and signals. I am not saying that neural structures violate the Conservation of Energy Principle; I am saying that it doesn't apply to neural impulses since they are not bodies with mass. Neural impulses are somewhat like magnetic fields and light waves in that they don't fit standard conceptions of what a body with mass is. An action potential is a kind of wave that travels through an axon; it is not a body with mass.

Levin's argument from the conservation of energy does not show that a thought cannot initiate neural activity since initiating neural activity does not require that some body with mass provide the acceleration. However, perhaps we can give a similar argument. In general, neural impulses are generated either by other neural impulses or spontaneously, probably by factors within the neuron itself. In the former case, a neuron receives a large number of graded potentials or action potentials and either generates a potential or does not, depending on its threshold and other factors. In the case of spontaneous activity, there is no reason to believe that an immaterial object is the cause. Hence in neither case does there appear to be any explanatory power added by postulating an immaterial object as the initiator of neural activity.

Perhaps a more sophisticated form of dualism which recognizes the distinction between the action of one body with mass on another

as it occurs in physics and the way in which neural activity occurs in the brain could explain interaction from a dualist viewpoint. However, any successful explanation will have to take into account the intentional and biological nature of the brain. The brain operates according to signals and rules, and it is part of a living biological system. Hence mechanical models of explanation imported from physics will not explain neurological processes. On the other hand, a thought has a neurological component in that it originates in the brain. Most forms of dualism have too simple a conception of "mental" and "physical" to accurately describe or explain these facts.

The real problem with dualism is that it divides the universe into two mutually exclusive and exhaustive categories. According to dualism, one conception of what is physical can apply to all cases in which the term 'physical' applies, while 'mental' refers to a distinct type of phenomenon. The motivation for this way of conceiving physical things may lie in Descartes' desire to bring human bodies and animals under the same laws as the entities of physics. Whatever the motivation, this view obscures important distinctions and creates unresolvable dichotomies. On the other hand, conceiving of a mental event such as a thought as an immaterial entity distinct from neurological (read "physical") events places a metaphysical synapse between them that no transmitter will ever cross. Only by revising its conception of

what is mental and what is physical can dualism hope to account for interaction.

A conceptual revision of the dualist conceptions of what is mental and what is physical which would possibly enable dualism to account for the interaction of mental and physical phenomena would have to take into account several factors. First, it would have to take into account the intentional and biological nature of the brain as well as the neurological aspects of mental events. Secondly, it would have to take into account the distinctions between the types of explanation and description in physics as opposed to neurology and biology. Third, it would have to give a consistent account of phenomena which appear to be non-physical and non-mental, or at least not purely physical or mental, such as persons, social conventions, and moral rules. As I will argue later, I do not see how such a conceptual revision can be made without abandoning dualism for some form of pluralism.

In conclusion, it seems that some materialist arguments against interaction are oversimplified in that they try to import laws and principles from physics which govern matter in motion into a neurological context. On the other hand, in just as oversimplified a manner, dualists accept this framework and try to reply within it. Because the problem of interaction is based on the false conceptions that the brain and matter as it occurs in physics are of the same type and obey the same laws, and that mental phenomena have no physical properties and physical phenomena have no mental

properties, neither argument can be conclusive. Both sides need to recognize that the universe is more complex than their theories allow.

I conclude, then, that the problem of interaction will remain unresolvable for dualistic theories unless and until a conceptual revision in the notions of "mental" and "physical" is made. Unless this can be done, or unless new arguments are presented, the problem of interaction may be fatal to dualism.

### 6.3. Three Conceptions of the Physical

What sort of conceptual revision is needed in the notions of "mental" and "physical" if the problem of interaction is to be resolved? I argued in Chapter 5 that it is unlikely that there is a single criterion for identifying all and only mental phenomena. Is there such a criterion for physical phenomena? Is it possible to give a single definition of 'physical' that identifies all and only physical phenomena? I will now discuss these questions.

#### 6.3.1 The Scientific Conception

In Chapter 3, I argued that Davidson's requirement that scientific laws draw on a homonomic conceptual domain has the result that the physical conceptual domain is restricted to the languages of chemistry and physics. Yet Davidson also cites a person's bodily movement as an example of a physical event. However, the conceptual



domains of biology, neurology, and human bodily movement are not homonomic since they include references to mental terms and to terms which are functionally defined such as 'life'.<sup>14</sup> Hence terms that describe human bodily movement belong to a different conceptual domain than the scientific one.

The attempt to interpret 'physical' as meaning "belonging to the conceptual domain of physics and chemistry" is not novel. This is a common conception of "physical". Hilary Putnam implicitly appeals to this in these remarks:<sup>15</sup>

Now, what do I mean by saying that the brain has non-physical properties? I mean properties which are definable in terms that do not mention the brain's physics or chemistry.

According to this conception, an object has physical properties just in case it has properties which are definable in terms which mention the physics or chemistry of the object.

Although this conception of what is physical as being that which is definable in terms from physics or chemistry is one that is often appealed to or implicitly assumed in mind-body discussions, it is too restricted to cover a number of central cases in which the term 'physical' is used. Herbert Feigl also saw the need of identifying this conception of "physical" as one to be distinguished from others. In his article "The 'Mental' and the 'Physical'" he distinguishes two concepts, which he calls "physical<sub>1</sub>" and "physical<sub>2</sub>":<sup>16</sup>

By "physical" I mean the type of concepts and laws which suffice in principle for the explanation and prediction of inorganic processes. If emergentism is not required for

the phenomena of organic life, "physical" would mean those concepts and laws sufficient for the explanation of inorganic as well as of biological phenomena. In accordance with the terminology of Meehl and Sellars, I shall henceforth designate this concept by "physical<sub>2</sub>" in contradistinction to "physical<sub>1</sub>", which is practically synonymous with "scientific", i.e. with being an essential part of the coherent and adequate descriptive and explanatory account of the spatio-temporal-causal world.

Feigl sees the need for distinguishing a scientific from a biological conception of "physical" due to the possibility of emergent phenomena. While it is true that emergent properties are a feature of biological and organic phenomena, they occur at the chemical level and the physical object level as well. Hence, emergence may not be the best choice to distinguish the two conceptions of "physical". As we shall soon see, however, there are good reasons for distinguishing two conceptions of "physical".

Davidson, Putnam, and Feigl all agree that there is a conception of "physical" according to which the term 'physical' is closely related to physics and chemistry or, more broadly, to the natural sciences. This is the conception of 'physical' which occurs in the phrases "the physical sciences". Here 'physical' contrasts with "non-physical sciences" such as biology and neuroscience and other life sciences, sociology, economics, and psychology. The reason Davidson uses this conception of "physical" is due to his views about scientific laws, and certainly laws, explanations, and theories are different in the natural sciences than in psychology and sociology. If Davidson is right that the physical forms a closed homonomic domain, then to say of something that it is

physical in this scientific sense has a very different meaning from saying that a bodily movement is physical.

I will now give a definition of the meaning of 'physical' which involves physics and chemistry. Because it is far from obvious what is meant by phrases such as "physical state" and "physical event", I state the definition in terms of the notions of physical property and physical entity. I choose 'entity' rather than 'object' because 'object' contrasts with 'subject' and so is more restricted. I call this the "scientific" conception of what is physical and denote it as follows: (s)physical. This conception can be defined in the following way:

P(s). x is a purely (s)physical property = df. x is a property and x is uniquely specifiable in terms from only the languages of physics and chemistry.

P(s)'. x is a uniquely (s)physical property = df. x is a property and x is partially but not completely specifiable in terms from only the languages of physics and chemistry.

E(s). x is a purely (s)physical entity = df. x is an entity and all of x's essential properties are purely (s)physical properties.

E(s)'. x is an impurely (s)physical entity = df. x is an entity and at least one but not all of x's essential properties are (s)physical properties.

The above four definitions are all ontological versions of (s)physical. A conceptual version can be created as follows:

T(s). 't' is an (s)physical term just in case 't' refers to some (s)physical property or (s)physical entity.

Using T(s), it is also possible to say what the (s)physical conceptual domain looks like, based on earlier work in Chapter

3.<sup>17</sup> The (s)physical conceptual domain will contain at least the following elements:

1. Logical terms ('and', 'or', 'if', etc.)
2. Mathematical terms ('number', 'line', 'point', etc.).  
The mathematics chosen can be governed by pragmatic considerations.
3. Syntactic rules.
4. (s)physical terms which are descriptive. These can be nouns, verbs, adjectives, or adverbs.

What terms qualify as "the terms of physics and chemistry"?

Some terms of physics or chemistry such as the term 'neutron' are originally introduced into the language of these sciences as part of an (s)physical or chemical theory. Other terms such as 'force' and 'acceleration' are terms taken from ordinary language and given a special meaning in an (s)physical or chemical theory. There is no single rule for determining whether a term is a term of physics or chemistry; one must consult the work of physicists and chemists and determine what language they use in formulating their theories. Even the term "the language of physics" is misleading since this language is composed of many subsets of smaller languages. However, in many examples such as 'neutron', 'force', 'mass', and 'gravity', there is no dispute that these terms are (s)physical terms.

What properties are (s)physical properties? The following equation, which is a formulation of Newton's Second Law, states a mathematical relationship using only (s)physical terms:<sup>18</sup>

$$\text{Force} = \text{Mass} \times \text{Acceleration}$$

Another example involves the gravitational attraction of a planet:

$$\text{Force due to gravity} = \frac{GMn}{r^2}$$

Here G is the gravitation constant, M is the mass of the sun, n is the mass of the planet, and r is the radius of the orbit.<sup>19</sup> The properties of force, mass, acceleration, and gravity are all (s)physical properties, uniquely specifiable in the languages of physics or chemistry.

What entities are (s)physical entities? Entities whose essential properties are all (s)physical are purely (s)physical entities. Atoms, along with protons, neutrons, and electrons are (s)physical entities under this criterion. However, a cup is not purely (s)physical since it is essential to an object's being a cup that someone has intended to use it for drinking or has intended that it be used for drinking. The term 'intention' is a mental term and not a term of physics or chemistry, so a cup is not purely (s)physical. However, it may be essential to a cup's being the cup it is that it have a certain chemical makeup. If this is true, a cup is an impurely (s)physical entity.

Other terms mix an (s)physical term with non-(s)physical terms. The phrase 'apparent shape' refers to the shape something appears to have. The property of apparent shape is a complex property involving the (s)physical property of shape and a mental property of appearance. Such a property is impurely (s)physical

since complete specification of it makes reference to a non-(s)physical (in this case mental) property.

### 6.3.2 The Personal Conception.

Much discussion relating to the mind-body problem centers around the body of a person or human being. Or, it might center around particular elements of a person's body such as his brain, a neural state, or his arm or leg. If we describe a human body in terms of the languages of physics and chemistry, we only describe chemicals, molecules, and certain other properties such as weight, mass, and heat. We can agree with Davidson that terms referring to these entities are part of a closed homonomic domain which forms the languages of these two physical sciences ((s)physical terms). When we give a physical description of a person or a human body in terms relating to a person we use terms such as 'beautiful', 'well-proportioned', 'tanned', 'muscular', and 'brunette' which come from a different conceptual domain: the domain of personal discourse. We also speak of a person's body as the physical aspect of a person and describe it as living, as being composed of organs such as the heart, and as having perceptual organs such as the eye, the ear, and the nose. These are not terms or entities of physics or chemistry ((s)physical terms or (s)physical entities).

This does not mean that one cannot make (s)physical statements about human bodies. This is possible, but only within the limited



(s)physical vocabulary. One recent text defines physical science as follows:<sup>20</sup>

The topic we consider in this book is Physical Science: the body of knowledge that concerns motion; gravitation; electricity; radiation; the structure of solids, liquids, and gases; radioactivity; astronomy; geology; and cosmology.

The difference between the (s)physical vocabulary and the vocabulary of personal discourse is one in points of view. From the viewpoint of physics, the human body is a mass of matter which differs only incidentally from other solid objects such as rocks and tables. As Russell says:<sup>21</sup>

There is...some conflict between what common sense regards as one thing, and what physics regards as an unchanging collection of particles. To common sense, a human body is one thing, but to science the matter composing it is continually changing.

Thus, it is possible to make (s)physical statements about the human body, but only about the physics or chemistry of the body.

Other examples show that there is a distinct conception of "physical" from "(s)physical" which relates to human (and animal) bodies. A person's physical needs are ones which relate to his body's proper functioning. Physical desires are desires relating to sexual desires and to other desires relating to fulfillment of bodily needs. If I am physically tired, my body is tired. Physical education involves instruction in the development and care of the human body. Physicality is an intensely physical orientation at the expense of the mental, spiritual, or social. A physical game (e.g. football) is one involving much body contact. The refrain "Let's

Get Physical" from a recent song exhorts intense body movements (viz. dancing). In none of these cases is a statement being made about a person's physics or chemistry, although in all cases an implicit or explicit reference is made to a person's body.

It might seem that I should call this the biological rather than the personal conception of "physical". This would not be inappropriate since a person's body is his biological aspect. However, as I have already noted, many words and concepts become entrenched in ordinary language long before scientists acquire an interest in studying the entities they denote. This is certainly the case with 'solid object' and with 'physical', 'mind', and 'thought'. Scientists are interested in explaining and predicting phenomena and events, and in giving descriptions of events which cohere with the theories necessary to explain and predict. Thus, words acquire new meanings in a scientific context and some well-entrenched descriptions are found not to apply at all.

Philosophical discussion of the mind-body issue often employs terms from science such as 'neuron' or 'brain state' but it also discusses everyday life examples such as a person raising his arm or drinking coffee. Thus there are terms used in this discussion which refer to an everyday point of view of the body, and also terms which are taken out of a highly specialized neurological or scientific context and whose meaning is seldom understood by anyone in the discussion. Probably the most correct view is that each science forms a conceptual domain to itself and that ordinary language has

its own conceptual domains which are distinct from the scientific domains. Because I cannot argue for this view here, I will simply define the personal physical domain in terms of properties which are uniquely specifiable in terms which make reference to a human or animal body or to elements of those bodies without specifying whether it has to be in a scientific (e.g. biological or neurological) context or an everyday life (personal) context. Hence, the personal conception of the physical will include biological, neurological, and ordinary language terms which refer to the body of a person or animal.

Some terms are used both in a scientific context and in an ordinary language context. An example is 'life'; biology is the study of living things, but we also speak of a living organ and of the life of the body. Other examples are terms that refer to parts of the body such as 'hand', 'heart', and 'brain'. If we describe an event such as a heart attack, the breaking of a leg, or a case in which someone suffers brain damage or brain-stem death as a physical event, whether in a scientific or an every day context, we don't mean that the properties of this event can be described in physics. Terms such as 'life', 'death', 'health', and 'brain damage' do not denote (s)physical properties. It is to account for the use of 'physical' in biological, neurological, and every day contexts in which human and animal bodies are discussed that we need a second conception of what is physical.

In order to understand this conception of "physical", it is helpful to notice that the term 'physical' refers to one of several elements or aspects of a person; others are social, intellectual, spiritual, mental, and moral. Because of the close connection between this notion of "physical" and the notion of person, I refer to this conception as the personal conception of what is physical, and denote it as follows: (p)physical. However, it does apply to animal bodies and to the bodies of human beings which are not persons, if there are any. It also applies in biological and neurological contexts in which human or animal bodies are discussed.

I define the personal meaning of 'physical' ((p)physical) as follows:

P(p). x is a purely (p)physical property just in case x is a property and x is uniquely specifiable in terms which only make reference to the body or to elements of the body of a person, human being, or animal.

P(p)'. x is an impurely (p)physical property just in case x is a property and x is partially but not completely specifiable in terms which only make reference to the body or to elements of the body of a person, human being, or animal.

E(p). x is a purely (p)physical entity = df. x is an entity and all of x's essential properties are (p)physical properties.

E(p)'. x is an impurely (p)physical entity = df. x is an entity and at least one but not all of x's essential properties are (p)physical properties.

We can create a conceptual version in the following way:

T(p). 't' is a (p)physical term just in case 't' refers to a (p)physical property or a (p)physical entity.

Using T(p), we can also give a reading of the (p)physical conceptual domain. The (p)physical conceptual domain will include at least the following elements:

1. Logical terms ('and', 'or', 'if', etc.)
2. Mathematical terms ('number', 'line', 'point', etc.). The mathematics chosen can be governed by pragmatic considerations.
3. Syntactic rules.
4. (p)physical terms which are descriptive. These can be nouns, verbs, adjectives, or adverbs.

Examples of (p)physical properties according to these definitions are the property of regeneration of the optic nerve of amphibians,<sup>22</sup> the homeostatic properties of the hypothalamus,<sup>23</sup> the (functional) property of the heart to pump the blood, and the property of every living body to grow. Examples of (p)physical entities are neurons, a human arm, leg, or bodily organ such as the heart, and the brain. All these properties and entities are (p)physical; none is uniquely specifiable in the terms of physics or chemistry.

### 6.3.3 The Objective Conception.

There is also a third conception of what is physical which has not been captured by either our scientific or personal conceptions. This is the term 'physical' as it occurs in statements such as "I can touch this thing so it must be physical" and "This table is a

physical object". This conception might be called a "common sense" conception of the physical, since it is the conception of "physical" most common in everyday life. Probably the term closest in meaning to this conception is 'objective'. To say something is objective is to say in part that it can exist independent of any mind. Criteria for applying the term 'physical' in this objective sense are as follows. If an object can be both seen and touched, this is reason to believe it is physical. Alternatively, if it can be seen and touched by several people at once, this is even more reason to believe that it is physical.

In everyday life, we regard physical objects as being solid, stable entities that endure through time. Physics has a different view of the matter. Bertrand Russell recognizes a gulf between the world of sense and the world of physics in these remarks:<sup>24</sup>

Among the objections to the reality of objects of sense, there is one which is derived from the apparent difference between matter as it appears in physics and things as they appear in sensation. Men of science, for the most part, are willing to condemn immediate data as "merely subjective", while yet maintaining the truth of the physics inferred from those data. But such an attitude, though it may be capable of justification, obviously stands in need of it; and the only justification possible must be one which exhibits matter as a logical construction from sense-data--unless, indeed, there were some wholly a priori principle by which unknown entities could be inferred from such as are known. It is therefore necessary to find some way of bridging the gulf between the world of physics and the world of sense.

Russell's concern here is epistemological; his entire lecture is devoted to discussing "the apparent difference between matter as it appears in physics and things as they appear in sensation".



Just as human bodies have properties which cannot be described by the languages of physics and chemistry, so physical objects have properties not describable in terms of their physics or chemistry. According to physics and chemistry, objects are not solid at all but are constantly changing masses of particles or waves. A table can be described as useful, beautiful, round, solid, and made of pine. According to the view of physics, physical objects are instances of matter in motion; the emergent properties physical objects have in virtue of being individual things which exist through time and which stand in economic, social, aesthetic, and historical relations to other objects cannot be uniquely specified in the terms of chemistry or physics. Other properties such as smoothness, roughness, having a grainy texture, a sensational color, of weighing more than a typewriter, and having a sweet taste are all properties we attribute to physical objects based on our perceptions of them. These are terms of ordinary language, not scientific language.

There are many physical objects which cannot be defined in terms which make an essential reference only to their physics or chemistry. For example, an artefact such as a cup is a physical object; yet it is essential to a cup that it have been in the past an object of the intention that it be used for drinking. And the term 'intention' is not a part of the language of physics.

Likewise, it is essential to my copy of Our Knowledge of the External World that it have the property that it was written by Bertrand Russell; yet this property is not part of its physics or

chemistry. It is essential to a tennis ball that it is intended to be used to play tennis. Artefacts have essential properties not definable in terms of physics or chemistry; yet they are physical objects according to some conception of "physical".

Physics and chemistry are concerned with physical objects, but only with certain of their aspects. In one physics text, I found the sentence "Blow up a small party balloon and you will have an excellent source of negative electric charge".<sup>25</sup> Does this make either 'balloon' or 'party balloon' a term of physics? Perhaps not; but the property of being a source of negative electric charge is an (s)physical property. Yet such a balloon will have other non-(s)physical properties such as frightening many people when it blows up and being the center of attention. Is the party balloon an (s)physical object because it has mass and weight? If we consider having mass an essential property of a balloon, then a party balloon is an impurely (s)physical entity because it has at least one essential (s)physical property. However, having weight should not be considered an essential property since a balloon in outer space is weightless but is still the same balloon.

I call this third conception of what is physical the "objective" conception, and denote it as follows: (o)physical.

This conception is defined as follows:

P(o). x is a purely (o)physical property = df. x is a property and x is uniquely specifiable in terms which refer only to properties that can exist independent of any mind and can be seen or touched by more than one person simultaneously.

P(o)'. x is an impurely (o)physical property = df. x is a property and x is partially but not completely specifiable in terms which refer only to properties that can exist independent of any mind and can be seen or touched by more than one person simultaneously.

E(o). x is a purely (o)physical entity = df. x is an entity and all of x's essential properties are (o)physical properties.

E(o)'. x is an impurely (o)physical entity = df. x is an entity and at least one but not all of x's essential properties are (o)physical properties.

A conceptual version can be created in the following way:

T(o). 't' is an (o)physical term just in case 't' refers to an (o)physical entity or to an (o)physical property.

The (o)physical conceptual domain will include at least these elements:

1. Logical terms ('and', 'or', 'if', etc.)
2. Mathematical terms ('number', 'line', 'point', etc.). The choice of mathematics can be governed by pragmatic considerations.
3. Syntactic rules.
4. (o)physical terms which are descriptive. These can be nouns, verbs, adjectives, or adverbs.

Examples of purely (o)physical entities are the sun, a rock, a mountain, a tree, and other natural objects. Artefacts are impurely (o)physical since they have an essential property that makes reference to an intention. Since light exists independent of the mind and can be seen simultaneously by two persons, it is purely (o)physical. I have already said that an atom is an (s)physical entity; however, the properties of an atom can exist independent of

any mind. If any of these mind-independent properties are essential and can in principle be seen or touched by more than one person at once, an atom is impurely (o)physical. The wind is a difficult example because while it exists independent of any mind, its effects can be seen but it cannot. It can be felt simultaneously by more than one person. However, the definition says 'touch' not 'feel', and allowing 'feel' would create even more problems. Perhaps the best solution is to say that the wind satisfies some but not all the criteria for (o)physical entities and hence should not be counted as (o)physical.

How do these definitions of 'physical' relate to each other? I will use the phrase '(x)physical property' to mean "either an (s)physical or a (p)physical or an (o)physical property". Some entities have different (x)physical properties. For example, a rock has the (o)physical property of hardness but it also has the (s)physical property of being composed of certain chemical elements. If it is essential to a rock that it be composed of certain chemical elements and it is essential to it that it exists independent of the mind, then a rock is impurely (s)physical and impurely (o)physical. Assuming that it is an essential property of a human body that it is the body of a person or human being, a human body is either purely or impurely (p)physical. If it is either an essential property of a human body that it can exist independent of any mind, or that it can be touched, then the human body is impurely (p)physical and impurely (o)physical. If it is also essential to a

human body that it be composed of certain chemical elements such as carbon or  $H_2O$ , then a human body is impurely (s)physical, impurely (p)physical, and impurely (o)physical. Hence it is likely that some single entities are (o)physical, (s)physical, and (p)physical.

It is likely that whatever analysis is given of a human body can also be given of a human brain, since a human brain is an element of a human body. Hence it is likely that a human brain is (s)physical, (o)physical, and (p)physical. If this is so, then trying to state the relations between a mental phenomenon and a human brain is more complex than it may appear. If a human brain is physical according to all three meanings of 'physical', then a mental event may stand in different relations to each of these three elements of a human brain. However, it is likely that the most interesting relations from the viewpoint of mind-body interaction will obtain between a mental phenomenon and the (p)physical elements of the brain. For it is this element that presupposes the concept of "person" and thus holds the most hope of a link between mental and physical phenomena. I discuss this subject as it relates to mind-body interaction in Chapter 7.

#### 6.4 Conclusion

Section 6.2 discussed mind-body interaction as a problem for dualism. The most important conclusion from this discussion is that the concepts of "mental" and "physical" used in the discussion of

this issue need to be revised before an account of mind-body interaction can be given. In Section 6.3 I proposed a revision of the concept of "physical". I distinguished three meanings of the term 'physical'. If ontological dualism is the view that there are two kinds of entities in the world, mental and physical entities, then accepting the results of Section 6.3 requires abandoning dualism. If the term 'physical' has three meanings, then there are more than two kinds of entities in the world, there are at least four: mental entities, (s)physical entities, (p)physical entities, and (o)physical entities. Since any view according to which there are more than two fundamentally distinct kinds of entities in the world is pluralistic, Section 6.3 is an argument for a form of pluralism.

In conclusion, it appears that the problem of mind-body interaction is fatal to dualism. Dualism as a philosophical view about mind and body is caught in the following dilemma:

- D. Either dualism can account for mind-body interaction or it cannot. If dualism cannot account for mind-body interaction, then dualism is not a viable theory. In order to account for mind-body interaction, a conceptual revision in the notion of "physical" is required which involves abandoning dualism for pluralism. Hence either dualism is not a viable theory or dualism must be abandoned for pluralism.

In chapter 7, I show how distinguishing three meanings of 'physical' helps give an account of mind-body interaction.



## FOOTNOTES

1. Benedict de Spinoza, *Ethics*, ed., James Gutmann, (New York: Hafner Publishing Company, 1953), pp. 79-126.
2. See G. E. Moore, "A Defense of Common Sense" in Philosophical Papers (New York: Collier Books, 1966), pp. 32-59. In this article, Moore identifies a number of propositions which he claims to know with certainty. The first one is "There exists at present a living human body, which is my body". These propositions are ones which are fundamental to our total conceptual domain. This article served as the starting point for Wittgenstein in On Certainty. Wittgenstein claimed that Moore's propositions are so fundamental that to say we know them is inappropriate. We might describe them as the fundamental assumptions of our conceptual domain. The appeal to "common sense" is sometimes associated with the "Ordinary Language" movement which can be traced back to Moore and through Wittgenstein and John L. Austin. This movement has waned in recent years but is alive in spirit in Kripke's emphasis on the importance of the appeal to intuition as a philosophical datum. One important criticism of the common sense-ordinary language-intuition methodology is that if there are inconsistencies in our use of language, then its usefulness as a test of correct usage and of truth is suspect.
3. Michael E. Levin, Metaphysics and the Mind-Body Problem (Oxford, England: Clarendon Press, 1979), p. 82.
4. Ibid., p. 83.
5. Ibid., pp. 84-5.
6. See Clifford E. Swartz, Phenomenal Physics (New York: John Wiley & Sons, 1981), pp. 112-35 for an account of centrifugal force and centripetal force and their application to the pendulum case.
7. Levin, pp. 85-6.
8. This formulation of Newton's Third Law is found in Swartz, p. 147.
9. Jacqueline Ludel, Introduction to Sensory Processes (San Francisco: W. H. Freeman and Company, 1978), p. 315.
10. Ludel, p. 18. The illustration of a neuron is also from Ludel, p. 17.

11. William Uttal, The Psychobiology of Mind (Hillsdale, N.J.: Lawrence Erlbaum Associates, 1978), p. 423.
12. Charles R. Noback and Robert J. Demarest, The Human Nervous System, Basic Principles of Neurobiology (New York: McGraw-Hill Book Company, 1981), p. 63.
13. Ibid., pp. 106-7.
14. See Chapter 3, Section 3.5.
15. Hilary Putnam, Reason, Truth, and History (New York: Cambridge University Press, 1981), p. 78. The emphasis here is Putnam's.
16. Herbert Feigl, "The 'Mental' and the 'Physical'", in Minnesota Studies in the Philosophy of Science, ed. Herbert Feigl, Michael Scriven, and Grover Maxwell, Vol. II, Concepts, Theories, and the Mind-Body Problem (Minneapolis: University of Minnesota Press, 1957), p. 377.
17. See Chapter 3, Section 3.5.
18. Ernest C. Pollard and Douglas C. Huston, Physics, An Introduction (New York: Oxford University Press, 1969), p. 69.
19. Ibid., p. 277.
20. Ibid., p. 3.
21. Bertrand Russell, Our Knowledge of the External World (New York: The New American Library, 1956), p. 86.
22. See Steven Rose, The Conscious Brain (New York: Alfred A. Knopf, Inc., 1973), p. 167 for an account of the experiments by Roger Sperry on the regenerative properties of the optic nerve in frogs, toads, newts, and salamanders.
23. Ibid., pp. 225-9. Here Rose gives an account of how the hypothalamus regulates bodily temperature. Mechanisms that regulate bodily temperature are called homeostatic mechanisms.
24. Russell, 1956, p. 81.
25. Clifford E. Swartz, Phenomenal Physics (New York: John Wiley & Sons), p. 505.

C H A P T E R   V I I  
AN ACCOUNT OF MIND-BODY INTERACTION

7.1   Introduction

In Section 6.2 I argued that if dualists are to give an account of mind-body interaction, they must accept a conceptual revision in the notions of "mental" and "physical". In Section 6.3 I provided such a revision in the notion of "physical" by distinguishing three conceptions of the physical: the scientific ((s)physical) conception, the objective ((o)physical) conception, and the personal ((p)physical) conception. In Section 6.4 I argued that accepting the view that there are three conceptions of the physical requires abandoning dualism for a form of pluralism.

In this chapter I return to the problem of giving an account of mind-body interaction. I return to the argument by Michael Levin against the possibility of mind-body interaction which I discuss in Section 6.2. I concede that Levin's argument may be successful in showing the impossibility of causal interaction between mental phenomena and (p)physical entities, and that a variation of Levin's argument may show the impossibility of causal interaction between mental phenomena and (o)physical entities. However, I deny that either Levin's argument or a variation of it shows the impossibility

of causal interaction between mental phenomena and (p)physical entities.

The main result needed to support the claim that Levin's argument against the possibility of mind-body interaction is unsuccessful if applied to (p)physical entities is that some (p)physical entities have mental properties or aspects and some mental phenomena have (p)physical properties or aspects. I support this result by citing discussions of the visual system which show that some neurons have properties which are uniquely specifiable by reference to mental terms. I also cite Davidson's contention, discussed in Chapter 3, that an event is mental if and only if it can be picked out by a mental description in support of the claim that (p)physical entities have mental properties. That mental phenomena have (p)physical properties follows from the claim that (p)physical entities have mental properties. I also cite an independent argument for this claim according to which mental phenomena that originate in (p)physical entities have at least one (p)physical property: the property that they originate in a (p)physical entity.

In Section 7.3 I give an account of interaction by using the result that (p)physical entities have mental properties or aspects and mental phenomena have (p)physical properties or aspects. According to this account, when a mental entity such as a thought causally acts on the brain or initiates a neural event, it is the neural aspect of the thought that acts on some other neural aspect.

This second neural aspect may or may not be associated with some other mental phenomenon. I cite a study by Claude Ghez of voluntary movement in support of this account. My account of interaction applies only to mental phenomena with (p)physical aspects: if there is any mental phenomenon with no (p)physical aspect, then my account of interaction does not apply to this mental phenomenon.

In Section 7.4 I cite views by Jack Ornstein, G. D. Wassermann, and John Searle which are similar in one or more respects to the view I present. In Section 7.5 I sketch a more general position that links a plural aspect view with Leibniz's pluralism of viewpoints. I call this position Viewpoint Pluralism. Section 7.6 summarizes the chapter. Section 7.7 briefly summarizes earlier chapters, as well as Chapter 7.

## 7.2 Levin's Argument Against Mind-Body Interaction

In Chapter 6, I discussed an argument stated by Michael Levin purporting to show that causal relations between mind and body are impossible. This argument appeals to the law of the conservation of energy. According to Levin:<sup>1</sup>

Another traditional argument, however, does seem to show decisively that the causal relations between the mind as the dualist conceives it and the body are impossible. Originally my leg, which has mass  $w$ , is motionless. Then it moves distance  $d$ . Where did the energy come from? It is an empirical fact, the law of the conservation of energy, that the energy had to come from somewhere, and not only does it come from some preceding physical event, no immaterial substance could possibly supply mechanical energy. The mind can apply no physical force to the leg,

while the contraction of the hamstring muscle supplies just the right amount - and the bodily event just is the motion of an object of mass w through distance d. Thus the contraction of the muscle is the cause of e. If there were non-physical causes of physical events, the energy in the universe would increase.

The dualist may complain that this appeal to the conservation of energy reifies energy as a kind of stuff passed from object to object, whereas legitimate talk of energy (and mass) is shorthand for facts about acceleration. The conservation of energy in particular is shorthand for an empirical fact about constancies in such ratios. But even if energy is analytically eliminable, the present objection stands. Suppose we 'operationalize' energy and mass, explaining the mass of a body as its (constant, additive) tendency to induce acceleration. My leg, which has a constant additive tendency to induce acceleration w, moves with average velocity v during time-interval  $\Delta t$ . It is a hard empirical fact that some object with a constant additive tendency w' to induce acceleration must have moved with such an average velocity v' during  $\Delta t$  that  $\int_{\Delta t} 1/2 w v^2 dt = - \int_{\Delta t} 1/2 w' v'^2 dt$ . Since...immaterial objects have no tendency to induce acceleration, it cannot have been an immaterial object that slowed down. But the cause of my leg's moving was the slowing down of the object with mass w'.

### 7.2.1 Criticisms of Levin's Argument

One criticism of the argument Levin states in the above quotation is that Levin's insistence that there be one cause of the leg's moving and that this cause be the immediate physical cause ignores the role of the nervous system in voluntary motion. The movement of a hamstring muscle may be the immediate cause of my leg's moving, but this muscle moves in response to neural commands. These neural commands occur as a sequence or series of events which culminates in an instruction to the hamstring muscle to contract. The first event of this series or sequence of neural events is as



important a cause of my leg's moving as the movement of the hamstring muscle, even though it is not the immediate cause.

A second criticism of Levin's argument is that Levin has taken a mechanical principle of motion applying to bodies with mass and applied it to the human body which contains structures whose motion is not just governed by force and acceleration, but by codes, rules, and signals. Now that I have distinguished (s)physical from (p)physical entities, I can state this criticism in relation to this distinction. The law Levin cites, the law of the conservation of energy, is a law of physics that is formulated in (s)physical terms. The laws that govern neuronal interactions are formulated in (p)physical terms. Levin's conservation of energy argument does not apply to a sequence of neural events because whether a neuron fires and the amplitude of the impulse is not simply a function of the sum total of the inputs. Whether a neuron fires depends on the way in which the neuron integrates and sums its inputs.<sup>2</sup> The conservation of energy argument Levin cites applies to bodies with mass, and an action potential is not a body with mass.

After stating these criticisms, I formulated another argument which is similar to Levin's.<sup>3</sup> According to this argument, neural impulses are generated either by other neural impulses or spontaneously. In the former case, a neuron receives a large number of graded potentials or action potentials and either generates a potential or does not, depending on its threshold and other factors. In the case of spontaneous activity, there is no reason to

believe that an immaterial object is the cause. Hence in neither case does there appear to be any explanatory power added by postulating an immaterial object as its cause. Hence, no mental event plays a causal role in neural activity.

If this argument appears irrefutable, it is only because the conceptions of "mental event" and "physical event" it employs are dichotomous and inaccurate. In Chapter 6 I distinguished three conceptions of the physical, and gave three corresponding definitions of "x is a physical property" and "x is a physical entity". I called these three conceptions the scientific ((s)physical) conception, the objective ((o)physical) conception, and the personal ((p)physical) conception. I now discuss how distinguishing these three conceptions of the physical affects the problem of mind-body interaction.

#### 7.2.2 Levin's Argument Applied to (s)physical, (o)physical, and (p)physical Entities

Levin's original argument purported to show that no immaterial object could impart mechanical energy to someone's leg. Levin says "the energy had to come from somewhere, and not only does it come from some preceding physical event, no immaterial substance could possibly supply mechanical energy".<sup>4</sup> A human leg has impurely (s)physical properties since it has weight and mass. A human leg has impurely (o)physical properties since it can exist independent of any mind and can be seen or touched by more than one person at once. A human leg is impurely (p)physical since it is partially

specifiable in terms which make reference to a human being. It seems safe to say that it is essential to a human leg that it is the leg of a human being. Some of the (s)physical properties of a human leg are probably essential. It is probably essential to a human leg that it have mass and that it contain certain chemical elements such as carbon and water. It is probably also essential to a human leg that it can exist independent of a human mind. If these (s)physical and (o)physical properties are essential, as seems likely, a human leg is an impurely (s)physical entity, an impurely (o)physical entity, and an impurely (p)physical entity.

Levin's argument against the possibility of interaction seems addressed to those properties of a human leg in virtue of which it is an (impurely) (s)physical entity. The concept of mechanical energy is a concept of physics, and Levin claims that no immaterial object can supply mechanical energy to an object with mass. If we accept the correctness of the terms Levin uses to phrase his argument, his argument is correct. If a mental phenomenon is an immaterial object, it is difficult to see how a mental object could supply mechanical energy to at least a purely (s)physical object. However, a human leg is an impurely (s)physical object, so I will consider its non-(s)physical properties in order to determine whether his argument succeeds.

Levin's argument could be rephrased as an argument purporting to show that no mental phenomenon can interact causally with an (o)physical object. If we consider an impurely (o)physical object

such as a rock, it seems clear that only the application of some kind of physical energy can cause a rock to be moved. This energy is (s)physical energy since it can be uniquely specified by using terms of physics. It takes (s)physical energy to move an (o)physical object. And no immaterial object can exert (s)physical energy. Such an argument seems to show the impossibility of someone's moving a chair or a fork purely by mental concentration. While such feats have been attempted,<sup>5</sup> the correct explanation of these feats and whether they show anything is far from clear. If the case for mind-body interaction rests on the ability of some people to perform these dubious feats, the case for mind-body interaction is weak at best. However, there is one more type of physical property to consider: (p)physical properties.

Standard examples of mind-body interaction typically involve the interaction of a mental phenomenon with one or more (p)physical entities. A feeling of embarrassment causes my face to turn red, or my thought that I will move my leg causes my leg to move. Or, my feeling of love for someone causes my heart to beat rapidly. While my face, my leg, and my heart are all impurely (s)physical and impurely (o)physical entities, it is the impurely (p)physical properties of my face, my leg, and my heart that are particularly relevant to the issue of mind-body interaction. I now consider these (p)physical properties.

I now repeat my definition of '(p)physical property' and '(p)physical entity':

- P(p). x is a purely (p)physical property = df. x is a property and x is uniquely specifiable in terms which only make reference to the body or to elements of the body of a person, human being, or animal.
- P(p)'. x is an impurely (p)physical property = df. x is a property and x is partially but not completely specifiable in terms which only make reference to the body or to elements of the body of a person, human being, or animal.
- E(p). x is a purely (p)physical entity = df. x is an entity and all of x's essential properties are uniquely specifiable in terms which only make reference to the body or to elements of the body of a person, human being, or animal.
- E(p)'. x is an impurely (p)physical entity = df. x is an entity and at least one but not all of x's essential properties are uniquely specifiable in terms which only make reference to the body or to elements of the body of a person, human being, or animal.

We can create a conceptual version in the following way:

- T(p). 't' is a (p)physical term just in case 't' refers to a (p)physical property or a (p)physical entity.

Since both a human brain and a human eye are elements of a human body, both are impurely (p)physical entities. In Chapter 3<sup>6</sup> I argued that some terms for elements of the human body are defined by their mental function. Examples are terms for sense organs such as 'eye', 'ear', and 'nose'. I also stated that if scientists were to study these sense organs apart from their mental function they would be unable to account for many of their interesting properties. Viewing sense organs apart from their mental functions also strips science of much of its explanatory and descriptive

ability. I now support these claims by considering an account of how the eye works.

The visual system is a very complex system that includes many structures both in the eye and in different parts of the brain. Eric R. Kandel presents a schematic diagram of the visual system, shown in Figure 7-1.<sup>7</sup>

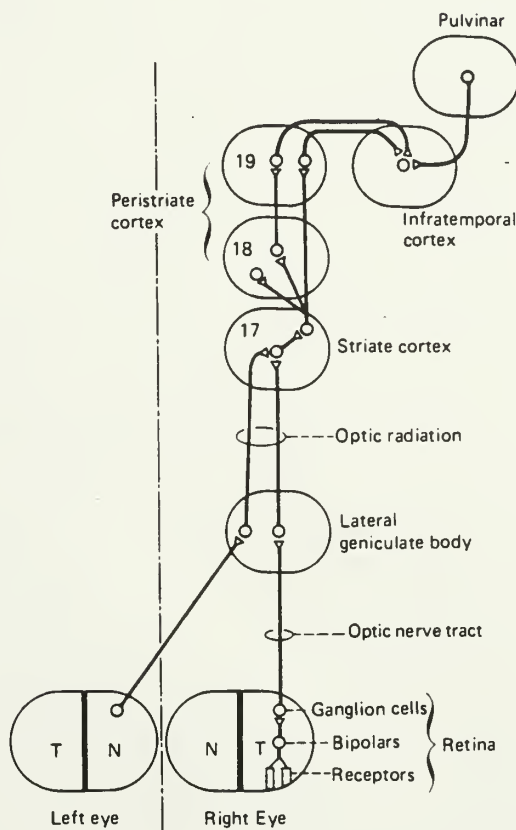


Figure 7-1. A Schematic Diagram of the Visual System



The following selection is taken from a discussion by Kandel of the striate cortex, a structure which plays a key role in human vision:<sup>8</sup>

The structure of the striate cortex is more complex than that of the lateral geniculate nucleus. Large numbers of cells are arranged in several different ways. Most of the input from the geniculate comes in layer IV of the cortex and then is conveyed to the critical layers above and below. Paralleling the increase in anatomical complexity, Hubel and Wiesel found an increase in complexity of the stimulus requirements needed for cells to respond. Small spots of light that are effective in the retina and the lateral geniculate are not as effective in the striate cortex. To be effective, a stimulus must have linear properties....It must be a line, a bar, or something that has a clear edge. Hubel and Wiesel categorized the cortical cells in area 17 into two major groups, simple and complex. Both respond to linear stimuli....

Some support for the hypothesis that complex cells are connected to simple ones comes from the finding by Hubel and Wiesel that the visual cortex is organized into narrow columns (much as is the somatic sensory cortex), running from about the surface to the white matter. Each column is about 10-30  $\mu$ m wide and 2 mm deep, and in each column there are many simple cells with almost identical retinal position and identical axes of orientation. In any column there are also complex cells. The properties of these complex cells in a column could most easily be explained by postulating that each complex cell receives direct connections from the simple cells in the column. This finding gives us a completely new insight into the importance of cortical columns. In the visual system columns seem to serve as anatomical devices for bringing cells together so as to interconnect them and to generate, by means of these interconnections, a new level of abstraction of visual information. For example, the columns allow cortical cells to generate linear receptive field properties from a geniculate input that responds best to small spots of light.

The discovery of columns in the various sensory systems is perhaps the most important single advance in cortical physiology in the past several decades and has led to a number of related discoveries. One can ask: given that cells with the same axis of orientation tend to be grouped into columns, how are columns of cells with different axes of orientation organized in relation to one another?

Detailed mapping of adjacent columns by Hubel and Wiesel, using tangential microelectrode penetrations, has revealed a very orderly shift of axis orientation. Every 20-50  $\mu$ m, the electrode encounters a new column and a shift in axis of orientation of about  $10^\circ$ ....

In addition to columns devoted to axis of orientation there are columns devoted to eye preference or ocular dominance (properties related to binocular interaction)....Hubel and Wiesel have designated the term hypercolumn to refer to the whole set of columns for analyzing lines of all orientations from a particular region in space via both eyes. From these studies it is apparent that the striate cortex carries out two major functions: (1) it combines the input from the two eyes, a step in a sequence of transformations necessary for depth perception; and (2) it decomposes the visual world into short line segments of various orientations, an early step in the process thought to be necessary for pattern discrimination.

Hubel and Wiesel suggest that the interaction between simple and complex cells may be important for the [following] issue:...the perception of patterns irrespective of where they fall on the retina. Take a vertical line--or an object with vertical edges--that is located in front of you. The vertical edge (or the line) will excite a population of simple cells and a population of complex ones, each with a vertical axis of orientation. A slight (saccadic) movement of your eye or of the object will call into play a new population of simple cells (because these cells are very sensitive to the exact position of the line in the receptive field). However, for a small movement the stimulus will still excite the original population of complex cells (because these cells are less sensitive to movement within the receptive field)....

An unresolved question that is now receiving increasing attention is, how far can this hierarchy go? (Is there a grandmother cell?) Is there a group of cells that observe the hyper-hypercomplex cells and make one aware of the total pattern? If so, is there a still higher group in the hierarchy that looks at combinations of complex patterns as these enter our awareness?

There may indeed be other higher order cells combining the computational results of the inferotemporal, peristriate, and striate cortices to produce even more elaborate abstractions. However, to discern the relatively simple features we have thus far considered has already required an enormous proportion of visual brain. It would appear curious to attribute progressively more important processing to a relatively small group of cells

and expect of them this enormously complex abstraction. An alternative to this supposition would be that at higher levels the mechanism of transformation changes, so that single units no longer serve to represent feature states but represent only selected aspects of the percept. To represent a familiar face or a landscape may require parallel processing, that is, activity in cells in different areas in the inferotemporal, peristriate, and striate cortices, with the cells in each area coding for a particular aspect of the stimulus: for shape, depth, movement, and color. At this higher level cells in many parallel visual areas are likely to be involved, and their simultaneous activity may serve as the feature detector. The states of the parts taken separately may not represent the whole; rather, it is the relationship among them that may be important. One can draw an analogy to the individual silver halide grains of a photograph: these do not represent the photograph of a face, but the ensemble of grains does.

There is in fact some support for this view. We have seen that even as early as the retina, visual information is carried to the lateral geniculate nucleus and then to area 17 of the cortex by two parallel systems, the X and Y cells. Moreover, we described area 17 above as if it sent one main projection to areas 18 and 19. Actually, S. M. Zeki at University College London has found that area 17 sends five separate projections to area 18 alone. Zeki found that the cells in each of the five regions that receive input from area 17 have different functional properties, a finding that supports the idea that each region handles a different aspect of visual information. For example, several regions in area 18 contain cells that code for color, other regions contain cells that code for movement, others still for binocular disparity necessary for depth perception. Interconnections among these various regions could generate a very large variety of response types suitable for analyzing in almost endless detail the intricacies of the visual world.

Explaining and understanding the experience of color vision also requires viewing the visual system in relation to the qualities of subjective experience. The following discussion of color vision by Peter Gouras illustrates this fact:<sup>9</sup>

Animals with color vision simply have great perceptual capability, and this presumably has great survival value. Think for a moment of a black and white version of a work by a colorist such as Turner, Monet, or Renoir; almost a million nuances of contrasting shapes are lost by an achromatic rendition of one of their works. This high degree of discrimination in color vision can actually be understood in quantitative terms. The subjective experience of color can be broken down into three semi-independent sensibilities: (1) hue, (2) saturation, and (3) brightness. All color experience is composed of these three psychological impressions.

Hue has the strongest effect on color and is the major determinant of principal colors such as red, yellow, green, and blue. It is what we ordinarily mean by "color". This impression is determined by the proportion in which the three cone mechanisms are activated by the object and its background. The brain must keep track of how much each of the three photoreceptor systems contributes to the detection of an object. Most of us have names for only a family of hues, but actually about 200 varieties can be distinguished. The second distinct quality of color is saturation, which reflects how much a hue has been diluted by grayness; this is determined by the degree to which all three cone mechanisms are stimulated in common by the object and by the background. There are about 20 distinguishable steps of saturation for each hue.

The third quality of color, brightness, is a sensation shared with achromatic visual systems. It is due to the total effect on all three cone mechanisms of an object relative to its background. We shall see later that one of the three cone mechanisms (the so-called blue mechanism) makes little or no contribution to brightness. It is the brightness factor that turns orange into brown and gray into black or white. There are about 500 distinguishable steps of brightness for every hue and grade of saturation. In contrast to achromatic vision with only 500 steps, color vision has more than 1 million gradations with which to detect the contours of shapes in the external world....It is no wonder that natural selection uncovered its power.

By the latter part of the eighteenth century it was realized that the ability to experience myriad different colors might not be due to an equivalent number of detectors in the retina but to a minimum number--three--with every color determined by the proportions in which each of these three detectors responds to light.



The preceding quotations from Kandel and Gouras show that in describing and explaining the visual system, including its neural components, it is necessary to make use of concepts that contain a reference to mental terms. Kandel's discussion contains terms such as 'shape', 'depth perception', 'binocular movement', and 'color' which refer to aspects of visual experience. Gouras states that "the subjective experience of color can be broken down into three semi-independent sensibilities: (1) hue, (2) saturation, and (3) brightness. All color experience is composed of these three psychological impressions".<sup>10</sup> Even the concept of a visual system is defined by reference to vision, which is a mental phenomenon. Someone who tried to account for the visual system in terms of only its (s)physical or (o)physical properties could not explain how the neural structures of the visual system encode a retinal image and function to make possible the experience of sight. The neural structures relating to color vision would be particularly impossible to account for using only (s)physical and (o)physical terms. It seems clear, then, that viewing the visual system and its components apart from their mental functions strips science of its explanatory and predictive power.

I now return to Levin's argument against the possibility of interaction between mental phenomena and physical entities. I have conceded that Levin's argument may show that mental phenomena do not interact causally with (s)physical and (o)physical entities. However, it is not clear that it shows that interaction between

mental phenomena and (p)physical entities is not possible. Earlier in this chapter<sup>11</sup> I formulated an argument similar to Levin's which purports to show this. According to this argument, neural impulses are either generated by other neural impulses or spontaneously. In neither case is it necessary to postulate an immaterial object as a cause. I now reconsider this argument in light of my discussion of (p)physical entities and (p)physical properties.

The discussion of the visual system by Kandel and Gouras makes it clear that some neurons have properties that are uniquely specifiable by reference to mental terms. Kandel says:<sup>12</sup>

Several regions in area 18 contain cells that code for color, other regions contain cells that code for movement, others still for binocular disparity necessary for depth perception.

These (neural) cells that code for color have the property that they code for color. Likewise, other (neural) cells have the property that they code for the binocular disparity necessary for depth perception. According to Gouras:<sup>13</sup>

Hue has the strongest effect on color and is the major determinant of principal colors such as red, yellow, green, and blue. It is what we ordinarily mean by "color". This impression is determined by the proportion in which the three cone mechanisms are activated by the object and its background. The brain must keep track of how much each of the three photoreceptor systems contributes to the detection of an object.

The three cone mechanisms have the property that the proportion in which they are activated by the object and its background determines



the impression of color the person experiences. These examples show that some (p)physical entities have mental properties.

The result that some (p)physical entities have mental properties is consistent with the argument by Davidson which I discussed in Chapter 3 that every physical event can be picked out by a mental description.<sup>14</sup> If we accept Davidson's view that an event is mental if and only if it can be picked out by a mental description, the result that some (p)physical entities such as neurons have mental properties is even less surprising. Presumably the descriptions given by Kandel and Gouras can be used to pick out individual neurons, or at least groups of neurons. However, these neurons will be impurely mental since they are also (p)physical entities.

I now define the terms 'purely mental entity', 'impurely mental entity', 'purely mental property', and 'impurely mental property'.

M(m). x is a purely mental property = df. x is a property and x is uniquely specifiable in terms which only make reference to mental phenomena.

M(m)'. x is an impurely mental property = df. x is a property and x is partially but not completely specifiable in terms which only make reference to mental phenomena.

E(m). x is a purely mental entity = df. x is an entity and all of x's essential properties are mental properties.

E(m)'. x is an impurely mental entity = df. x is an entity and at least one but not all of x's essential properties are mental properties.

I have shown that some (p)physical entities have mental properties, and have defined 'mental entity'. Do some mental

entities have (p)physical properties? That this is so follows from the fact that some (p)physical entities have mental properties. If a particular neuron has the property that it codes for a particular experience of color, then that particular experience of color has the property that it is coded for by a particular neuron. A particular impression of color has the property that it is determined by the proportion in which the three cone mechanisms are activated by an object and its background. Hence mental entities have (p)physical properties.

It is possible to give an independent argument to show that some mental entities have (p)physical properties. If I have a toothache, it is a property of this toothache that it originates in a particular tooth. A toothache originating in a different tooth would not be the same toothache even if it is in the same phenomenal location. Since this is so, the property of originating in a particular tooth appears to be essential to the toothache's being the toothache it is. If this property is essential, it is a (p)physical property and a toothache is an impurely (p)physical entity. If a feeling of love makes my heart beat rapidly on a particular occasion, then it is a property of my experience of a feeling of love on that occasion that it makes my heart beat faster. It is generally believed that thoughts have a neural basis. Hence if I have a thought about my plans for the weekend, a property of this thought is that it is associated with some series

of neural events. These examples show that some mental events have (p)physical properties, some essential and some accidental.

I now return to Levin's argument as it applies to the interaction of mental phenomena and (p)physical entities. According to this argument, a neural impulse is generated either by other neural impulses or spontaneously. In neither case does positing an immaterial object have any additional explanatory power. One problem with this argument is that it employs an inaccurate conception of a mental phenomenon: that a mental phenomenon is an immaterial object. An immaterial object is one with no physical properties. However, I have just shown that some mental phenomena have (p)physical properties. It is the conception of a mental event as a (purely) immaterial object which is employed in this argument against the possibility of interaction which gives the argument its plausibility since it does not seem possible for an immaterial object to act on a (p)physical entity. Once we realize that this conception of a mental phenomenon is a misconception, the argument loses its plausibility.

The conception of "physical entity" this argument employs is equally misleading. I have conceded that the argument has some merit as applied to interaction between mental phenomena and (s)physical and (o)physical entities. However, (p)physical entities have mental properties, so they are not purely (p)physical. Levin's argument against the possibility of mind-body interaction turns on the ability to draw a dichotomous distinction between purely mental

(immaterial) phenomena and purely physical (non-mental) entities. However, this dichotomy is a false one since some (p)physical entities have mental properties and some mental phenomena have (p)physical properties.

### 7.3 An Account of Mind-Body Interaction

According to Levin's argument as it applies to the interaction of (p)physical entities and mental phenomena, a neural impulse is caused either spontaneously or by other neural impulses. There is no need to postulate an immaterial object as a cause. However, if the other neural impulses that cause some neuron N to fire are themselves aspects of a mental phenomenon, then it is correct to attribute a causal role to the mental phenomenon. For example, in the case of intentional action, the neural patterns associated with my decision to move my leg initiate the neural activity necessary to carry out this action. This consists in sending signals to motor neurons. My thought that I will move my leg has a neural aspect and a phenomenal aspect, and the neural aspect of my thought initiates the neural activity required for moving my leg.

When a mental entity such as a thought acts on the brain or initiates a neural event, the neural aspect of the thought acts on some other neural aspect. This second neural aspect may or may not have an associated mental aspect. When someone is engaged in continuous thought, neural aspects associated with thought are

continuously generating other neural aspects associated with other thoughts. When someone forms an intention to raise his arm, on the other hand, the resulting neural activity may be a signal to the motor cortex to initiate arm-raising. This motor cortex activity need not have mental aspects.

We can picture this as follows:

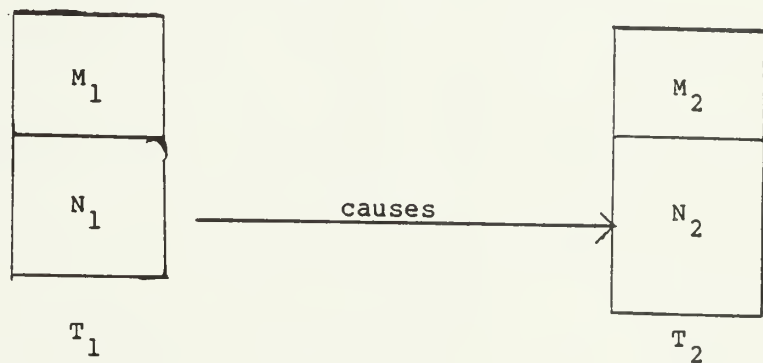


Figure 7-2. A Diagram of Mental-Physical Interaction

In the above diagram,  $T_1$  is a thought having mental aspect  $M_1$  and neural aspect  $N_1$ .  $T_2$  is a different thought having neural aspect  $N_2$  and mental aspect  $M_2$ . The neural aspect  $N_1$  initiates causal neural activity, e.g. by firing an action potential, resulting in neural state  $N_2$  which is a neural aspect of thought  $T_2$ . In this way, through its neural aspect, a thought can act directly on the brain. As I said above, there is no need for  $N_2$  to be an aspect of a mental event.

Neurologists still do not fully understand all the neurological mechanisms of voluntary movement, but there have been many studies.

One especially interesting account of several studies is given by Claude Ghez, and it bears full citation:<sup>15</sup>

In voluntary movements pyramidal tract and rubrospinal neurons start firing 20-50 msec before any sign of muscle contractions, but other groups of neurons fire earlier. The activity of neurons in the basal ganglia and the deep cerebellar nuclei precedes the activity of neurons in the motor cortex and red nucleus. During many voluntary movements activation of the basal ganglia and deep cerebellar nuclei represent the end result of complex processes that start hundreds of milliseconds earlier. Kornhuber and his collaborators in Germany have shown that changes in the potentials recorded from the human scalp over the motor cortex begin about 800 msec before a voluntary finger movement [Figure 7-3]. Although the mechanisms producing these potentials remain controversial, their occurrence underscores the fact that activity in the nervous system begins long before movement is actually performed. During this period the central nervous system is thought to set up a program which specifies how its neurons will respond to a stimulus.

Evarts and Tanji have investigated this program by recording the discharges of pyramidal tract neurons in monkeys that were trained to perform one of two different responses according to a preceding signal when a sudden perturbation was delivered to a handle that they were holding [Figure 7-4]. The instruction to either push or pull the handle was given by illuminating one of two lamps in front of the animal. The signal to move was a sudden jerk either toward or away from the animal, produced by a motor attached to the handle itself. Pyramidal tract neurons were recorded in the motor cortex and the direction of limb movement with which the unit increased its activity was also determined.

Two interesting observations emerged from these experiments. First, the instruction itself elicited different changes in activity in the neuron according to the direction of movement called for. The effects also differed according to the muscle groups whose contraction was associated with activity in that cell. For example, a pyramidal tract neuron that was phasically active when the animal flexed its arm showed tonic increases in activity after the instruction to pull (i.e., to contract the flexor muscle) was given. This change in neural activity was, however, not associated with any change in muscle activity. Second, the perturbation delivered to the handle, constituting the signal to move, elicited two distinct bursts of discharge in the same neurons. The



first burst, termed the reflex discharge, occurred whenever the biceps was stretched by sudden movement of the handle. This burst of activity is merely the result of the stimulus applied to the receptive field of the neuron, a reflection of a feedback system similar to that described by Asanuma. The later burst (appearing at a latency of about 50 msec), termed the intended discharge, was contingent on the instruction. The intended discharge occurred when the animal was instructed to pull (i.e., contract biceps); the discharge did not occur when the animal was to push. This intended discharge actually produces the intended response to the stimulus. The fact that its occurrence is dependent on context suggests that the sensory input can be gated through the pyramidal tract neurons by neuronal events set in motion by the instruction. The instruction is thought to put the nervous system in a specific state or voluntary set.

It is thought that the relationships between stimulus and response that are appropriate to a given context or instruction are established during an initial stage of learning. The learned relationship determines which muscle groups should contract and how much contraction should occur relative to the locus and the intensity of stimulation. Once learned, the instruction signal puts the nervous system into the appropriate set, which in turn operates on a neural gate and enables sensory inflow to be channeled to the proper muscles. One of the principal reasons for invoking a gating mechanism is that the interval between the handle jerk and the intended discharge of the cortical neurons is very brief. Recent experiments suggest that similar mechanisms operate in the quickest responses we make to any stimulus be it visual, cutaneous, or auditory.

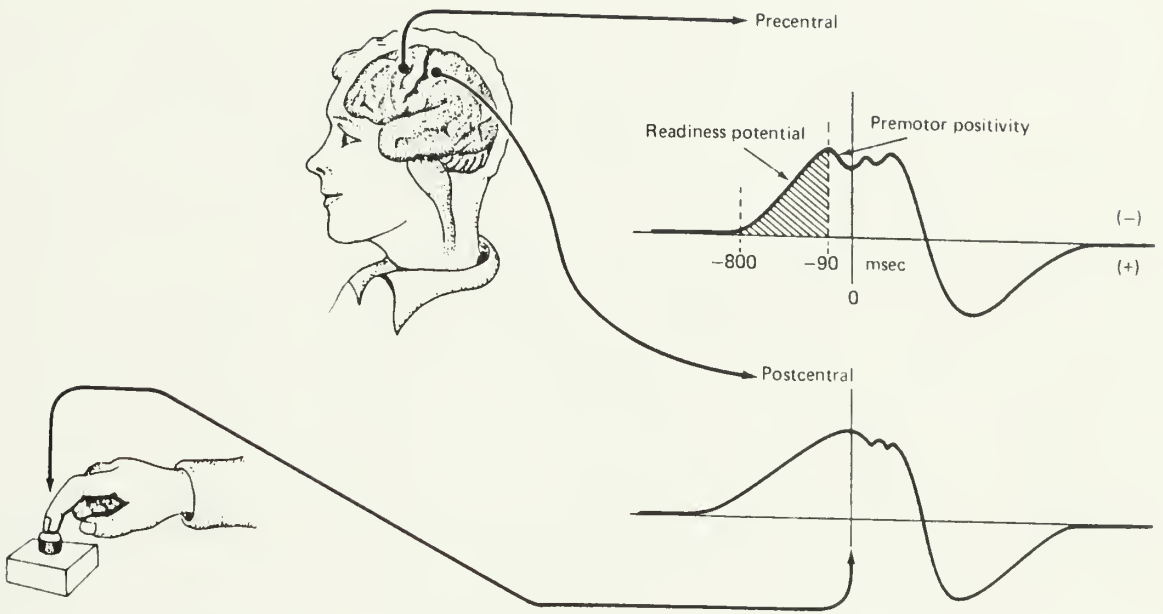


Figure 7-3. Potentials Recorded Before a Voluntary Finger Movement

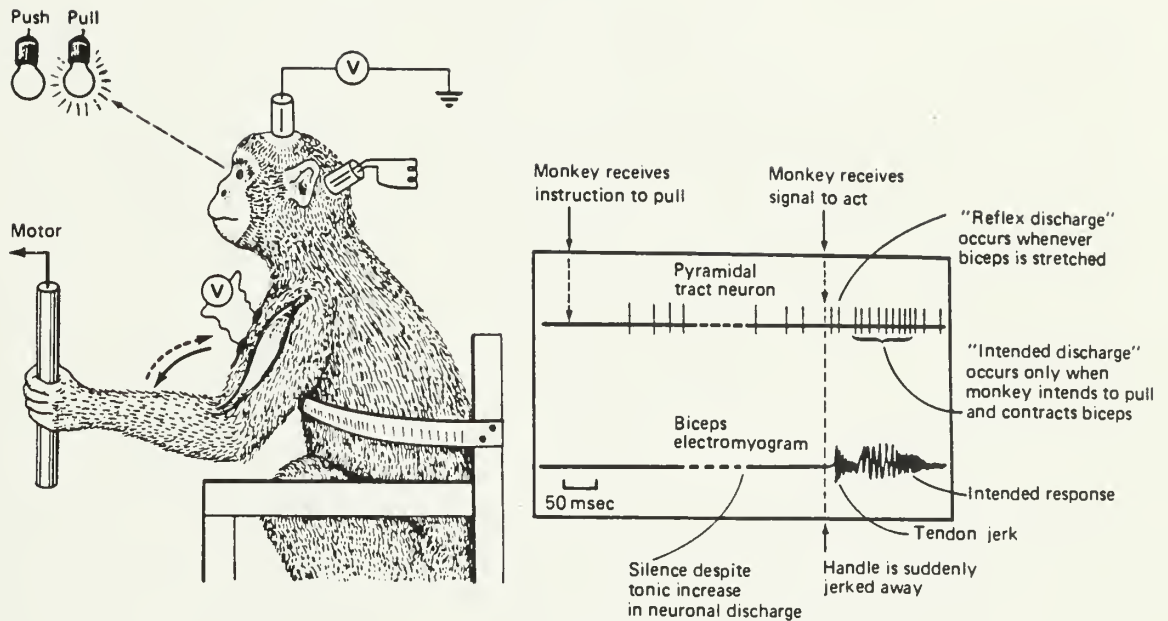


Figure 7-4. Cortical Activity Prior to Voluntary Movement

There are several points worthy of note about the preceding quotation. First, Ghez says "the instruction itself elicited different changes in activity in the neuron according to the direction of movement called for". This can be viewed as the neurological aspect of the understanding by the monkey of what he is

to do. Secondly, the burst of neural activity termed the intended discharge is the neurological aspect of the intention to pull the handle. Third, according to Ghez, learning "determines which muscle groups should contract and how much contraction should occur relative to the locus and intensity of stimulation. Once learned, the instruction signal puts the nervous system into the appropriate set". Ghez also cites the studies of Kornhuber et al., which show that "changes in the potential recorded from the human scalp begin about 800 msec before a voluntary finger movement". This activity may be regarded as the neurological aspect of the initial contemplation or consideration of the action (thinking about it).

The preceding study shows that thinking about something, forming the intention (deciding) to do it, and proceeding to do it is not an instantaneous action but is rather a complex neurological and intentional process which occurs over a period of time. Furthermore, it also shows that accomplishing such a task involves relying on neurological coding which takes place during earlier learning. By viewing these neurological aspects as aspects of a complex process containing neurological, intentional, phenomenal, and subjective aspects we can explain how thoughts and decisions can interact with neurological processes. If we instead view thoughts and decisions as completely distinct entities from neurological processes in the manner of dualism, we cannot explain interaction.

Complex patterns of neural states have phenomenal aspects and a person learns by experience what neural patterns produce what mental

phenomena and learns by experience how to produce these neural states. Speech and thought seems automatic until we try to speak another language or watch a child learning his first words. The commands which must be issued by the cerebral cortex to the mouth muscles are very complex and must be learned through trial and error and through imitation. There are "command" neurons in the cerebral cortex whose function is to issue motor commands and commands of other types to other elements of the nervous system. The signals to these "command" neurons originate in other areas of the cerebral cortex. A person can issue these signals by exerting a kind of internal effort, but we must learn what effort will produce what results. Learning to use our brains, like any other conscious activity, takes attention and experience.

It might be that the states  $N_1$  and  $N_2$  in the diagram in Figure 7-2 can be functionally defined. For example,  $N_1$  might be "the complex pattern of neurons which has as its phenomenal aspect the sound of the word 'Fire'". The brain then calls a routine producing  $N_2$  which could be "the complex pattern of neurons which has as its phenomenal aspect the sound of the word 'Run'". One could define other neural states according to their mental function, including neural states which issue certain specific signals to the "command" neurons in the cerebral cortex such as "Raise the right arm". In this way, one could give a functional analysis of the brain, not of the mind, and define these functional neural states by reference to mental phenomena. Such a view is not a reintroduction

of parallelism, but an account of mind-body interaction that uses functional techniques of analysis to show that mind and brain do not form a dichotomy but rather are aspects of one another.

The account of interaction I have given applies only to mental phenomena with (p)physical aspects. If there is a mental phenomenon with no associated (p)physical aspects, my account does not explain how such a phenomenon can interact with (p)physical entities. Because such a mental phenomenon would be one with no neural or (p)physical basis, it is doubtful whether there are any such mental phenomena. If one construes the phrase 'mental phenomenon' broadly so as to include phenomena such as sentence meanings (intensions) or particular features of mental experience such as the intensity of anger or fear, these intensions or mental features may not have a particular (p)physical entity or property associated with them. However, it is not clear that an intension can interact with a neural event, and in the case of the intensity of fear or anger the interaction is between the fear or anger itself and the neural event or entity; we do not have to say that every feature of the fear or anger also interacts with a neural event or entity.

The view I advocate is not a form of parallelism. According to a parallelist view, mental states and brain states are distinct and they occur parallel to each other. According to the view I advocate, neural and other (p)physical entities are properties or aspects of mental states and hence it is correct to attribute a causal role to the mental phenomenon in virtue of the causal action



of its neural aspect. We often attribute a property to a whole based on the action of one of its aspects. For example, we might say "The water in Reading makes me thirsty" when it is the sodium in the water that causes the thirst. Or, we might say "That car is making a lot of noise" when it is the car's exhaust system that is making the noise. In both of these examples we attribute a causal role to an entity in virtue of the action of one of its aspects. Likewise, the neural aspect of a mental phenomenon acts causally on the brain and we can attribute a causal role to the mental phenomenon in virtue of the action of this neural aspect. This view is not parallelism because a parallelist would deny that mental phenomena have (p)physical aspects and that (p)physical phenomena have mental aspects.

#### 7.4 Similar Accounts

The account of interaction I have just presented and the multi-aspect view it depends on is not entirely novel. Jack Ornstein develops a multi-aspect view of mind in his book The Mind and the Brain: A Multi-Aspect View. According to Ornstein, an adequate analysis of a subject's being in pain must take five aspects into account:<sup>16</sup>

An adequate analysis of the subject's being in pain, however, would have to take into account the following aspects or features: experiential, neural, bodily, behavioral and verbal.

Ornstein then describes each of these aspects:<sup>17</sup>

The experiential aspect: This is the felt quality or the sensation of pain--the "raw feel" as it has sometimes been called. It is what one experiences or feels when one is in pain. This is the feature of pain which leads us to call pain a 'mental' process. It is because this feeling or sensation could not exist unfelt or unsensed that we call it 'mental'. That is, it could not be truly said to exist unless it were sensed; "I am in pain but I am not aware of the pain" is self-contradictory. It could not exist unless it existed as an aspect of consciousness.

The neural aspect: These are the brain processes which we have granted to be invariably correlated and concomitantly variable with specific experiences. As neurophysiologists will readily admit, extremely little is known about how the brain operates and very little investigation has been conducted into the phenomenon of pain. However, it was thought that criticisms of, and an alternative to, the Identity Theory, would be more readily accepted if as much as possible were granted to it.

The firing of the neurons in question, then, has been conceded to be necessary and sufficient for the occurrence of the experience. If the appropriate neurons were not firing in X's brain, and the rest of his body, then he would not be having this experience.

Bodily aspects of pain: An example of this aspect would be X's decayed tooth. Of course a decayed tooth is not, properly speaking, an aspect of mind. It is, however, very often a feature of toothaches--states which are characteristically mental in the senses outlined above. Thus the bodily aspects of being in pain are part of our concept of pain. When someone is in pain we naturally assume that there is something wrong--some physiological irregularity due to damage or disease which is the source of the pain. We naturally assume it because there characteristically is something amiss physiologically when one is in pain.

The behavioral aspects: This is an aspect of mind. That is, we often do refer primarily to someone's behavior when we speak of his mind, e.g. "He has a good mind". However,...consciousness is a sufficient condition for the presence of a mind. If a person were conscious but unable to behave at all due to complete paralysis, we would still say that he had a mind. In fact, even if we allowed that a computer or a robot "behaved" in many ways as a man, we would not say that it had a mind if we know that it was not capable of being conscious.

Behavior, then, is often what we are referring to when we speak of someone's mind. But we usually are referring in such cases not merely to what the person did or is doing but also to what he is capable of doing.

The verbal aspect: This is not what X feels, not what is occurring in his brain, not what happens to him and not what he does--it is what he says. It may be objected that verbal behavior is just a species of behavior and, therefore, that it should not be classified separately.

I think, however, that the verbal aspect of pain merits separate classification. It is the one aspect of pain which is uniquely human. The other animals share with man all the other features of pain. Only man verbalizes his pain.

These, then, are the five aspects or features of being in pain--experiential, neural, bodily, behavioral and verbal. The first two aspects are the sine qua non of pain but some or all of the last three are characteristically present. Pain involves neither solely neural processes nor solely an experience. It is conceivable that someone could possess the latter of these features without the other but so far as we can tell this could never occur. Neither, of course, is pain solely behavior. Once it is learned that pain always has a neural component, any theory which identifies pain with actual and potential behavior is rendered untenable.

To say that one or the other of these five aspects is what pain (or the mind) really is, is to ignore the other four aspects. Identity theorists ignore all but the neural aspect. Cartesian Dualists deny all but the experiential aspect and Behaviorists deny all but the behavioral aspect.

Ornstein does not to my knowledge address the problem of mind-body interaction. However, his view that a mental phenomenon such as pain has a neural aspect is consistent with the view I present above. He also believes that mental phenomena have a bodily, a behavioral, and a verbal aspect. While I have not made this claim, and will not defend it here, such a claim is clearly in the spirit of the account I have given of mind-body interaction. Ornstein's view is similar to the view I present in that he rejects

the claim that the mind-body distinction is a dichotomy and recognizes that mental phenomena have (p)physical (neural and bodily) aspects.

G.D. Wassermann presents an account of mind-body interaction that resembles the one I give. Interestingly, Wassermann takes himself to be defending a view he calls materialistic dualism. Wassermann states this view as follows:<sup>18</sup>

Materialistic dualism is a metaphysical doctrine, which hypothesizes that in addition to physical properties some or conceivably all components of matter are endowed with mental properties, and that the conceptual modes of symbolizing physical properties of matter could never suffice for symbolizing its mental properties.

Wassermann states his view of interaction as follows:<sup>19</sup>

An antimaterialist cynic might ask how anyone could theorize about brain representations of, say, visual images, without having consciously experienced such an image, so that his conscious experience of images has a bearing on his theorizing. To this one can simply reply that it is our physical brain representations of theories--and not our conscious experiences of visual images--that lead us to theorize about brain representations. Let us also return to the apparent dilemma, mentioned earlier in this section. It concerned the problem of how people could claim to be consciously aware of something without their conscious awareness "causing" the claim to be made. This, however, poses no problem for the materialist when it is recognized that the conscious awareness of anything could be associated with a specific physical state of a particular brain structure....This awareness-related brain state could then become associated with appropriate brain representations that encode verbal claims, leading thereby to a verbal claim of a conscious awareness. In this case, however, it is not the awareness that "causes" the activation of the brain representation of the claim of that awareness. Instead it is the brain representation that is associated with the conscious awareness that causes the brain representation of the claim of that awareness to be activated.

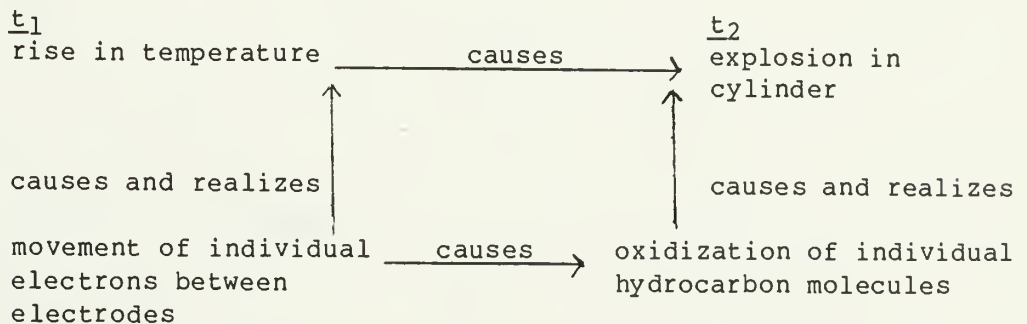
Here Wassermann states a view very similar to the one I advance above. Instead of supposing that conscious awareness considered as a purely mental phenomenon with no physical aspects causes someone to make a verbal claim, he supposes that the awareness-related brain state becomes associated with brain representations that encode verbal claims, leading to a verbal claim of conscious awareness. Wassermann does not make the further claim that the brain state is an aspect of conscious awareness as I do, but still his account is quite similar to the one I present.

In his recent book, Intentionality,<sup>20</sup> John Searle presents an account of interaction that resembles the one I have given. Searle makes use of the notion of "levels of description" in stating his view:<sup>21</sup>

But this model of "caused by" and "realized in" only raises the next question, how can Intentionality function causally? Granted that Intentional states can themselves be caused by and realized in the structure of the brain, how can Intentionality itself have any causal efficacy? When I raise my arm my intention in action causes my arm to go up. This is a case of a mental event causing a physical event. But, one might ask, how could such a thing occur? My arm going up is caused entirely by a series of neuron firings. We do not know where in the brain these firings originate, but they go at some point through the motor cortex and control a series of arm muscles which contract when the appropriate neurons fire. Now what has any mental event got to do with all of this? As with our previous questions, I want to answer this one by appealing to different levels of description of a substance, where the phenomena at each of the different levels function causally; and as with our previous question I want to make clear the relations involved by considering completely banal and unproblematic examples. Consider the explosion in the cylinder of a four-cycle internal combustion engine. The explosion is caused by the firing of the spark plug even though both the firing and the explosion are caused by and realized in phenomena



at a microlevel, at which level of description terms like "firing" and "explosion" are entirely inappropriate. Analogously I want to say that the intention in action causes the bodily movement even though both the intention in action and the bodily movement are caused by and realized in a microstructure at which level terms like "intention in action" and "bodily movement" are inappropriate. Let us try to describe the case a little more carefully--and again it is not the particular case or its details that matter but the type of relations that are exemplified. The aspect of the spark plug firing which is causally relevant is the rise in temperature in the cylinder between the electrodes to the kindling point of the air fuel mixture. It is this rise in temperature which causes the explosion. But the rise in temperature is itself caused by and realized in the movement of individual particles between the electrodes of the spark plug. Furthermore the explosion is caused by and realized in the oxidization of individual hydrocarbon molecules. Diagrammatically it looks like this:

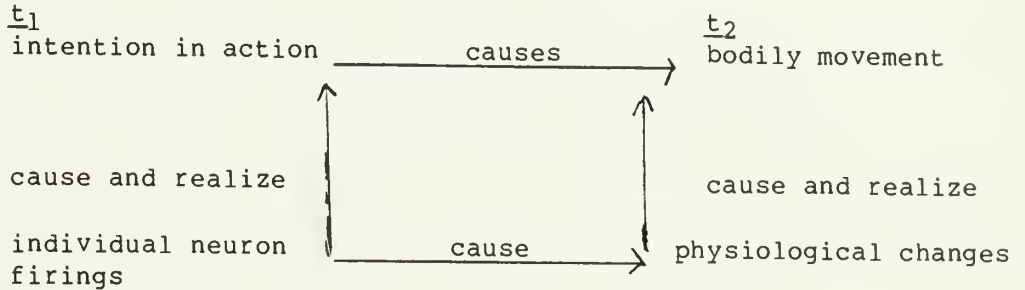


The phenomena at  $t_1$  and  $t_2$  respectively are the same phenomena described at different levels of description. For that reason we could also draw diagonal arrows showing that the movement of the electrons causes the explosion and the rise in the temperature causes the oxidization of hydrocarbon molecules.

Though we know very little about how intentional action originates in the brain we do know that neural mechanisms stimulate muscle movements. Specifically they stimulate calcium ions to enter into the cytoplasm of a muscle fiber, and this triggers a series of events that result in the movement of the myosin cross bridges. These cross bridges connect myosin filaments to actin filaments. They alternately attach to actin strands, exert pressure, detach, bend back, reattach and exert more pressure. This contracts the muscle. At the microlevel then we have a sequence of neuron firings which causes a series of



physiological changes. At the microlevel the intention in action is caused by and realized in the neural processes, and the bodily movement is caused by and realized in the resultant physiological processes. Diagrammatically it is formally similar to the diagram of an internal combustion engine:



Notice that on this model, as with our earlier model, we could also draw diagonal arrows which in this case would show that the intention in action causes physiological changes and that the neuron firings cause bodily movements. Notice also that on such a model the mental phenomena are no more epiphenomenal than the rise in temperature of the firing of a spark plug.

Let us now return to our 'dilemma'. The first horn claims that if we think of the relation between the mental and the physical as causal we are left with a mysterious notion of causation. I have argued that this is not so. It only seems so if we think of mental and physical as naming two ontological categories, two mutually exclusive classes of things, mental things and physical things, as if we lived in two worlds, a mental world and a physical world. But if we think of ourselves as living in one world which contains mental things in the sense in which it contains liquid things and solid things, then there are no metaphysical obstacles to a causal account of such things. My beliefs and desires, my thirsts and visual experiences, are real causal features of my brain, as much as the solidity of the table I work at and the liquidity of the water I drink are causal feature of tables and water.

Searle's account appeals to the notion of "different levels of description of a substance". On his view, a description of the neural state associated with an intention occurs at a different

level of description than a description of the intention. I discussed the notion of a level of description in Chapter 4 in the context of a discussion of Dennett.<sup>22</sup> There I was not able to provide a very complete account of this notion. However, Searle's use of this phrase seems very similar to my use of the phrase 'conceptual domain', which derives from Davidson. I see no reason why Searle's claim could not be formulated as a claim about different conceptual domains, with a unique level of description being roughly equivalent to a unique conceptual domain. If this is so, his account of interaction is very similar to the one I have presented with the (p)physical (neural) aspect of a mental phenomenon causing other (p)physical (neurological) changes. Searle's view of the relation between a neural state and an intention is that individual neural firings cause and realize the intention. While I have not discussed this relation in detail, Searle's view is a promising way to analyze this relation.

Searle also criticizes the view that 'mental' and 'physical' name two ontological categories that are mutually exclusive. He appears to endorse the view that a mental phenomenon together with its neural aspect form a single substance. On this view, the neural aspect of an intention is as much a part or element of an intention as its phenomenal aspect. It is the neural aspect of the intention that acts causally on the brain. By rejecting the conception that the mental-physical distinction forms a mentally exclusive dichotomy, such a view can account for mind-body interaction.

### 7.5 Viewpoint Pluralism

The account of mental and physical phenomena and of mind-body interaction I present preserves the insights of dualism, materialism, and functionalism while rejecting features of those theories which make them untenable in their current form. The view I present preserves the insight of dualism that there is a mental-physical distinction, though this distinction is not a dichotomy and is a distinction of aspects and conceptual domains. It preserves the insight of materialism that neural activity plays a causal role in mental phenomena. It preserves the insight of functionalism that it is useful to take a functional view of the brain. According to the view I advocate, brain states are defined by their mental function. Because the key feature of this view is its plural aspect analysis of mental and physical phenomena, it is a plural aspect view.

I have not discussed pluralism of entities or aspects in this or in preceding chapters in any detail. However, in Chapter 1 I took note of Spinoza's plural attribute view and Leibniz's pluralism of viewpoints. Spinoza's use of the term 'attribute' was similar to our contemporary use of the term 'property'. I have spoken of pluralism of aspects, which seems to introduce still another term to describe the same phenomenon. However, the term aspect does have a unique meaning apart from 'attribute' and 'property'. The term

'aspect', unlike either 'attribute' or 'property', contains a reference to the idea that something is perceived from a particular view. An aspect of  $x$  can be defined as a property of  $x$  as perceived from a point of view:

- V.  $A$  is an aspect of  $x$  just in case  $A$  is a property of  $x$  as perceived from a particular point of view.

This definition of 'aspect' provides a link between Spinoza's plural attribute (property) view and Leibniz's pluralism of viewpoints. If an aspect of  $x$  is a property of  $x$  as seen from a particular point of view, then it should be possible to perceive different aspects of an entity by changing points of view of the entity. For example, changing one's point of view of a house allows one to perceive its different sides, its different colors, and both its outside and inside. Conversely, an entity that has many aspects is one that can be perceived from many points of view. Principle V thus allows us to provide a link between a plural aspect view and pluralism in points of view.

I cannot fully explicate the link between aspects and points of view here. However, there is another connection between the notion of viewpoint and the concept of a conceptual domain as Davidson introduces it and as I have used it in preceding chapters. I stated that the (s)physical domain includes at least the following elements:

1. Logical terms ('and', 'or', 'if', etc.)
2. Mathematical terms ('number', 'line', 'point', etc.).  
The mathematics chosen can be governed by pragmatic considerations.
3. Syntactic rules.
4. (s)physical terms which are descriptive. These can be nouns, verbs, adjectives, or adverbs.

I gave similar definitions of the (p)physical and (o)physical conceptual domains. I argued that these constitute three distinct conceptual domains.

My definition of the (s)physical domain may appear to be incomplete in that it does not contain a specification of the meaning of the term '(s)physical'. In Chapter 6 I defined the notion of an (s)physical term as follows:

T(s). 't' is an (s)physical term just in case 't' refers to some (s)physical property or (s)physical entity.

I defined the notions of "(s)physical property" and "(s)physical entity" as follows:

P(s). x is a purely (s)physical property = df. x is a property and x is uniquely specifiable in terms from only the languages of physics and chemistry.

P(s)'. x is a uniquely (s)physical property = df. x is a property and x is partially but not completely specifiable in terms from only the languages of physics and chemistry.

E(s). x is a purely (s)physical entity = df. x is an entity and all of x's essential properties are purely (s)physical properties.

E(s)'. x is an impurely (s)physical entity = df. x is an entity and at least one but not all of x's essential properties are (s)physical properties.

The above five definitions are implicit in the specification of the conceptual domain I gave above (conditions 1 - 4) in that these conditions implicitly rely on a conception of what an '(s)physical' term is even though this conception is not explicitly stated. If we conjoin conditions 1 - 4 with the above five definitions we have a more complete specification of the (s)physical conceptual domain. This specification explicitly states what an (s)physical term is. One way to describe this is by saying that it formulates a membership condition for the class of (s)physical terms.

It should be possible to state such a membership condition for every conceptual domain that can be specified in a way that parallels my specification of the (s)physical domain. My specification of this domain relies on the notion of (s)physical terms. It should be possible for any parallel specification of a conceptual domain to provide definitions that state the meaning of a term which is relied on in this way. Thus, I also give definitions of '(p)physical' and '(o)physical'. In light of the preceding paragraph, we can say that it should be possible to specify a membership condition for any conceptual domain that can be specified in a way that is parallel to my specification of the (s)physical domain.

When a conceptual domain includes a specification of a membership condition as specified in the preceding paragraph, I will say that this conceptual domain expresses a point of view. The point of view of the (s)physical domain is that some entities have



properties that can be uniquely specified in terms from only the languages of chemistry and physics. While I cannot fully defend or explicate this notion of a point of view and its relation to conceptual domains in this chapter, I believe I have said enough to give a sketch of a position.

If my argument that there are at least three distinct conceptual domains is correct, and I am correct in saying that each of these conceptual domains can be specified in such a way as to formulate a unique point of view, it follows that there are at least three distinct physical points of view in terms of which the world can be viewed. These, together with the mental conceptual domain (point of view), constitute four points of view of the world. While I believe this argument can be generalized to generate a bolder, more full-blooded ontological and conceptual pluralism, I content myself here with the conclusion that from these four points of view one can perceive at least four irreducibly distinct aspects of the world. I call this view Viewpoint Pluralism.

## 7.6 Summary

I began this chapter by considering an argument by Michael Levin advanced in Chapter 6 against the possibility of mind-body interaction. Using concepts developed in the second half of Chapter 6, I conceded that Levin's argument may provide an argument against the possibility of interaction between mental phenomena and

(s)physical entities and between mental phenomena and (o)physical entities. However, I argued that Levin's argument does not show the impossibility of interaction between mental phenomena and (p)physical entities. The basis for this claim is in part that some mental phenomena have (p)physical properties or aspects and that some (p)physical entities have mental properties or aspects. I then gave an account of mind-body interaction according to which the neural aspect of a mental phenomenon interacts causally with the distinct neural aspect of a distinct mental phenomenon. I then cited three writers who have presented views similar to the one I advocate: Jack Ornstein, G.D. Wassermann, and John Searle. I concluded by citing the relation between the view I advocate and dualism, materialism, and functionalism. I then briefly stated a more general theory that links a plural aspect view with Leibniz's pluralism of viewpoints. I call this theory Viewpoint Pluralism.

## 7.7 Conclusion

In Chapter 1 I presented a historical survey of the mind-body problem. In Chapter 2 I defended Saul Kripke's dualistic views against criticisms by Fred Feldman. In Chapter 3 I discussed Donald Davidson's anomalous monism. While I later utilize some notions Davidson develops such as that of a conceptual domain, I argued that Davidson has not established anomalous monism and that there are

good reasons to doubt that it is true. In Chapter 4, I criticized Daniel Dennett's defense of functionalism.

In Chapter 5 I discussed two criteria for identifying mental phenomena. The first, incorrigibility, is advocated by Richard Rorty. The second, intentionality, is advocated by Roderick Chisholm. In the first half of Chapter 6 I discussed the problem of mind-body interaction and concluded that a conceptual revision in the notions of "mental" and "physical" is necessary before an account of interaction can be given. In the second half of Chapter 6, I attempted such a revision by distinguishing three conceptions of the physical. I called these conceptions the scientific ((s)physical) conception, the objective ((o)physical) conception, and the personal ((p)physical) conception. In Chapter 7 I returned to an argument against the possibility of mind-body interaction I discussed in Chapter 6 and showed how distinguishing the three conceptions of the physical helps given an account of interaction. According to this account, the neural aspect of a mental entity acts causally on the distinct neural aspect of a distinct mental entity. I concluded with a brief statement of Viewpoint Pluralism.

## FOOTNOTES

1. Michael E. Levin, Metaphysics and the Mind-Body Problem (Oxford, England: Clarendon Press, 1979), pp. 84-6.
2. Refer to the quotation from Noback and Demarest in Chapter 6, Section 6.2.
3. Refer to Chapter 6, Section 6.2.
4. Levin, *Ibid.*, p. 84.
5. By Yuri Gellner and others. I have not discussed alleged cases of mind moving matter, extra sensory perception, or other forms of psychic phenomena anywhere in these chapters not because I think that nothing can be learned from them but because those who are predisposed not to accept their validity are usually not convinced by these phenomena. On the other hand, many of those who are predisposed to accept the validity of psychic phenomena accept their validity uncritically.
6. Refer to Chapter 3, Section 3.2.
7. Eric R. Kandel, "Visual System III: Physiology of the Central Visual Pathways", in Principles of Neural Science, ed. Eric R. Kandel and James H. Schwartz (New York: Elsevier/North Holland, 1981), p. 237.
8. *Ibid.*, p. 240, pp. 243-5, and pp. 245-7.
9. Peter Gouras, "Visual System IV: Color Vision", in Kandel and Schwartz, *op. cit.*, p. 253.
10. *Ibid.*
11. Refer to Chapter 7, Section 7.2.1.
12. Kandel, *op. cit.* p. 247.
13. Gouras, *Ibid.*
14. Refer to Chapter 3, Section 3.4.
15. Claude Ghez, "Cortical Control of Voluntary Movement", in Kandel and Schwartz, *op. cit.*, pp. 331-3.
16. Jack Ornstein, The Mind and the Brain A Multi-Aspect Interpretation (The Hague: Martinus Nijhoff, 1972), p. 122.

17. Ibid., pp. 122, 130, 132, 134, 138, 141.
18. G. D. Wassermann, "Materialism and Mentality", Review of Metaphysics 35 (June, 1982), p. 716.
19. Ibid., pp. 721-22.
20. John Searle, Intentionality (New York: Cambridge University Press, 1983).
21. Ibid., pp. 268-71.
22. Refer to Chapter 4, Section 4.2.1.

## BIBLIOGRAPHY

### ARTICLES

Arbib, Michael A., "Consciousness: the Secondary Role of Language", Journal of Philosophy, 69, pp. 579-91, 1972.

Block, Ned J., and Fodor, Jerry, "What Psychological States are Not", The Philosophical Review, 82, pp. 159-81, 1972.

Charlton, William, "Aristotle's Definition of Soul", Phronesia, 25, pp. 170-86, 1980.

Churchland, Patricia Smith, "A Perspective on Mind-Brain Research", Journal of Philosophy, 77, pp. 185-207, April, 1980.

Cooper, W. E., "Four Anti-Materialist Positions", Philosophical Forum, 11, pp. 103-11, Winter, 1979.

Davidson, Donald, "Mental Events", in Foster, Lawrence and Swanson, Joe W. (eds.), Experience and Theory (Amherst, Massachusetts, 1970).

Dennett, Daniel C., "Current Issues in the Philosophy of Mind", American Philosophical Quarterly, 4, pp. 249-61, October, 1978.

Feigl, Herbert, "The 'Mental' and the 'Physical'", in Feigl, Herbert, and Scriven, Michael, and Maxwell, Grover (eds.), Minnesota Studies in the Philosophy of Science, Vol. 2 (Minneapolis, Minn., 1958).

Feldman, Fred, "Kripke and the Identity Theory", Journal of Philosophy, 71, pp. 665-76, 1974.

Gunderson, Keith, "Content and Consciousness and the Mind-Body Problem", Journal of Philosophy, 69, pp. 591-604, 1972.

Johnston, Patricia, "Origin and Necessity", Philosophical Studies, 32, pp. 413-8, November, 1977.

Kraemer, Eric Russert and Sayward, Charles, "Dualism and the Argument from Continuity", Philosophical Studies, 37, pp. 55-59, January, 1980.



Kripke, Saul, "Naming and Necessity", in Davidson, Donald and Harman, Gilbert (eds.), Semantics of Natural Language, (Dordrecht, 1972).

Lycan, William G., "On Intentionality and the Psychological", American Philosophical Quarterly, 6, pp. 305-12, 1969.

Lycan, William G., "Mental States and Putnam's Functionalist Hypothesis", Australasian Journal of Philosophy, 52, pp. 48-62, 1974.

Lycan, William G., "Kripke and the Materialists", Journal of Philosophy, 71, pp. 677-89, 1974.

Nagel, Thomas, "Physicalism", The Philosophical Review, 74, pp. 339-56, 1965.

Nagel, Thomas, "What is it Like to be a Bat?", The Philosophical Review, 83, pp. 435-50, 1974.

Pribham, Karl H., "Unpacking Some Dualities in a Mind/Brain Dualism", Philosophy of the Social Sciences, 10, pp. 295-308, Spring, 1980.

Putnam, Hilary, "Minds and Machines", in Hook, Sidney (ed.), Dimensions of Mind (New York, 1960).

Putnam, Hilary, "Robots: Machines or Artificially Created Life", Journal of Philosophy, 61, pp. 668-91, 1964.

Puntam, Hilary, "The Mental Life of Some Machines", in Castaneda, Hector-Neri (ed.), Intentionality, Minds, and Perception (Detroit, 1967).

Rorty, Richard, "Functionalism as the Mark of the Mental", Journal of Philosophy, 18, pp. 237-49, Summer, 1980.

Rorty, Richard, "Incorrigibility as the Mark of the Mental", The Journal of Philosophy, LXVII, 12 (June 25, 1970): pp. 399-424.

Tietz, John, "Davidson and Sellars on Persons and Science", Southern Journal of Philosophy, 18 pp. 237-49, Summer, 1980.

# BOOKS

- Anderson, Alan R. (ed.), Minds and Machines (Englewood Cliffs, New Jersey, 1964).
- Block, Ned, Readings in Philosophy of Psychology (Harvard University Press, 1980).
- Boden, Margaret, Artificial Intelligence and Natural Man (Hassocks, England, 1977).
- Borst, C. V., (ed.), The Mind/Brain Identity Theory (New York: St. Martin's Press, 1975).
- Broad, C. D., The Mind and Its Place in Nature (London: Kegan Paul, Trench, Trubner, & Co., Ltd., 1947), esp. pp. 95-133.
- Chisholm, Roderick M., Perceiving: A Philosophical Study (Ithaca, New York: Cornell University Press, 1957).
- Davidson, Donald, Essays on Actions and Events (Oxford: Clarendon Press, 1980).
- Dennett, Daniel C., Content and Consciousness (London, 1969).
- Dennett, Daniel C., Brainstorms (New York, 1979).
- Descartes, Rene, The Philosophical Works of Descartes, trans. Elizabeth S. Haldane and G. R. T. Ross, Vol. I and II (New York, Cambridge University Press, 1972).
- Ducasse, C. J., The Belief in a Life After Death (Springfield, Illinois: Charles C. Thomas, 1961).
- Fodor, Jerry, The Language of Thought (New York, 1975).
- Kandel, Eric R. and Schwartz, James H. (ed.), Principles of Neural Science, (New York: Elsevier/North Holland: 1981).
- Levin, Michael, Metaphysics and the Mind-Body Problem (Oxford, England: Clarendon Press, 1979).
- Ornstein, Jack, The Mind and the Brain A Multi-Aspect Interpretation (The Hague: Martinus Nijhoff, 1972).
- Putnam, Hilary, Mind, Language and Reality, Philosophical Papers (Cambridge: Cambridge University Press, 1975).

Ringle, Martin (ed.), Philosophical Perspectives in Artificial Intelligence (Humanities Press, 1979).

Rorty, Richard, Philosophy and the Mirror of Nature (Princeton: Princeton University Press, 1979).

Rosenthal, David M. (ed.), Materialism and the Mind-Body Problem (New Jersey: Prentice Hall, 1971).

Searle, John, Intentionality (New York: Cambridge University Press, 1983).

Snell, Bruno, The Discovery of the Mind (Cambridge, Massachusetts: Harvard University Press, 1953).

Strawson, Peter F., Individuals (London, 1959).

Wittgenstein, Ludwig, Philosophical Investigations (Oxford, 1953).



