

PHYSICALISM, INTENTIONALITY, MIND:  
THREE STUDIES IN THE PHILOSOPHY OF MIND

by

STEPHEN PRIOR

B.A., COLUMBIA COLLEGE

1968

SUBMITTED IN PARTIAL FULFILLMENT  
OF THE REQUIREMENTS FOR THE  
DEGREE OF  
DOCTOR OF PHILOSOPHY  
at the  
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 1977

Signature of Author . . . . .  
*Stephen Prior* Department of Linguistics and Philosophy

Certified by . . . . .  
*[Signature]* Thesis Supervisor

Accepted by . . . . .  
*[Signature]* Chairman, Departmental Committee



PHYSICALISM, INTENTIONALITY, MIND:  
THREE STUDIES IN THE PHILOSOPHY OF MIND

by  
STEPHEN PRIOR

Submitted to the Department of Philosophy of Linguistics  
and Philosophy on May 5, 1977, in partial fulfillment of  
the requirements for the Degree of Doctor of Philosophy

ABSTRACT

This thesis consists of three related but fundamentally distinct parts. They will be summarized separately.

"Physicalism and Reduction" argues that it is useful and historically plausible to consider the physicalist program in the philosophy of mind as the attempt to show that all nomic properties are physical properties. (Nomic properties are those mentioned in the laws of any science.) Three more familiar doctrines can be seen to derive from this broad definition. Inter-theoretic reduction is in part an attempt to show that the properties mentioned in, for example, psychological law are coextensive with physical properties, and hence dispensable from science. The mind-body identity theory, as originally formulated, is just the claim that the psychological properties are (identical to) physical properties. Logical behaviorism can be understood as an attempt to show, without inter-theoretic reduction, that the putatively mental properties of psychology are physical properties because the predicates ascribing psychological properties are synonymous with predicates ascribing physical properties. These three are all shown to be quite different from the more modest claim that all events are physical events. These distinctions from the historical background for Fodor's new model of the relationship of the sciences, which in turn permits rigorous formulation and support for anti-reductionist theses common in the literature.

"Dennett on Intentionality" argues that D. C. Dennett's attempt to give a reductionistic, neurological account of intentionality is a failure. The notion of reduction Dennett suggests is too weak to eliminate intentional states or properties from psychology. His argument to the effect that intentional psychology is non-empirical or fictional is shown to contain serious errors, and, even if sound, to fail to demonstrate that there is some distinctive problem with

intentional psychology that does not pertain to other special sciences. His argument that intentional psychology is question-begging is shown to rest on the familiar and illicit inference from the denial of substance dualism to the assertion that there must be reductionist accounts of mental states. It is suggested that Quine's more radical strategy of accepting the irreducibility of intentionality yet denying it scientific status (ultimately by reason of indeterminacy) is perhaps now the only move a true physicalist can make.

"Other Minds and the Argument from Analogy" argues that those who use the argument from analogy in fact accept the basic portrait of mental states propounded by other minds skeptics. It is suggested that the dissatisfaction with the argument from analogy lies just in that acceptance of a dubious theory of our knowledge of the mental states of others. Some typical skeptical arguments about privacy and observation are examined and rejected. Strawson's thesis that behavior is an intersubjectively available appearance of a mental state is favorably received and augmented with an argument to the effect that if mental states are not observable neither are many of the states which comprise animate behavior. A limited privacy of sense qualia is upheld, but the skeptical portrait of mental states in general is rejected.

Thesis Supervisor: J. A. Fodor, Professor of Philosophy and Psycholinguistics

## TABLE OF CONTENTS

|  |     |
|--|-----|
| Title Page . . . . .                           | 1   |
| Abstract . . . . .                             | 2   |
| Table of Contents . . . . .                    | 4   |
| Acknowledgments . . . . .                      | 5   |
| <br>   |     |
| PHYSICALISM AND REDUCTION . . . . .            | 6   |
| Footnotes . . . . .                            | 38  |
| Bibliography . . . . .                         | 42  |
| <br>   |     |
| DENNETT ON INTENTIONALITY. . . . .             | 45  |
| Footnotes . . . . .                            | 89  |
| Bibliography. . . . .                          | 90  |
| <br>   |     |
| OTHER MINDS AND THE ARGUMENT FROM ANALOGY. . . | 92  |
| Footnotes . . . . .                            | 132 |
| Bibliography. . . . .                          | 134 |
| <br>   |     |
| Biographical Sketch . . . . .                  | 136 |

## ACKNOWLEDGMENTS

There are many people without whose assistance this thesis would never have reached completion. I wish to thank Jerry Fodor for the depth and insight of his criticisms. Ned Block commented carefully on every draft and taught me a good deal of philosophy in the process. I am especially indebted to James Thomson for helping me to understand my own thoughts better.

The friends and relations who provided moral support are too numerous to mention; but I do especially want to thank Sylvain Bromberger for years of advice and encouragement. And most of all, I wish to thank my parents for all the many ways they have helped me during my work on this thesis.

PHYSICALISM AND REDUCTION

## What is Physicalism? Preliminary Characterization

Physicalism is perhaps best understood as a family of related (and sometimes obscurely related) claims and arguments. In works on physicalism or on issues said to bear on the truth or falsity of physicalism, one finds discussions and claims about ontology, the unity of the sciences, mechanism, teleology, vital forces, organic wholes, irreducible properties, identity, explanation, the nature of psychology, the meaning of psychological predicates, the future of science, et al. It is not my intent to unravel all the issues and display their proper relationship (or lack thereof). I wish to consider three specific theses:

(A) Let  $M_i$  be a mental state; let  $P$  range over physical states. Then, for every  $M_i$

$$(x) \left[ \text{if } x \text{ is a token of } M_i, \text{ then } (\exists P)(\exists y) \right. \\ \left. (y \text{ is a token of } P \text{ and } x = y) \right]^1$$

(B) For every mental state  $M_i$  there is a physical state  $P_i$  such that state  $M_i$  is identical to physical state  $P_i$ .

(C) Psychology is reducible to neurology.

## Some Basic Clarifications

(A) and (B) are two quite distinct versions of an identity theory of mind and body. (A) asserts the identity of the events which are tokens of various state types, whereas (B) asserts the identity of certain states. That is, (A) asserts that when a particular organism is in some particular mental state at time  $t_i$ , the having of this mental state  $M_i$  at  $t_i$  is an event which is the organism's being in some physical state or other. (B) asserts something much stronger, viz., that  $M_i$  and  $P_i$  are the same state. It does not assert that to say of something that it is in state  $M_i$  is

synonymous with saying that it is in state  $P_i$ , but that the property denoted by ' $M_i$ ' is the property denoted by ' $P_i$ '. Thus the essential difference between (A) and (B) is that (B) requires the identity of the denotatum of every universal or type term of mental states with the denotatum of some universal or type term of physical states. (A) can be read as asserting that all events are physical events, whereas (B) asserts that for every mental type predicate such as '. . . is in state  $M_i$ ' there is a physical type predicate '. . . is in state  $P_i$ ' such that both predicates ascribe the same property. Hence (B) asserts that there are no properties which are non-physical because they are mental. Or, to state it positively, (B) says that all properties used to characterize mental states and events are physical properties.

Which properties are physical properties? Let us begin by noticing that on the strictest interpretation, physical properties are those used in the laws of physics (or some standard formulation thereof). But this seems unduly restrictive, for it is quite possible that what many would call a physicalistic account of mental events and behavior could obtain without the state types of the relevant science being state types of physics. Therefore let us provisionally take 'physical property' to mean 'property used to define a state type present in the laws of any of the sciences of the inorganic and biology.'

Which properties are mental properties? A workable answer is those properties mentioned in the laws of psychology (including psycho-physiology), and mentioned in the laws of no other science. With this criterion we can reformulate (B) as

(B') All nomic properties of psychology are physical properties

There may be other criteria than nomicity in psychology for determining which are the mental properties, but I prefer this one because it is clear and because it generates a realistic goal for physicalism. The demonstration of (B') would be a remarkable achievement, not one that should be dismissed on the grounds that some non-nomic mental properties had not been shown to be physical properties.

Since properties are difficult to inspect, let us turn our attention to predicates. Let us say that a predicate is nomic if it occurs in some good formulation of the laws of some science. The special sciences can be roughly distinguished by the families of nomic predicates they use. (We have to say 'roughly' because we cannot differentiate laws, theories, and scientific domains by predicates alone. For example, identity relations make it difficult to differentiate laws properly: are  $E = h\nu$  and  $E = h\frac{f}{c}$  indeed two different laws because the predicates are distinct? It hardly seems likely, for frequency and wave length are not the same property.<sup>2</sup> This problem in differentiating laws with predicates migrates to the problem of differentiating theories and domains as well, if we attempt (as seems reasonable) to differentiate domains by the theories, and theories by the laws they contain. And additional difficulties arise: many laws and predicates are used by more than one special science, as in the use of what we usually call the laws of physics by chemistry. Nonetheless, we do have some usable intuitions. No one is going to think that '... believes that p' is prima facie a nomic predicate of physics or chemistry.) If we have some notion of which are the predicates of psychology we can say that those predicates which are nomic in

psychology and not elsewhere are the mental predicates. Then (B') can be reformulated as

(B'') All nomic predicates of psychology denote physical properties

Now thesis (B'') has a task which (A) does not. It must attempt to show that all the mentalistic predicates (such as 'believes that p') which, as seems likely, will be nomic predicates of psychology, denote physical properties. To prove (B'') one must show that the prima facie distinctness of mental predicates (e.g., the fact that only mental predicates are intentional) is not reason to believe that the properties denoted by these predicates are non-physical properties. The physicalist must show that all such predicates denote physical properties; I wish to call this program the physicalization of mentalistic predicates. It should be obvious that neither intentional nor phenomenal predicates appear to denote physical properties and hence that special account of them is owed us by physicalism.

It might well be thought that (A) will ultimately entail (B') and (B''). For it is at least plausible to follow Brandt and Kim's suggestion (Brandt & Kim, 1967) that we define an event token as the instancing of a property at a certain time and at a certain place. In this case, an event token identity thesis for mental and physical events will be stated as the generalization of the schema

$$(1) \quad (M_i, t_i, L_i) = (P_j, t_j, L_j) \text{ iff. } \begin{aligned} L_i &= L_j \\ t_i &= t_j \\ M_i &= P_j \end{aligned}$$

where  $L_n$  are locations,  $t_n$  are times, and  $M_n$  are mental properties and  $P_n$  are physical properties. But this criterion of event identity runs into trouble even before one comes to the mind-body identity

theory. We usually think, for example, of the battle between the Monitor and the Merrimack as an event. But it instances many properties: it was the first battle of iron-clad ships and it was a sea-battle off the coast of Virginia, etc. Are we to say that in fact many events occurred in the same place at the same time? This is a hard line to follow, for we will then be forced to say that what we thought was one event was in fact as many events as there are properties true of it and that infinitely many events can occur in the same place at the same time. Brandt and Kim's criterion of event identity is most counterintuitive when we consider acts or behaviors (the sort of events psychological theory will discuss): where we would normally say that a rat's choosing the third object of a set is an event which may itself have many properties, such as being the choice of an object which is white, being the seventeenth such choice done that day, being the rat's mental state at three o'clock, or being what the rat thought would get a reward, Brandt and Kim will presumably have to say that the instancing of each of these properties is a different event. But then this token of the choosing of an object which is the third member of a set is not the same event token as the choosing of the white object, and not is the same event token as the choosing of what the rat thought would get a reward, and so on. Does this mean that the rat has performed as many acts as there are properties true of the choice? We more naturally think of the animal as having done one thing, but that this act has many properties. Event (or act) individuation by property individuation is much more finely grained than our ordinary notion of event or act individuation. Brandt and Kim admit that their individuation of events by properties is "strong" (p. 215), and that

they have used it in formulating the identity theory because any weaker formulations "leaves unresolved most of the puzzles about the relation of mind and body that have agitated philosophers historically (p. 216). This claim has some justification, but it certainly seems wiser to differentiate property and event identity theses. An event identity theory would presumably prevent a dualism of substances, which has traditionally been the major worry. And in any case, we want to determine which forms of physicalism are true, which not, and in what relations they stand. It is self-defeating to formulate the criterion of event identity to fit the needs of mind-body identity theory. At least as a first step, one wants to see whether or not mental events (both types and tokens) can coherently and perhaps truly be said to be identical on some reconstrual of our ordinary notions of event identity. Assuming that Leibniz' Law is true, an event type claim implies the property identity claim, without our having to employ an unlikely theory of event individuation.

Since property individuation is more finely grained than event individuation, we should say that one and the same event will include the instancing of many properties. And it is entirely possible that it will instance more than one nomic property, where a nomic property is one mentioned in the laws of any science. And if one of the nomic properties is not a physical property, (A) could be true and (B') false. Another way of putting the same point is to note that (A) presumably entails that all events are physical events.<sup>3</sup> If, as seems reasonable, we assume the generality of physics-- i. e., that every event has a nomological deductive explanation using only the laws of physics--, then every event must instance

a nomic physical property. But it is entirely consistent with this claim that some events instance nomic non-physical properties as well. Thus the two physicalist theses (A) and (B') differ in a way which pertains not just to ontology but, as we shall see, to the structure of science as well.

### Nomic Properties and Predicates

Since an event may instantiate more than one nomic property, events will cross classify into different nomic classes (a nomic class is all those events possessing one and the same nomic property). For example, one and the same functional state might be instantiated by entities with very different neurophysiologies or electronics. That is, functional state F might be instantiated by entity X when it was in neurological state N and by entity Y when it was in 'electronic' state E; the states F, N, and E are not co-extensive.<sup>4</sup> Similarly, there might be no property nomic in physics and coextensive with a given mental property. In fact, if (A) is true, and if physics is complete in the sense that all event tokens instance a nomic property of physics, and if we have no further constraints on the relation of mental and physical properties, every  $M_1$  event might instance a different nomic physical property. The most we can infer from the completeness of physics and (A) is

$$(2) (x) (M_{1x} \text{ iff. } (P_{1x} \vee P_{2x} \vee P_{3x} \dots))$$

where  $P_1, P_2$ , etc. are nomic physical predicates (each denoting a nomic physical property), where  $x$  ranges over events, and where  $M_1$  is a nomic mental property. But why can we not say that (2) contains in fact a constructed or complex predicate, viz.,

$$(P_1 \vee P_2 \vee P_3 \dots)$$

which we call  $P_1^*$  and which denotes a nomic physical property?<sup>5</sup>

The difficulty with this apparently unproblematic strategy can be seen if we consider these constructed predicates in greater detail. We have, let us assume, a law of psychology which relates two mental properties. The exact nature of this law is unimportant so let us express it schematically as

$$(M_1) \text{ --- } (M_2)^6$$

Let us assume that  $M_1$  is coextensive with a disjunction of physical predicates, as is  $M_2$ . Let us designate the disjunction coextensive with  $M_1$  as  $(P_1 \vee P_2 \vee P_3 \dots)^7$ . Let us designate the disjunction coextensive with  $M_2$  as  $(P_1' \vee P_2' \vee P_3' \dots)$ . Now  $P_1, P_2, P_3, P_1', P_2'$  are all nomic predicates. That is, there is a law-like statement relating  $P_1$  with  $P_1'$  (let us say), and  $P_2$  with  $P_2'$ , etc. (If this were not the case,  $(M_1) \text{ --- } (M_2)$  would express a law which was not expressed in physical predicates.) What we have, then, is the following situation:

$$\begin{aligned} &(M_1) \text{ --- } (M_2) \\ &(M_1) \text{ iff. } (P_1 \vee P_2 \vee P_3 \dots) \\ &(M_2) \text{ iff. } (P_1' \vee P_2' \vee P_3' \dots) \\ &(P_1) \text{ --- } (P_1') \\ &(P_2) \text{ --- } (P_2') \\ &\text{etc.} \end{aligned}$$

We have defined

$$\begin{aligned} (P_1^*) &= (P_1 \vee P_2 \vee P_3 \dots) \\ (P_2^*) &= (P_1' \vee P_2' \vee P_3' \dots) \end{aligned}$$

Now, finally, the question is, can we conclude from this that there is a scientific law

$$(P_1^*) \text{ --- } (P_2^*)?$$

In the first place, it doesn't follow from the fact that

$$(M_1) \text{ --- } (M_2)$$

is a law and the true biconditionals

$$M_1 \text{ iff. } P_1^*$$

and

$$M_2 \text{ iff. } P_2^*$$

that

$(P_1^*) \text{ --- } (P_2^*)$  is a law. For law-likeness is not preserved in what is implied truth functionally by the conjunction of a law and non-law-like but true statements. For example, if it is a law that all rational creatures have large brains, and true that all creatures with large brains have two legs, it does not follow that it is a law that all rational creatures have two legs.

But then can we consider  $P_1^*$  nomic because each of its disjuncts is nomically related to one of the disjuncts of  $P_2^*$ ? The answer is almost surely No. The disjunction of predicates  $(P_1 \vee P_2 \vee P_3 \dots)$  is seriously heterogeneous. It includes predicates not only of neurology but electronics as well. And what we know of these fields suggests that one and the same mental state may be instantiated by neurologically or electronically type distinct events. For example, the state of remembering that '2 + 2 = 4' may well be instantiated by neurologically distinct events each time it occurs; Lashley's principle of equipotentiality implies as much. The electronic variability of computer systems is obvious enough, for one and the same computation state may be instantiated by computers in different physical states.

But the problems with  $(P_1^*)$  have just begun. Since

$$(P_1^*) \text{ --- } (P_2^*)$$

is supposed to be a law, each disjunction must comprise not only

those states which are in fact instantiated when  $M_1$  or  $M_2$  is, but also all those states which are nomologically possible co-instantiators with  $M_1$  or  $M_2$ , respectively. The disjunctions must include, then, not only the states which are instantiated by the creatures and machines we know, but also all the states which could in some nomologically possible universe be the neurology-like states instantiated when  $M_1$  or  $M_2$  are. As Fodor has suggested following Davidson (Fodor, 1974), the most rational assumption is that a disjunction like  $(P_1 \vee P_2 \vee P_3 \dots)$  is susceptible only to brute enumeration; and sufficiently heterogeneous that no amount of enumeration would suffice to convince us that we had indeed enumerated all of  $P_1^*$ . We are convinced that the relevant properties exist because we are convinced of the completeness of physics and the modest demands of event token identity theory. But there would be no general method for determining which  $P_i$  was instantiated during any  $M_1$  event.

If  $P_1^*$  is the sort of predicate just described, it seems immensely unlikely that it could be a nomic predicate. Scientific laws typically are expressed in predicates which pick out what Fodor has called "natural kinds" (Fodor, 1974, p. 7). Disjunctive predicates do not typically pick out such natural kinds. We quite naturally think that 'irradiation of green plants by sunlight causes carbohydrate synthesis' is a law and that 'freezing causes the destruction of protoplasm' is a law, but not that '(irradiation of green plants or freezing) causes (carbohydrate synthesis or the destruction of protoplasm)' is a law.<sup>8</sup> We tend to think that 'all emeralds are green' is law, whereas 'all emeralds are grue' is not.<sup>9</sup> Since  $P_1^*$  contains a probably infinite disjunction of hetero-

geneous terms, its status seems even more dubious than, say, 'irradiation of green plants or freezing'. It seems very unlikely indeed that such a predicate would turn out to be a nomic predicate.

A constructed predicate like  $P_1^*$  would be nomic only if we could find a law

$$(P_1^*) \text{ --- } (P_2^*)$$

Of course it is always possible that a new scientific theory would supply such a law; but it is equally possible that the only nomic class of the  $P_1^*$  events is precisely the  $M_1$  class. Or in other words, there may be psychological laws whose nomic classes are not coextensive with the nomic classes of physical law. Or, more positively, there are laws of psychology which are not expressed in, nor derivable from the laws of any other science. We have special sciences precisely because we want to find the laws that we can, not just those of physics. Constructed predicates would obliterate the differentiation of the sciences which the polynomicity of events permits. And to no end, for all events are already covered by physics.

Some Fundamental Confusions

Token and type physicalism have not, until recently, been distinguished. The cost of this oversight will become clear as we proceed. Another failure of discrimination, one which is more readily apparent, permeates the writings of many modern physicalists (and identity theorists). Even in careful works we find a failure to distinguish between an ontological thesis--type or token physicalism, or some thesis which is vague as to this distinction--and a thesis about the explanation of mental or behavioral events.

For example, we find the following pairs of claims made together without any expression of how distinct the one member is from the other:

- (1) a. "There does seem to be, so far as science is concerned, nothing in the world but increasingly complex arrangements of physical constituents."  
(Smart, 1962, p. 34)
- b. ". . . even the behavior of man himself will one day be explicable in mechanistic terms."  
(Smart, ibid.)
- (2) a. ". . . materialism. . . assumes that the only entities existing in the world are atoms, aggregates of atoms, and the relations between such aggregates."  
(Feyerabend, 1963, p. 83)
- b. Materialism can give us a "correct account" of human behavior. (Feyerabend, ibid.)

(And similarly, Feigl, who never gives a clear statement of his version of the identity theory does manage to claim that "normal inductive extrapolation from the successes of psychophysiology to date make it plausible that an adequate theory of animal and human behavior can be provided on a neurophysiological basis" (Feigl, 1956a, p. 382) in an article which is generally taken as one of the seminal works of the identity theory.) I think it fair, then, to say that at least some of the original identity theorists confused their ontological thesis with some other thesis, one which deals with the explanation of mind and animate behavior. This thesis has no constant form, as should be clear from the quotes above. I wish to abstract (or construct) one thesis which

asserts at least a good deal of what the varying claims express, and is relatively clear, plausible and decidable. The best I can do is thesis

(C) Psychology is reducible to neurology.

How is it that claims apparently so disparate as ((A) or (B)) and (C) are confused or seem to stand in some obvious logical relation, a relation apparently so evident that most authors don't even trouble to state it? A plausible answer, an answer that has some historical justification is that the possible falsity of (A) or (B) was seen as the only likely reason for thinking that (C) might not be true. It is a commonplace today to take physicalism as simply an ontological thesis. But the traditional breadth of discussion hardly seems to be captured in this one claim. And before the rise of modern identity theory in the 1950's physicalists commonly claimed that a thesis such as (C) was in fact the fundamental claim of physicalism. In 1939, Feigl stated that "physicalism in the strict sense" asserts "the potential derivability of all scientific laws from the laws of physics" (Feigl, 1939, p. 382, his emphasis); no mention is made of an ontological thesis. In 1963 Feigl stated that the unity of science thesis espoused by the Vienna circle consisted in (or can now be seen to consist in) two theses: (1) the "unity of the language of science" (usually taken to mean that there is some universal criterion of empirical meaningfulness) (1963, p. 227) and (2) "the facts and laws of the natural and social sciences can be derived--at least in principle--from the theoretical assumption of physics" (1963, p. 228) or (2') "the facts and laws of mental life can be given a 'physical' explanation, . . ." (1963, p. 242). In sum, "physicalism amounts

to a monistic view of scientific explanation, and therefore-- in a sense--also of the universe." (p. 266) (It is to be noted that Carnap, in commenting on this article now admits that the second thesis of physicalism is not "firmly established knowledge but /a/ sweeping extrapolating hypothesis" (Carnap, 1963, p. 883).

Whatever the historical relationship may be, it is more important now to determine the relationships which hold among (A), (B), and (C). What we shall see is that (B)--or a thesis identical to (B) except that  $M_i$  and  $P_i$  are said to be nomically coextensive rather than identical--is a necessary condition for (C), and is, from the standpoint of our present ignorance of both psychology and neurology, the only formally necessary condition for (C) about which there is a clear doubt. (A), it will turn out, cannot carry the weight the (B) or even its weak sister (biconditional) claim can in a proof of (C). Thus it might well turn out that (A) is true, that token ontological physicalism is true, but (B) and (C) are not. To demonstrate these points we must discuss reduction.

### Reduction

Traditional physicalist portraits suggest a hierarchy of the sciences. But there are two conflicting interpretations of this image. One is of a hierarchy where disciplines are differentiated by the different laws and theories (and theoretical entities) they discover. Of course, the special sciences must be closely related if Science is to be that unified explanation of Being which was desired. The unity is provided by suitable relationship

of the laws of the various special sciences. But if 'suitable relationship' and 'unity' are interpreted in plausibly strict ways, the hierarchy collapses, for all the laws are shown to be laws of physics or logically derivable from the laws of physics with the addition of certain definitions or definition-like statements. Then the hierarchy telescopes into physics; or more precisely, the hierarchy is seen to be essentially pragmatic. And this is the second interpretation of the physicalist image of the relationship of the sciences. The interpretations conflict; the image is unstable. To understand traditional claims about the unity of science clearly, we must examine the relationship of the laws and theories of the various sciences--what is commonly called the problem of reduction.

I will use E. Nagel's model of reduction as the standard model of reduction because it is more complete, e. g., in its discussion of the informal conditions a putative reduction must meet in order to be a useful scientific advance and in its discussion of definition relationships; and because other, superficially less stringent models are forced to introduce essentially the same principles Nagel states explicitly, and introduce a bit of confusion as well.

Nagel maintains that

A reduction is effected when the experimental laws (and if it has an adequate theory, its theory as well) /of the secondary science/ are shown to be the logical consequences of the theoretical assumptions (inclusive of the coordinating definitions) of the primary science.

(Nagel, 1961, p. 352)

If the secondary science contains terms which do not occur in the primary science, and if 'A' is such a term, then there must be "suitable relations between what is signified by 'A' and traits represented by theoretical terms already present in the primary science." (Nagel, 1961, p. 353)

But now what precisely is this "suitable relation"? What precisely are the "coordinating definitions"? Nagel suggests three possibilities. One is "logical connection" in which 'A' is shown to be synonymous with or implied by theoretical terms of the primary science. The second is "convention", where by decision we connect 'A' with theoretical terms of the primary science. And third, "factual or material connection" in which

asserting that the occurrence of the state of affairs signified by a certain theoretical expression 'B' in the primary science is a sufficient (or necessary and sufficient) condition for the state of affairs designated by 'A'. It will be evident that in this case independent evidence must in principle be obtainable for the occurrence of each of the two states of affairs, so that the expressions designating the two states must have identifiably different meanings.

(pp. 354-355)

Deciding between the second and third can be difficult. In the case of thermodynamics and the condition of connectability which identifies temperature with mean kinetic energy, we might want to say that this is by convention if the only experimental procedure statistical mechanics had for determining kinetic energy were temperature measurements. It would be the third if we have other connections between the two theories (pp. 356-357).

One point of Nagel's portrait of reduction is vague. He fails

to specify clearly that in his model the "factual or material connection" is a law-like statement. The more common name for these "conditions of connectability" is 'bridge-laws' and that term expresses the matter better. For the difference between the standard model described by Nagel and that suggested by Fodor (Fodor, 1974) can be understood to lie just in this, that the disjunctive, reduction-like statements of Fodor's model are not law-like, whereas those of Nagel's are.

In addition to these formal conditions for reduction, there are also informal conditions, such as whether (i) the reduction makes possible a unification of the apparently unrelated laws of the secondary science, e. g., the way the kinetic theory of gases can relate the second law of thermodynamics and the Boyle-Charles law; (ii) the reducing theory explains the limit conditions or numerical constants of the reduced theory; and (iii) whether the reducing theory allows for less restrictive formulations of the laws of the reduced theory (or even new laws), e. g., as the kinetic theory of gases can formulate gas laws for non-ideal gases (Nagel, 1961, pp. 358-360). Nagel takes these informal conditions quite seriously:

for a reduction to mark a significant intellectual advance, it is not enough that previously established laws of the secondary science be represented within the theory of the primary discipline. The theory must also be fertile in usable suggestions for developing the secondary science, and must yield theorems referring to the latter's subject matter which augment or correct its currently accepted body of laws. (p. 360)

One can certainly appreciate Nagel's informal conditions from a scientific standpoint. A merely formal reduction of

psychology to neurology would be of no use either to psychologists or neurologists. Yet in this particular case, even a formal reduction would be of importance because it would prove (C) "in principle", which is all the philosophical physicalist need worry about. If it could be demonstrated that psychology is not now reducible to neurology and that there are reasons for believing that it will never be reducible, then (C) is false.<sup>10</sup>

Therefore let us return to the formal conditions. Other models of reduction have a principle similar to Nagel's conditions of connectability. This is not immediately apparent. Oppenheim and Putnam define reduction thusly: If  $T_1, T_2, \dots, T_n$  are theories,  $T_2$  is reduced to  $T_1$  iff.

- (1) "The vocabulary of  $T_2$  contains terms not in the vocabulary of  $T_1$ ."
- (2) "Any observational data explainable by  $T_2$  are explainable by  $T_1$ ."
- (3) " $T_1$  is at least as well systematized as  $T_2$ " (i. e. , "the 'ration' . . . of simplicity to explanatory power should be at least as great in the case of the reducing theory as in the case of the reduced theory")

(Oppenheim & Putnam, 1956, p. 5)

A branch of science (i. e. , a special science)  $B_2$  reduces to another branch iff. at time  $t$ , all the theories of  $B_2$  are reduced to theories of  $B_1$ . Now note that this sense of reduction does not require bridge laws connecting the distinctive properties of the laws of the secondary science with nomic properties of the primary science. Two theories could, on Oppenheim and Putnam's account, stand in a relationship of reduction even if no such bridge laws were obtainable: thermodynamics could reduce to statistical mechanics

even if it were the case that temperature and mean kinetic energy were not related by a bridge law. In fact a mentalistic theory of behavior could be said to reduce to a strictly behavioristic theory if they explained the same things and if the behavioristic theory was just a bit more systematized than the mentalistic one. This reduction would hold despite the lack of any specifiable logical relation between the properties, or states, of the two theories. But then this account fails to tell us how we would insure that indeed the two theories explain "the same things." Without bridge laws, the reduced theory might ascribe properties and use predicates not present in the reducing theory; and thus the laws of the secondary science cannot be derived from those of the primary science. If specific laws cannot be derived, it is difficult to see in what sense reduction, rather than replacement by a different theory has been accomplished. The Oppenheim-Putnam model of reduction does not give us a way of distinguishing replacement from reduction. Surely a model of reduction should do this. 'Reduction' by replacement may indeed describe an interesting transformation of theories in the sciences, for example, the replacement of the phlogiston theory of heat with the kinetic theory. In a loose sense, the two theories are dealing with the "same things." But this reduction and this sense of 'same things' is not that of the canonical case, the reduction of thermodynamics to statistical mechanics. It is possible that we will come to regard mental states as we now regard phlogiston. But given that there are mental states and laws thereof, reduction by replacement leaves unanswered the question that now most wants answering, viz., how are psychology and neurology related?

It is the looseness of this reduction by replacement, I think, which leads Oppenheim and Putnam to emphasize another notion of reduction, namely micro-reduction. "B<sub>2</sub> micro-reduces to B<sub>1</sub> iff. B<sub>2</sub> is reduced to B<sub>1</sub>; and the objects in the universe of discourse of B<sub>2</sub> are wholes which possess a decomposition into proper parts all of which belong to the universe of B<sub>1</sub>" (p. 6). This is both stronger and more obscure than Nagel's formulation of inter-theoretic reduction. Nagel's conditions of connectability require only that there be definition relations of some sort between the terms of B<sub>1</sub> and B<sub>2</sub>, not that the things referred to stand in the 'Pt' relation. Oppenheim and Putnam's micro-reduction thesis reads in a kind of atomism into the idea of reduction. Moreover, it is not clear what this strict 'Pt' relation gives to reduction programs: the 'Pt' relation may express something important about the reduction of thermodynamics to statistical mechanics, but why should we limit ourselves to such a narrow relation for the reduction of biology to biochemistry or psychology to neurology? And in any case, the 'pt' relation of entities does not seem to be sufficient to generate the bridge laws which relate nomic properties of the two fields; the 'pt' relation is not sufficient for reduction, if the reduction of thermodynamics is taken as the paradigm case. Yet Putnam and Oppenheim assert that micro-reduction is "the only method of attaining unitary science that appears to be seriously available at present" (p. 8) and that "when we come . . . to branches with different universes--say physics and psychology--it seems clear that the possibility of reduction depends on the existence of a structural connection between the universes via the 'pt' relation" (ibid). (Of course, what may underlie their

insistence on the 'pt' relation is the intuition that there must be appropriate "structural connections" between the property "universes" of the two fields, i. e. , that there be some strong connection of the states or properties mentioned in the laws of the two disciplines. But the clearest form of such a view is just the bridge-law thesis of Nagel's model.)

It is clear, then, that all adequate formulations of reduction must use bridge laws or statements. For without these, the laws of the secondary science cannot be derived as laws (in the case of bridge laws) or cannot be derived at all (in the case of bridge statements). A failure to find bridge laws can be used to generate a clear definition of emergence:

The Boyle-Charles law cannot be deduced from the assumptions of statistical mechanics unless a postulate is added relating the term 'temperature' to the expression 'mean kinetic energy of molecules'. This postulate cannot itself be deduced from statistical mechanics in its classical form; and this fact--that a postulate (or something equivalent to it) must be added to statistical mechanics as an independent assumption if the gas laws are to be deduced--illustrates what is perhaps the central thesis in the doctrine of emergence. . . .

(Nagel, 1961, p. 372)

Emergence in this sense, the non-existence of bridge laws, implies the failure of reduction, though the converse does not hold. To say that a certain property is emergent is to say that a condition of connectability to some physical property does not exist. Physicalism (B') or (B'') obviously maintains that no nomic property (or predicate) of psychology is emergent. Of course, it is consistent with this claim that some psychological

properties are now emergent--or perhaps we should say 'apparently emergent'. No current failure is sufficient to guarantee the falsity of (B') or (B''). Nonetheless, there are principled reasons for believing that at least some psychological properties will (always) be emergent. And in any case, it is rather unfair for physicalism to define itself in such a way that it is not falsifiable. If physicalism is to be an empirical hypothesis about the sciences, one must be willing to accept whatever evidence there is about the emergence of psychological properties: (B') may still be true, but it is at least worthwhile to see what follows, empirically and philosophically, from the fact that it may be false.<sup>11</sup>

#### The Return to (A), (B), and (C)

Now we can play our trumps. Claim (B) is a general schema for bridge laws between psychology and neurology. In fact, (B) is stronger than any bridge law need be, for it asserts an identity of properties, whereas a bridge law is only a law-like biconditional. Thus, ironically, if the identity claim (B) is used as a premise for (C), it goes further than is needed for inter-theoretic reduction; because (again) all one needs for the inter-theoretic reduction claim is the thesis that every mental property is nomically coextensive with a property of physical law. It is now hard to see what motivated the insistence by the early identity theorists on what in effect were property identity claims. The existence of non-physical properties becomes a serious problem for physicalism only when the idea that such properties are not even coextensive with physical properties is seriously entertained--a possibility the early identity theorists rejected on supposedly empirical grounds.<sup>12</sup>

(A), it is important to note, is not a bridge-law. Even though identity is a stronger relation than the nomic biconditional necessary for reduction, (A) fails the reductionist because it does not relate the appropriate, i. e., kind, terms of the two sciences. (A) says only that every instance of a mental state is an instance of some physical state or other, and a bridge law must be stronger than that. Reduction requires law-like statements which coordinate nomic terms of the two sciences. This last point--that the coordinating definitions be law-like--seems fairly obvious in discussions of reduction, but was commonly overlooked in formulations of the identity theory or psychophysical parallelism. It is not sufficient for inter-theoretic reduction that every instance of a mental state is an instance of some brain state or other (i. e., some form of the (A) thesis). What is required is that each mental state type be coordinate with a nomic neurological type term specifiable within an adequate neurological law or law-like generalization. If a term which appears in any laws of the secondary science is not a physical term, nor stands in at least a law-like biconditional relation to any such term, then that term will designate an emergent psychological property. And in this case token physicalism (A) will be true, though (B), (B'), and (C) will not. Differentiating (A) and (B) allows us to determine whether there are emergent properties in psychology without prejudicing us towards dualism.

### Other Confusions

(A) and (B) differ (in word at least) from the formulations of identity theory one frequently encounters in the recent literature. There one finds that an identity claim is usually made about 'raw feels' or sensations (Cf. Place, 1956; Smart, 1962; Feigl, 1956) on the one side and brain processes on the other. Why, we might well ask, are these formulations limited to sensations or occurrent mental states? Why aren't desires, beliefs, thoughts, etc. included? Presumably they too are identical with brain processes, so why are they not included in the formulations of the identity theory?

One reason for the limitation to occurrent mental states seems to be the belief that an identity claim is most difficult to prove in regard to occurrent mental states and that other types of mental state will be covered by the claim, a fortiori. This does have a certain intuitive force, for everyone has probably experienced that sense of a chasm between electromagnetic radiation, neurons, electro-chemical changes, etc., and the experience of red.

But it is important to ask why did it appear to everyone (objectors and defenders alike) that the strongest arguments against identity theory turn on occurrent mental states. Or alternatively, why did it appear to everyone that if identity theory could prove its case with occurrent mental states, other mental states would be covered, too? Were philosophers simply relying on their intuition about the paramount difficulty of occurrent mental states?

The answer to the last question seems to be an unequivocal No.. Most identity theorists have given reasons for believing

that non-occurrent mental states need no consideration at all. The basic reason is the rather startling claim that non-occurrent mental states are not, in a proper scientific ontology, mental states at all. They are, we are told, "dispositions to behave." The commitment by many identity theorists to some form of behaviorism is undeniable.<sup>13</sup> The importance of this point cannot be overstressed. Logical behaviorism, if it were true, would show that certain expressions which we take to be mentalistic have the same meaning as expressions which clearly are not mentalistic. For example, the property designated by ' . . . believes that p' is (so it was said) the same property as that designated by ' . . . has a disposition to behave in way  $B_p$ , or is actually so behaving.' Since it was believed that dispositions and behaviors were safely physical properties, these synonymy relations were thought to demonstrate that mentalistic expression in fact ascribed physical properties. But the behaviorist program runs into serious trouble with mentalistic predicates like ' . . . sees a yellowish after-image', for this does not seem to mean ' . . . is in neurological state  $N_i$ '. The heart of the traditional (type) identity theory is the claim that these two expressions, though not synonymous, ascribe the same property (Brandt and Kim, 1967, pp. 226-228). (Note that a token identity thesis such as (A), since it is a claim about event tokens, need not claim any identity of properties.)

Now we can see why the special attention given to occurrent mental states was well justified within the theoretical context the identity theorists had constructed. Whereas predicates ascribing non-occurrent mental states were said to be synonymous with

some set of predicates ascribing physical states, predicates which ascribe phenomenal or occurrent states were not considered synonymous with neurological predicates. Type identity theory claims that the same property is ascribed even though the mentalistic and physicalistic predicates are not synonymous. And this claim was understood to have a much weaker status than the behavioristic reductions of the intentional predicates.

The identity theory arose, then, as part of a more comprehensive plan, that of showing that all mental state types are physical state types. But the state types are (probably) defined by nomic properties, hence this larger plan is nothing other than (B').

Identity theorists saw occurrent states and the predicates ascribing them as the most difficult part of their general solution to the mind/body problem. But now identity theorists and physicalists generally are faced with the failure of the logical behaviorist program. Sentences ascribing non-occurrent mental states turn out not to be synonymous with sentences ascribing behaviors and dispositions. The upshot for the general program is just this, that it is back at ground zero, where all genuine mental predicates suggest the existence of distinctively mental properties. If anyone takes this general physicalist program seriously, he must now begin again and see if there is any way to account for whatever predicates (and properties) logical behaviorism was thought to render unproblematic.

Now can an account based on a claim of same property ascription by non-synonymous predicates satisfy the physicalist? We saw above (pp. 10 and 27) that the most general way to characterize the physicalist program is to say the physicalism

wants to show that the mentalistic predicates do not imply that there are distinctively mental properties (See (B'') p. 10 . )

A type identity theory will, if true, surely accomplish this, but it is doubtful that an event token identity theory even could. In other words, if an identity claim like (A) is true and (B) false, then the identity theory will not assert the identity of, nor the coextensivity of nomic psychological and neurological properties. And thus, event token identity theory will not demonstrate that psychological properties are physical properties. Now if physicalism amounts only (as is now often said) to an event token identity thesis, then physicalism turns out to be a much weaker thesis than it often presented itself to be, for token event identity, though it is sufficient for denying any entity dualism, permits emergent mental properties in psychology. This conclusion would certainly displease some physicalists, who present physicalism as asserting "a monistic view of explanation" (Feigl, 1963, p. 266), or as asserting that "we can give a complete account of man in purely physio-chemical terms" (Armstrong, 1965, p. 67). The physicalist may wish to withdraw these claims, and at a minimum admit that event token identity is consistent with their falsity. And thus now we can see again the importance of the type/token distinction applied to mental and physical states. As we saw above, as long as it is believed that the ontological claim of physicalism is a claim about the identity of nomic properties, the ontological thesis (viz. (B)) removes what seems to be the only formal barrier to the truth of the reduction thesis (C), for the identity statements are or can be used as bridge laws. Once it is seen that there is a way of putting the anti-dualist ontological claim which does not generate

bridge laws, we have the very interesting possibility that an ontological thesis may be true and the reduction claim false.

### Theoretical Windfall

The relationship between different sciences' explanations of behavior has been a focus of concern for some time. A number of authors argue for vague and rather radical theses about the relationship of psychology and the (presumptively) physical sciences. Norman Malcolm has argued that neurological explanations are "no more basic" than those of psychology (Malcolm, 1968). Charles Taylor has claimed, among other things, "that the shape of developments in the cortex can themselves not be fully explained without reference to the /psychological/ level," (Taylor, 1970), thereby suggesting that neurology must in some way be founded on psychology--a reversal of the physicalist view. It is not my concern here to evaluate such claims fully, but to show that their basic intent can be made more clearly and with more warrant if we attend to the differences between the varieties of physicalism.

For clarity's sake, I shall focus entirely on the work of Charles Taylor. Taylor has attempted to show that at least some psychological laws are teleological in a way no law of inanimate processes is and that experimental behaviorism has been unsuccessful in its attempt to eliminate, reduce, or replace such laws (Taylor, 1964). In "The Explanation of Purposive Behavior" (Taylor, 1970) he considers the claim that there might be non-teleological neurological laws which would explain animate behavior or to which psychological laws would reduce. But it

turns out that some of his claims, which at first blush appear to turn on teleology in fact do not require anything more than the claim that psychological state types are not identical to neurological states.

Taylor begins with what he calls the "thesis of expression"

To say that all thoughts, etc., must have a neural expression is not necessarily to say that a given thought has a characteristic neural expression which always pertains to it whenever it appears in the mind of any human being. The expression thesis need involve no guarantee that the criteria for identity of mental states will parallel the criteria of identity of brain states. What is required for the thesis, however, is that there be no disembodied thought.

(p. 68)

The thesis that there is no disembodied thought is of course weaker than even an event token identity theory; a substance dualist could consistently maintain that no thoughts were in fact disembodied. But the spirit of Taylor's work is certainly anti-dualist, and his earlier article "Mind-Body Identity: A Side Issue?", explicitly propounds an event token identity theory.<sup>14</sup>

Taylor then goes on to assert, and argue in a general sort of way, for what certainly appear to be distinct claims: (a) "that explanation by purpose is the appropriate model for human and animal behavior" (p. 68), (b) "for this range of behavior /higher animate behavior/, the key level of explanation is a psychological one" (p. 72), (c) "that the shape of developments in the cortex can themselves not be fully explained without reference to /the psychological/ level" (p. 72), (d) "for this range of behavior, the most basic level of explanation is psychological" (p. 72) (my emphases). Taylor argues that the thesis of expression does not

entail the thesis which he variously expresses as (a), (b), (c), or (d): "that there is no disembodied mental life tells us nothing per se about the laws governing this life" (p. 73), and thus there is no a priori reason to assume that (a), (b), (c), <sup>or</sup> (d) is false (p. 73). Thus Taylor, in fact, was the first to assert what Fodor was to present more rigorously, viz., that token event identity does not entail reduction, and that there is no compelling argument, a priori or empirical, for the reduction of neurology to psychology. What is lacking, however, in Taylor's work is a clear statement of the relationship of psychology and neurology (or explanations derived therefrom) and what sort of priority psychological explanations have over neurological ones for the explanation of animate behavior. It is by no means obvious exactly what each of (a)-(d) asserts, nor, on plausible interpretations, that they are even coextensively true. To say that psychology is the appropriate level for the explanation of behavior seems to be a merely pragmatic point, one that even a reductivist could agree to. To say that neurological events cannot be fully explained without reference to psychology seems to be a very strong thesis indeed, stronger than the obscure claim that the "key" level of explanation of some behavior is psychological.

With the type-token distinction and Fodor's model of reduction we can make Taylor's basic claim more precise and more viable. We can simply say that there are nomic state types in psychology which are not coextensive with nomic neurological types. Then, instead of (a)-(d), with pragmatic "appropriateness" or an unargued reverse reduction, or a vague "key level" thesis, we have the claim that some psychological event types have, as

types, no nomological deductive explanation on the neurological level. Psychological laws of behavior are basic, in this portrait, in that they have no expressions as laws in neurology, and are not derivable as laws from the laws of neurology. All of Taylor's terms for describing the primacy of psychology--"key", "basic", "appropriate"--suggest that in his view both neurology and psychology explain the same things though psychology is said to be "more basic" or whatever. The type-token distinction allows us to avoid the claim that neurology and psychology explain the same things. This, and the irreducibility of psychology obviate the need to compare psychology and neurology in terms of their basicness, keyness, or whatever; they don't have to be compared at all in these vague respects, for there is a definite sense in which psychology and neurology do not explain the same things. And thus the fundamental point that Taylor wishes to make--that the reduction of psychology to physical science is not a priori guaranteed on the neurological level (p. 67)--is made without obscure claims about their relationship.

### Footnotes

1. In other words, each event token which is an instance of  $M_i$  is also an instance of some physical state; but it cannot be concluded that every event token which is an instance of  $M_i$  is an instance of the same physical state  $P_i$ .
2. I owe this example, and a good deal of this paragraph, to Ned Block.
3. All that is needed is the highly plausible definition  
(P) A physical event is one which instances some physical property or several such properties.
4. The example is Putnam's. See Putnam, 1960.
5. I have eliminated the single quotes commonly placed around predicates mentioned rather than used. The extra punctuation only complicates the already difficult symbolism. The context makes clear in each case whether use or mention is at issue.
6. In this formula and those that follow I have eliminated various terms which would surely be present in standard formulations of any scientific laws, but are irrelevant to my purposes. In particular I have eliminated all quantifiers and variables, since their inclusion would only detract from clarity of presentation. It must be remembered, however, that when I speak of laws I mean something which relates events and not properties. ' $(M_1) \text{ --- } (M_2)$ ' is shorthand for a statement about events, not a logical relation of properties. Properly it would read ' $(x)(M_1)_x \text{ --- } (M_2)_x$ '.  
In addition, I make no specification about the nature of the ' --- ' connective.
7. When I say that  $M_1$  is coextensive with  $(P_1 \vee P_2 \vee P_3 \dots)$  I mean  
$$(x)(M_1)_x \text{ iff. } (P_1)_x \vee (P_2)_x \vee (P_3)_x \dots$$
8. The example is derived from Fodor.
9. See Goodman, 1955.
10. To the extent that identity theorists have confused (A), (B), and (C), they will find the falsity of (B) and (C) either

disastrous--if what they really wanted to prove all along was (C)-- or reason to make a strategic retreat to the less exciting thesis (A).

11. There are several terms which must be kept in balance to generate an interesting problem: if physicalism is defined too rigidly, emergence appears in trivial places, e. g., with 'temperature'; if physicalism is defined too loosely, so that any emergence is consistent with physicalism, physicalism itself has been trivialized, probably watered down to the simple ontological claim (A). An inverse relation holds for definitions of 'emergence'. Some definitions of emergence are unrealistically strict. Feigl, for example, construes emergence as "the logical underivability" of  $p_1$  from  $p_2$  terms" where

'physical<sub>1</sub> terms' = df. 'all (empirical) terms whose specification of meaning essentially involves logical connections with the intersubjective observation language, as well as the terms of this observation language itself'  
(Feigl, 1956a, p. 424)

In other words, physical<sub>1</sub> concepts are part of a language game in which there are intersubjective criteria of truth; any science and parts of ordinary language contain physical<sub>1</sub> concepts (terms).

'physical<sub>2</sub> concepts' = df. concepts 'sufficient for the explanation. . . of the observation statements regarding the inorganic (lifeless) domain of nature'

True emergence is the "logical underivability" of all physical<sub>1</sub> concepts from physical<sub>2</sub> concepts (Feigl, 1956a, p. 424).

Feigl has always admitted the possibility of this genuine emergence, i. e., that certain biological or psychological concepts "could not all be defined explicitly on the basis of the primitives sufficient for the physics and chemistry of inorganic phenomena" (Feigl, 1963, p. 265; cf. Feigl, 1956a, p. 424).

If one takes "logical derivability" to mean synonymy, then any psychological laws which cannot be translated into behaviorese has emergent terms. But by such a strict

criterion of derivability, it might well be the case that thermodynamics is emergent relative to statistical mechanics. After all, it does not seem to be the case that 'temperature' means the same as 'mean kinetic energy of molecules'. If it does not, and if we use Feigl's criterion of emergence, the emergence of psychological properties would not indicate that psychology was somehow importantly different from the sciences of the inorganic generally, for similar emergence would appear within the sciences of the inorganic themselves.

12. Perhaps the motivation was an allegiance to some (B) thesis, (though, again, it is hard to see what motivates this allegiance) or to an event individuation theory like Brandt & Kim's (see above, p. 10).
13. "Modern physicalism. . . unlike the materialism of the seventeenth and eighteenth centuries is behavioristic. Consciousness on this view is either a special type of behavior, 'sampling' or 'running-back-and-forth' behavior as Tolman has it, or a disposition to behave in a certain way, an itch for example being a temporary propensity to scratch. In the case of cognitive concepts like 'knowing,' 'believing,' 'understanding,' 'remembering,' and volitional concepts like 'wanting' and 'intending,' there can be little doubt, I think, that an analysis in terms of dispositions to behave is fundamentally sound. On the other hand, there would seem to be an intractable residue of concepts clustering around the notions of consciousness, experience, sensation, and mental imagery, where some sort of inner process story is unavoidable." (Place, 1956)

"A man is a vast arrangement of physical particles, but there are not, over and above this, sensations or state of consciousness. There are just behavioral facts about this vast mechanism, such as that it expresses a temptation (behavior disposition) to say /e.g./ 'There is a yellowish-red patch on the wall' . . . ." (Smart, 1962, p. 35)

And more recently, J. Shaffer says that he agrees with Ryle that in attributing most mental predicates to a person "we are attributing to him some bit of behavior. . . or a disposition toward behavior or both" (Shaffer, 1965, p. 81), but that "the predicates which refer to feelings and sensations are not analyzable into public, overt pieces of behavior, nor propensities toward such acts" (p. 82).

14. In fact, this article was one of the first to call attention to the importance of a type-token distinction for the identity theory. It is unfortunate that Taylor did not recognize how significant this distinction is for the claims he makes about the relation of psychology to the physical sciences.

### Bibliography

(Armstrong, 1965)

Armstrong, Dan. "The Nature of Mind", inaugural lecture, reprinted in Borst, 1970.

(Borst, 1970)

Borst, C. V. The Mind/Brain Identity Theory, Macmillan, London, 1970.

(Brandt and Kim, 1967)

Brandt, R. and Kim, J. "The Logic of the Identity Theory," The Journal of Philosophy, LXIV (Sept. 1967), pp. 515-37. Reprinted in O'Connor.

(Carnap, 1963)

Carnap, Rudolf. "The Philosopher Replies" in The Philosophy of Rudolf Carnap, ed. by P. A. Schilpp, New York: Tudor, 1963, pp. 227-69.

(Feyerabend, 1963)

Feyerabend, Paul. "Materialism and the Mind-Body Problem," The Review of Metaphysics XVII (1963), pp. 49-66. Reprinted in O'Connor.

(Feigl, 1939)

Feigl, Herbert. "Unity of Science and Unitary Science" (written for the International Congress for the Unity of Science), reprinted in Feigl, H. and Brodbeck, M. (eds.) Readings in the Philosophy of Science, New York: Appleton-Century-Crofts, 1953.

(Feigl, 1956a)

Feigl, Herbert. "The 'Mental' and the 'Physical'," in (Feigl, 1956b), pp. 370-497.

(Feigl, 1956b)

Feigl, Herbert. Minnesota Studies in the Philosophy of Science II, ed. by Herbert Feigl, Grover Maxwell, and Michael Scriven. Minneapolis: University of Minnesota Press, 1956.

(Feigl, 1963)

Feigl, Herbert. "Physicalism, Unity of Science and the Foundations of Psychology," in The Philosophy of Rudolf Carnap, ed. by P. A. Schilpp. New York: Tudor, 1963, pp. 227-69.

(Fodor, 1974)

Fodor, J. A. "Special Sciences (or: The Disunity of Science as a Working Hypothesis)," Synthese 23 (1974).

(Goodman, 1955)

Goodman, Nelson. Fact, Fiction and Forecast, Bobbs-Merrill Co., Inc., Indianapolis, 1955.

(Malcolm, 1968)

Malcolm, Norman. "The Conceivability of Mechanism," The Philosophical Review LXXVII (1968).

(Nagel, 1961)

Nagel, Ernest. The Structure of Science: Problems in the Logic of Scientific Explanation. Harcourt Brace, & World, Inc., New York, 1961.

(O'Connor, 1969)

O'Connor, John. Modern Materialism: Readings on Mind-Body Identity, Harcourt, Brace & World, Inc., New York, 1969.

(Oppenheim & Putnam, 1956)

Oppenheim, Paul, and Putnam, Hilary. "Unity of Science as a Working Hypothesis," in (Feigl, 1956b).

(Place, 1956)

Place, U. T. "Is Consciousness a Brain Process?" The British Journal of Psychology XLVII (Feb. 1956). Reprinted in O'Connor

(Putnam, 1960)

Putnam, Hilary. "Minds and Machines," in Dimensions of Mind, New York University Press, 1960; Collier Books, 1961.

(Shaffer, 1965)

Shaffer, Jerome. "Recent Work on the Mind-Body Problem," American Philosophical Quarterly II (April, 1965), pp. 81-104.

(Smart, 1959)

Smart, J. J. C. "Sensations and Brain Processes," Philosophical Review, LXVIII (1959), pp. 141-156.

(Smart, 1962)

Smart, J. J. C. "Sensations and Brain Processes" in The Philosophy of Mind, ed. V. C. Chappell, Prentice Hall, 1962. Reprinted in O'Connor, 1969. This article is a very slightly revised version of Smart, 1959.

(Taylor, 1964)

Taylor, Charles. The Explanation of Behavior, Routledge & Kegan Paul, London (New York, The Humanities Press), 1964.

(Taylor, 1967)

Taylor, Charles. "Mind-Body Identity, A Side Issue?" Philosophical Review LXXVI, April, 1967.

(Taylor, 1970)

Taylor, Charles. "The Explanation of Purposive Behavior" in Explanation in the Behavioral Sciences, ed. by Robert Borger and Frank Cioff, Cambridge University Press, New York, 1970.

DENNETT ON  
INTENTIONALITY

**PART A:**

**Introduction**

### Stage-setting

In an earlier work ("Physicalism and Reduction") I developed a relatively abstract definition of the physicalist program. In that paper I argued that it is useful and historically plausible to consider the physicalist program in the philosophy of mind as the attempt to show that all nomic properties are physical properties. (Nomic properties are those mentioned in the laws of any science.) The notion of physical property is dangerously obscure, but that problem will not be relevant in this paper, for my concern will be with a property, intentionality, which everyone considers at least *prima facie* non-physical.

Three more familiar doctrines or programs commonly called physicalist can be seen to derive from this broad definition. Inter-theoretic reduction is in part an attempt to show that the properties mentioned in, for example, psychological law are either coextensive with or identical to physical properties. If the former holds, the presumptively non-physical properties are dispensable from a rigorous but complete statement of the laws of science; if the latter, the psychological properties are physical properties. The mind-body identity theory (now often said to be the thesis of physicalism, despite the historical priority of others), as traditionally formulated, is just this property identity claim. The bridge-laws of the reduction of psychology to neurology were, it was said, going to support law-like generalizations of the form 'event  $x$  has property  $P_i$  iff. event  $x$  has property  $M_i$ '; and it was the claim of the identity theory that in fact every mental property was identical to some property of physical theory because, for example, predicate ' $M_i$ ' and predicate ' $P_i$ ' denote or express one and the same property. The new, more precise formulations of

identity theory differentiate between this claim and the claim that all events (dated individual events, that is) are physical, with no claim being made about properties. These weaker event-token theses do not even warrant the elimination of putatively mental properties from science, but do, if true, show that monistic ontology of events is in no way threatened by the putatively mental properties they may instantiate. Logical behaviorism, a third sort of physicalism, can be understood as an attempt to show, without inter-theoretic reduction, that the putatively mental properties of psychology are physical properties because the predicates ascribing the psychological properties are synonymous with predicates ascribing certain physical properties (motions and dispositions). The mind-body identity theory arose historically not from success in reduction but from the failure of behaviorism to physicalize in a linguistic manner phenomenal predicates like 'is in pain', thereby apparently leaving out of account certain decidedly mental properties. The identity theory was seen as a reasonable addition to the basically successful behaviorist analysis of the mental.

Various confusions and errors are to be found in these disparate doctrines and the relationships various philosophers seem to have believed they stood. The point that concerns us here, though, is that the presumption that logical behaviorism gives a correct account of all mental states and properties except the experiential or phenomenal ones is surely false. We are now in the luxurious position of having a rich variety of refutations<sup>1</sup>. If logical behaviorism is false (and let us assume that it is) the physicalist program is back at square one, for now not only are the phenomenal properties problematic, but so also are all the properties that logical behaviorism was thought to

have shown to be physical properties. Psychological states such as thinking, wanting, fearing, believing, et al. are now again problematic for physicalism. These states are the very heart of psychology, and some account of them is necessary if physicalism is to have any substance whatsoever. Identity theory is of little help, for empirical evidence and philosophical argument<sup>2</sup> clearly favor an event-token identity theory over the much stronger property identity theory.

### Intentionality

The psychological states of which behaviorism was thought to give a correct account--thinking, willing, etc.--all display a remarkable property, intentionality. Most concern has focused on this property, rather than on specific accounts of thinking, belief, or whatever. Intentionality was introduced into recent philosophical discussion by Brentano. He claimed that mental states are characterized

by what the scholastics of the middle ages referred to as the intentional (also the mental) inexistence of the objects, and what we, although with not quite unambiguous expression, would call relation to a content, direction upon an object (which is not here to be understood as a reality), or immanent objectivity.

(Psych. from Empir. Standpoint,  
quoted from Chisholm, 1957, p. 168)

In a very rough and intuitive way, I suppose everyone has a sense of what 'the directedness of the mind to an object' means: one can think about unicorns, round squares, possible states of affairs--things that do not exist. And it is easy to believe that one is thinking in each of these cases about some particular thing. One is tempted to say that one is thinking about some peculiar sort of object. For it might well

be argued that we have identity notions for such objects and our acts of thinking about them, just as we do for objects which do exist. For example, Rosenkrantz and Guildenstern are different objects, and thinking about Rosenkrantz and thinking about Guildenstern are two different mental acts; just as Quine and Chomsky are different things, and thinking about Quine a different mental act from thinking about Chomsky. (Needless to say, the existence of adequate identity criteria for propositional objects of thought--be they propositions or something else--is another matter entirely.)

Whether these intuitions can be built into a coherent and plausible theory of mental life is still very much an open question. But even if this vaguely realist talk of the objects of thought is eschewed, intentionality still poses a serious problem for the physicalist program in the philosophy of mind. There is another, far more perspicuous formulation of Brentano's thesis which suggests just how difficult a problem it is. Recent discussion of intentionality is immensely indebted to Chisholm for giving a clear linguistic formulation of intentionality, allowing us to avoid for many purposes the traditional and obscure vocabulary of mental relatedness. Chisholm proposes that a verbal predicate is intentional only if the sentence in which it is used is intentional. A sentence is intentional only if

(i) "it uses a substantival expression--a name or a description--in such a way that neither the sentence nor its contradictory implies that there is or that there isn't anything to which the substantival expression truly applies" (Chisholm, 1957, p. 170)

or (ii) if it contains a propositional clause, "neither the sentence nor its contradictory implies either that the propositional clause is

51

true or that it is false" (p. 171)

or (iii) substitution of identicals does not preserve truth value (p. 171). Let us add one further definition: an entity is in an intentional state  $I_m$  at time  $t_i$  iff. there is an intentional predicate 'm' true of the entity at time  $t_i$ .

Using this notion of intentionality, Chisholm reformulates Brentano's thesis in the following way:

Let us say (1) that we do not need to use intentional sentences when we describe nonpsychological phenomena; we can express all our beliefs about what is merely 'physical' in sentences which are not intentional. But (2) when we wish to describe perceiving, assuming, believing, knowing, wanting, hoping, and other such attitudes, then either (a) we must use sentences which are intentional or (b) we must use terms we do not need to use when we describe nonpsychological phenomena.  
(pp. 172-173)

The immense utility of logical behaviorism should now be fairly clear. If logical behaviorism is true, the second part of Chisholm's version of Brentano's thesis is false. If an intentional predicate e. g. , 'believes that p' is synonymous with 'does behavior  $B_p$  or is disposed to do  $B_p$  ', the intentional predicate is synonymous with a non-intentional one, thereby eliminating the intentional predicate from the list of what were said to be distinctively mental predicates. Since synonymous predicates denote or express the same property, and since the behavioristic properties are (it was assumed) physical properties, intentional predicates, it turns out, denote physical properties.

If logical behaviorism were true, believing (say) that the present king of France is bald need not be defined by the rather mysterious notion of relatedness to a content, nor even by the linguistic property

of intentionality. Believing that the present king of France is bald is nothing more (it was said) than a set of dispositions to say or do certain things and the actual doing of them. Thinking that the present king of France is bald is nothing more than saying, or being disposed to say, that he is, and acting accordingly. If all the apparently intentional states can be handled in this manner, the supposed relatedness to a mental or semantic content is shown to be nothing more than physical dispositions to behave and certain behaviors. Intentionality is then no more mental than any disposition or behavior.

It is now, of course, difficult to see how anyone could have believed any of this; or at least why it was believed with such fervor. With the decline of logical behaviorism and the growing suspicion that identity theory was not an adequate account of all the problems with intentionality for the physicalist approach to the philosophy of mind, intentionality should have received greater attention. But no direct consideration of the problem occurred till Dennett's pioneering Content and Consciousness (1969), perhaps because so many were mesmerized by the identity theory and perhaps also because interest in the traditional scope of physicalism had substantially diminished. To understand Dennett's approach--what he sees the problem to be, and what sorts of answers he thinks appropriate--we must briefly consider another recent contribution to the study of intentionality.

### Quine's Strategy

As I noted above, the return of intentionality as a problematic property for the physicalist program has caused surprisingly little distress. This is particularly odd in view of the fact that Chisholm's work now gives us such a tangible grasp of the problem. There are those who have tried (or can be understood as having tried) to give some kind of physicalist account of intentionality. The three I have

in mind are Quine, Davidson, and Dennett. Dennett is the most direct, being explicitly concerned with intentionality. But properly understood Davidson and Quine appear to have devised a most ingenious approach to intentionality which antedates and to some extent has influenced Dennett's work.

Quine's approach to intentionality is idiosyncratic, occurring in connection with his doctrine of the indeterminacy of translation, and not through a concern with traditional inter-theoretic reduction or identity theory. Thus I think we can begin to understand his strategy by Davidson's use of it (even though Davidson's work is contemporary with Dennett's and appears not to have influenced it to anywhere near the degree Quine's has, if at all), for Davidson is concerned with both of these more familiar ideas. In two important and influential papers "Mental Events" (1970) and "The Material Mind" (1971) Davidson argues in effect that the identity theory has failed to disambiguate an event token identity claim from a state type identity claim. According to the former, all event tokens are physical events. But this is consistent with any event,  $e_i$ , having a property  $P$  in virtue of which it falls under physical law and having a distinct property  $M$  in virtue of which it falls under a psychological law.  $M$  can be identical to  $P$  only if  $(e)(P_e \text{ iff. } M_e)$ , and there is nothing in even token identity theory which requires or demonstrates this claim. In fact, Davidson specifically argues that the instances of a mental state type will comprise a heterogeneous group of neurological or physical states susceptible only of brute enumeration. Even if such an enumeration were successfully completed, we would have no way of knowing we had done so, nor any reason to assume that the resultant statement ' $M_e \text{ iff. } P_e \text{ or } P'_e \text{ or } P''_e \text{ or } P'''_e \dots$ ' was law-like. Without such

a law-like statement, the laws of psychology could not be derived from those of neurology or any other science.

Thus far, Davidson's position is very much like the one Fodor developed somewhat later, and which was discussed in "Physicalism and Reduction" and briefly at the beginning of this paper. Now this position is tolerant of--in fact welcomes--irreducible psychological laws. This is contrary to both the spirit and letter (at least in some forms) of physicalism. But Davidson's intent seems to be something quite different. And Quine's intent certainly is. Davidson seems to argue, using Quinean doctrine, that something is not quite right with the intentional states and laws which occur in psychology. Quine uses the same argument to show that something is definitely wrong with them, but it is simply unclear how far Davidson is willing to go with Quine.

Davidson asserts, quite reasonably, that psychology must include an understanding of language and meaning. But, he says, to understand a single speech-act

we must have a grasp of the speaker's unrealized dispositions to perform other speech-acts. Indeed we may think of having, or knowing, a language as a simple, highly structured, and very complex, disposition of the speaker

(Davidson, 1971, pp. 718-719)

Thus in understanding the language use of persons

We start with wholes [the complex disposition], and infer (or construct) an underlying structure. Meaning is the operative aspect of this structure. Since the structure is inferred . . . we must view meaning itself as a theoretical construction. Like any construct it is arbitrary except for the formal and empirical constraints we impose on it. In the case of meaning the constraints cannot fix the theory of interpretation. The reason, as

Quine has convincingly argued, is that sentences a speaker holds to be true are determined, in ways we can only partly disentangle, by what the speaker means by how words and what he believes about the world. A better way to put this would be to say: belief and meaning cannot be uniquely constructed from speech behavior.

(Davidson, 1971, p. 720)

Davidson does not present his position so forcefully as he should. Radical claims are made with little argument. (Davidson is surely in the minority if he thinks that Quine has "convincingly argued" the view they propound.) The thrust of the position is better put if we say it is only "in the case of meaning [that] the constraints cannot fix the theory of interpretation." The view is that intentional theories--theories which ascribe intentional states to entities--are in a special sort of trouble. All theories are underdetermined by evidence, but intentional theory is said to be indeterminate in some greater or additional sense. It is not entirely clear just what this other sense is, though at a minimum it entails that if all physical theory were determinate, intentional theory still would not be. One consequence of this theory is that there can be no type-type identities from psychology to neurology because the psychological side--the intentional states of meaning--will always possess an indefinite number of interpretations, indistinguishable from the standpoint of neurology and physicalistic behaviorism, but 'intentionally distinct'. For example, one might mean by 'There is a shooting star' that a star has entered the earth's atmosphere, for one believes that some stars are tiny rocks; or one might mean that something or other is shooting through the atmosphere for one believes that it could not be a star which is so behaving for they are hot and very distant (Davidson, 1971, p. 720). This moderate

conclusion is about as far as Davidson goes. Quine, on the other hand, wants to go much further.

Davidson says very little about the status of intentional psychology. Given the distinctive indeterminacy of intentional theory in which Davidson believes, one would expect intentional psychology to have a problematic status. It is defective relative to physical theory; its mentalistic predicates cannot be physicalized either by synonymy relations or inter-theoretic reduction. Yet intentional theory--or at least intentional idiom--is firmly entrenched in ordinary and scientific praxis. Davidson says very little about these matters. But where Davidson is indistinct, Quine speaks with a loud voice. Intentional theory, he says, is not really science at all, but "projection" acceptable only as part of a "second-rate" formalization of our conceptual framework, a formalization which will include the vagaries of ordinary language (Quine, 1960, p. 221). Unlike all previous physicalists, however, Quine does not try to eliminate or physicalize intentionality. Instead he finds a brilliant way to both accept intentionality and (some version of) Brentano's thesis and at the same time reject them from science. He explicitly states that "Brentano's thesis is of a part with the indeterminacy of translation" (Quine, 1960). Quine's argument is roughly this: Genuine science is extensional, for only in extensional language do we have adequate identity criteria. Since there is "no entity without identity," meanings are not scientifically acceptable entities, or not entities at all. Because we lack extensional criteria for the identity of meanings, we lack extensional reductions of intentional idioms. But this is precisely the first part of Brentano's thesis. Quine admits that this is true, apparently giving intentional psychology

a special place. But, Quine continues, precisely because of those features in virtue of which intentional phenomena are specially emergent (their indeterminacy relative to extensional fact), intentional phenomena lack the characteristics necessary for regimentation into the body of science. The very features of language and meaning which give intentionality a special place are the very phenomena which prevent intentional psychology from being a genuine science.

The goal of this strategy, if not its argument, should now be obvious. The irreducibility of intentionality is fully admitted, but the conception of intentionality and science that is used permits Quine to defuse the impact of Brentano's thesis. Recall that that thesis, as formulated by Chisholm, explicitly stated

(2) When we wish to describe perceiving, assuming believing, knowing, wanting, hoping, and other such attitudes, then either (a) we must use sentences which are intentional or (b) we must use terms we do not need to use when we describe nonpsychological phenomena.

(Supra, p. 6)

Quine will admit that this is true, but defuses it by denying its apparent corollaries, that the science of mind is in good part the study of intentional states and that psychological law must be formulated in intentional terms. Intentionality is accepted, but its relevance to scientific psychology is denied--thereby saving a general sort of physicalism without denying or explaining away the obvious presence of intentional phenomena. As Quine puts it:

One may accept the Brentano thesis as showing either the indispensability of intentional idioms and the importance of an autonomous science of intention, or as showing the baselessness of intentional idioms and the emptiness of a science of intention. My attitude, unlike Brentano's, is the second.

(Quine, 1960, p. 221)

But there is a serious and immediate problem with this strategy: Quine has no scientific psychology at the present time. Any intentional psychology (as is much of cognitive and developmental psychology) is rejected as not genuinely scientific; and all the extant attempts at a non-intentional psychology, e. g. , behaviorism, have been dismal failures. Yet Quine continues to insist, and must insist, that some non-intentional psychology is possible, else he will have no scientific psychology at all. This, surely is the source of Quine's faith in behaviorism (and perhaps neurology also).

It is not, fortunately, my task to evaluate Quine's stunning attempt to save physicalism. What is important here is to get a general grasp of Quine's position, for it illuminates Dennett's strategy and some aspects of his otherwise obscure position.

**PART B:**  
**Dennett's Theories**

## Introduction

The first part of this section deals with Dennett's two somewhat distinct models of the relationship of intentional idioms (and science) to extensional idioms (and science), and with objections, conflicts, and problems inherent in the models themselves. The second part deals directly with Dennett's objection to the admission of intentionality to science.

## Dennett's First Model

Dennett is willing to meet the problem of intentionality head-on. He admits

1. Intentionality is an apparently non-physical property
2. Intentionality is now necessary for formulation of the laws of psychology.

But he believes that

3. A purely intentional science "would entail a catastrophic rearrangement of science in general."

(Dennett, 1969, p. 39)

(1) and (2) imply that psychology is not, as currently formulated, a physical science, and that surely is an honest place for a physicalist to begin. (Dennett's reasons for believing (3) are not immediately apparent; in fact, he never argues for the truth of (3), but seems to assume it. We shall get some hints of his concern when we examine his objections to intentionality in the second part, pp. 72ff. )

Since he sees the problem of intentionality as the relatively straightforward (3), Dennett begins with attempts at a straightforward answer. In Content and Consciousness he attempts to find a new way to eliminate intentional idioms from science. He appeals neither to

behaviorism nor (explicitly) to inter-theoretic reduction, but his goal is clear: to frame "within the scientific language the criteria--the necessary and sufficient conditions--for the truth of mental language sentences" and then to explain the apparently mental phenomena in this scientific language (Dennett, 1969, p. 19). Let us call this proposal the Dennett reduction (or D-reduction for short) of intentionality. Dennett does not specify what scientific language consists in, but he is quite clear that it is extensional, for he reads Brentano's thesis as saying that the scientific criteria of mentalistic idioms cannot be found because "no statement or statements about Intentional phenomena can have the same truth conditions as any statement about non-Intentional phenomena" (Dennett, 1969, p. 20). Obviously either D-reduction cannot be accomplished or Brentano's thesis (in any form) is false. Content and Consciousness and the subsequent articles can be read as attempts to refine both D-reduction and Brentano's thesis and to show that some form of D-reduction is possible.

Dennett hopes that psychologists will create "a truly predictive, extensional science of animal and human behavior (specified in pure motion terms and including all human behavior). . . ." (Dennett, 1969, pp. 32-33). Presumably after such a science has been created, philosophers can attempt to determine whether Brentano's thesis is true or not, and more generally to ascertain what relationship does hold between intentional and extensional idioms and between intentional psychology and physical science. Dennett thinks this the best way to demonstrate Quine's claim--which he wholeheartedly accepts--of "the baselessness of intentional idioms and the emptiness of a science of intention" (Quoted by Dennett from Quine, 1960, p. 221).

But Dennett's program is in fact more traditional than Quine's. He does not accept Brentano's thesis as true (as does Quine), nor does he advert to Quine's use of the thesis of the indeterminacy of translation to argue for the eliminability of intentional idioms from genuine science. Instead he sets out to show how Brentano's thesis might turn out to be false! Dennett is fully aware that there is a wealth of evidence for the truth of Brentano's thesis. But, as he points out, that evidence is roughly identical with the evidence for the falsity of logical behaviorism. There is, he thinks, one other way in which reductive definitions of intentional statements might be found. Dennett hopes that there will be appropriate relations between intentional idioms and, instead of "overt behavioral clues, covert, internal events serving as conditions of ascription" (Dennett, 1969, p. 39). That is, he wants to find statements which relate an intentional idiom and some neurological idiom in such a way that we can eliminate the intentional idiom from science.

Dennett is unfortunately most obscure about the nature of D-reduction. His claim seems to be that for every intentional property M (such as the property denoted by ' . . . believes that p'), M is true of entity A iff. extensional property P is true of A. But if the properties are coextensive, one wonders how this might be shown. The claim cannot be that the mental predicate ascribing M and the neurological predicate ascribing P mean the same. The only way I see to establish the coextensivity of the two properties is through finding appropriate relations between intentional states or properties and neurological ones, as the inter-theoretic reduction does by finding bridge-laws. But Dennett says nothing about the traditional intertheoretic reduction, and in fact tries to employ a

much weaker notion. He states:

The controls and activities of computers can certainly be given an extensional description, and if they can also be characterised justifiably in intentional terms, we shall have one case of an intentional-extensional reduction, and hence good reason for expecting a similar reduction in the case of animals and people.

(Dennett, 1969, pp. 44-45)

Now this sense of reduction is so weak that it does not require inter-theoretic reduction of intentional laws. All it says, in effect, is that if an event has an intentional explanation (an explanation via a law which relates intentional states) it also has an extensional explanation --and that that is sufficient for reducing intentionality and intentional laws. But if event token identity theory is true and physics is complete, it is trivially true that an event which has an intentional explanation also has an extensional one. And, for reasons already mentioned, this D-reduction will not eliminate intentional laws. Yet Dennett specifically wants to conclude that this sense of reduction is strong enough to block inference from the irreducibility of intentionality to the thesis of

an unbridgeable gap between the mental and the physical--whether this is construed as a radical dualism of phenomena, or of sciences, or of modes of description and explanation.

(Dennett, 1969, p. 89, my emphasis)

But a "dualism" of the sciences is what we get if there is failure of traditional inter-theoretic reduction for psychology and neurology. And there is considerable reason to believe that such a failure is likely. Dennett's notion of reduction is too weak to prevent what he says it will prevent. The reason, I suspect, that Dennett opts for

such a weak sense of reduction is that, on the one hand, he thinks intentionality must be eliminated, but, on the other hand, does not want to claim outright that Brentano's thesis is false. The reason he does not explicitly attack Brentano's thesis (an attack we might well have expected, given his general program) is the hope that his vague sense of reduction can be made to work, allowing him to give provisional acceptance and independence to intentional psychology while maintaining its ultimate dispensibility. In this he wants to be like Quine, but his program is tied too closely to the older physicalism which wanted to eliminate intentionality directly and cleanly (as in behaviorism). Quine holds no hope for a reduction of intentionality, but Dennett still does, at least in places, and that is what creates so much confusion in his work. ( In his later works, Dennett has tended to downplay the reduction theme, trying in several ways to do what Quine and Davidson do with the argument from the indeterminacy of translation, namely to show that intentional psychology is inherently second-rate and dispensible.)

Dennett backs away from the obscurities of D-reduction almost immediately to make another proposal. D-reduction first looked like an ordinary biconditional reduction; then it seemed to shift to a notion too weak to be of use. Dennett makes a third, and equally puzzling, proposal. He suggests that instead of biconditional relations between intentional psychological idioms and extensional neurological idioms, we will find such relations with intentional neurological idioms! "Could there be a system of internal states or events, the extensional description of which could be upgraded into an Intentional description? . . . Yes." (Dennett, 1969, p. 40) This upgrading requires showing that there can be "conceptually trustworthy formulations

roughly of the form 'physical state S has the significance (or means, or has the content) that p' . . ." (Ibid.) Presumably a D-reduction sentence would look something like this

$$(x)\left((\text{Intentional state } M)_x \text{ iff. } (\text{Physical state } S)_x\right)$$

I being (say) 'believes that p' and where S "means p." Now Dennett's program began with a proposal for the elimination of intentionality through biconditional reduction sentences with intentional idioms on one side, extensional ones on the other. But now he shifts his claim. He maintains that both sides of the statement will be intentional. The brain state is said to be about something. But if this is the case, it is very difficult to see how the intentional idioms have been eliminated. Instead of saying the person believes, desires, etc. we are permitted to say that his brain is in a state of belief, desire, etc. How this eliminates intentional idioms is a mystery to me. To say that extensional descriptions could be "upgraded" into intentional ones is not only vague but utterly inappropriate to the original task. Either a sentence is intentional or it is not. It is not possible to make an extensional. If the physical side of a reduction is not extensional, a proper D-reduction has not been accomplished.

Perhaps Dennett's point is only that there is no reason not to ascribe intentional properties to brain events, and that therefore intentionality raises no problems for the identity theory via a Leibnitz' Law objection; and indeed Dennett spends a good deal of time attempting to show that the ascription of meaning to brain states is possible, spending very little time indeed on the problem of the reduction to extensional idioms. It is certainly refreshing to hear it openly discussed that neurology might now (now) couch its theories

in a physicalist language. But we are still left with the reduction problem, that of describing the relationship between extensional idioms and science and intentional idioms and science. Ascribing propositional content to the nervous system does not improve our chances of finding biconditional relations between intentional and extensional statements. Dennett appears to believe that either we will develop a non-intentional science of mind or that some kind of reduction statements will be discovered. (Perhaps he sees the two suggestions as going together, it is hard to tell.) But he does not offer such reduction statements, nor does he successfully describe the reduction relation he has in mind. This, as I see it, is a lacuna at the very point where we want the most thorough explication.

#### Dennett's Second Model

Dennett's later work avoids the mysterious upgrading of neurology. What he proposes is, instead, a hierarchy of languages or theories of the behavior of organisms (or any entity with sufficiently complex behavior). The prediction and explanation of more complex systems, he argues, is most felicitously accomplished with intentional idioms: the computer perceives that both its rook and its pawn are threatened; it perceives that the pawn is in a most strategic spot; it wants to win; therefore, it will protect the pawn. Dennett calls this sort of use of intentional language, regardless of the nature of the entity and its manner of processing information, the "Intentional Stance" (Dennett, 1971). One uses it, he says, because it is the easiest mode of explanation. No ontological questions (is the computer therefore animate?) nor epistemological questions (Does it have experience?) are to receive their answers

simply from the fact that we adopt the Intentional stance.

Theories which give us characterizations of the rules of computation that go on in a system when it behaves are called "design" theories. Dennett is not too clear about what design stance descriptions and explanations look like. Presumably they are non-intentional, though Dennett specifically states they may be "purpose-relative" (Dennett, 1971, p. 88). Perhaps a design stance explanation would look something like this: 'The computer moved the pawn because it computed the success probabilities of the possible moves according to rule R, and that move was maximal'.

Dennett is curiously ambiguous about the ontological status of design states. He often seems to contrast their genuineness with the "heuristic" character of intentional states; for example, he says what we want to do in psychology is account for an animal's behavior "in terms of its design" (Dennett, 1971, p. 96). Yet this would seem to be inconsistent with the fictionalist view of all non-physical theory Dennett encapsulates in his characterization of the "physical stance". The physical stance is said to tell us about "the actual, physical state of the particular object" via our "knowledge of the laws of nature" (Dennett, 1971, p. 88). All other states must then, be fictional in some strong sense (though Dennett never spells it out); they might be inventions designed to facilitate theory and not truly states of the things to which we loosely ascribe them.<sup>3</sup> I do not find this position very clear, for I am not sure as to what is acceptable to Dennett as a law of nature, nor why only physical states are "actual" (nor just what this claim amounts to). If the "actual, physical state" is the state described by the basic science,

physics, design stance theories--along with all the theories of the special sciences--are fictional. Since it certainly seems that we have laws of nature in other sciences--chemistry, for example--it is reasonable to assume that the states mentioned by such laws are "actual". But then it becomes most unclear why it is that only intentional states are fictional. Some sort of Quinean move would be logical here, but Dennett never makes it. (We shall have more to say about Dennett's fictionalism and his obscure hostility to intentionality in the next section of this paper.)<sup>4</sup>

This model of the sciences puts in even higher relief the difficulties with the original model. Dennett claims that

if one wants to predict and explain the 'actual empirical' behavior of believers, one must . . . cease talking of belief and descend to the design stance or the physical stance for one's account.  
(Dennett, 1971, p. 106)

This claim is certainly contrary to most of the available evidence, the evidence which has led psychology away from behaviorism and which suggests that at least for the present time successful prediction and explanation of animate behavior does require intentional description. Dennett is surely aware of this, and is presumably speaking of what will be the case, given the development of physicalist science. But then the deeper confusions on Dennett's part appear. He seems completely unaware of the recent work by Putnam, Fodor, and Davidson, which suggests that there is every reason to believe that psychology is not reducible to neurology because there will be event types in the laws of psychology which are not nomically coextensive with the event types in the laws of any other science. We must admit, surely, that what is described and explained in the

laws of any science is "actual". And we can do this without compromising a monistic ontology of physical events.<sup>5</sup>

What I think Dennett is in effect trying to do in introducing the hierarchy of stances is to propose a new model of reduction, one which is not so strong as to use bridge-laws, yet strong enough for us to speak of a genuine hierarchy of related sciences rather than the looser relation envisaged, for example, by Fodor. The levels must have some independence, but there must be some kind of reduction of intentionality, or intentional theory. Dennett wants to find a place between traditional inter-theoretic reduction and Fodor's theory of irreducibility. That is a worthy physicalist goal, for inter-theoretic reduction now seems too strong to be true, and Fodor's model decimates the traditional thesis of the unity of the sciences. Indeed, one can see why anyone concerned with the nature of science might be concerned to find some middle ground. But Dennett has not done so. For, as we saw above (p.66), he is unable to articulate a sense of reduction appropriate to this strategy. His sense of reduction is too weak to reduce laws and the apparently "actual" states mentioned in them. It in no way guarantees anything stronger than the claim that every event which has an intentional explanation also has a purely physicalistic one. No interesting inter-theoretic relations are even suggested.

There are yet other problems with Dennett's second model. As we saw, he wants to say that only the physical description tells us what is "actual", and that the intentional description is a "heuristic overlay" (Dennett, 1969, p. 80). But in "Intentional Systems", the design stance itself is characterized in such a way that Dennett endangers his own 'actuality' claim. As he defines it

(pp. 87-88), the design stance includes the program and hence (he says) design stance explanations rely on the "notion of function, which is purpose-relative or teleological" (p. 88). But then there is a very real question as to whether there is a non-intentional functional state explanation of human behavior. If the design stance is rich enough to intentional states, what is to prevent the possibility that the only true design stance descriptions of human beings include intentional states among the functionals? If the program is a description of the internal processing states of an entity and how these are related to one another, there is nothing to prevent the functional states from being intentional. Dennett seems to have assumed that the functional states of humans are non-intentional. We have no a priori reason to think that this is true.

Even if the functional states of humans are non-intentional, the design stance theory of functional states may still not reduce to a physical stance theory in the required way. If D-reductions of intentionality are to be obtained there must be some kind of fairly tight relations between the laws of the various sciences. (The fundamental problem, again, is that Dennett does not tell us what these relations are.) The work of Putnam and Fodor (See footnote 2) demonstrates quite effectively that functional states are in all probability not coextensive with physical states and that functional state theories are not reducible to physical theories according to the traditional model. Assuming humans could be described as organisms with functional states, we have not assumed intentionality, but we have assumed functional states which characterize the behavior of infinitely many different physical devices. What is to prevent a really hard-nosed physicalist from claiming that functional states are merely a heuristic device which presupposes

teleology and thus takes out a promissory note on the notion of purpose, and which, therefore, should be replaced by physical stance description and explanation? If Dennett can accept functional states which introduce purpose and falsify the traditional model of reduction, why is he so hard on Brentano's thesis, which could be interpreted as saying simply that the functional state descriptions of human beings require intentional idioms? Dennett has a strong allegiance to physicalism in some form, yet he is apparently willing to accept functional states which would falsify the traditional model of reduction. His claim that "we want to explain the intelligence of man (or beast) in terms of his design" (Dennett, 1971, p. 96), rather than use intelligence and other implicitly intentional terms is vacuous if the functional states (the program) of human beings is intentional. If the functional states are not intentional states, it is still by no means clear that the functional states are "purely mechanical regularities" given that they are probably "teleological" (Dennett, 1971, p. 88) and that functional states seem to be states whose laws are not obviously reducible to physical stance laws. Dennett's problem is that he wants to eliminate intentional characterizations on the grounds that they are just further interpretation of the extensional descriptions. But there seems to be no motivated reason in his theory for drawing the 'actuality' line against intentionality alone rather than against all non-physical (in the sense of physics) state types. But this brings us to what will be our second topic of consideration, Dennett's hostility to intentionality.

**PART C:**  
**Dennett's Objections**  
**to**  
**Intentional Psychology**

Dennett has several arguments (or suggestions for arguments) to the effect that an intentional psychology is somehow illegitimate. It is my concern in this section to try to determine just what those arguments are, and to determine if they establish what Dennett thinks they do.

### Intentional States Are Not "Empirical"

Dennett's first objection to intentional psychology is that intentional states are not "real" or "empirical" or "actual" features of any entity (Dennett, 1971). He presumably holds some belief like Quine's that intentional descriptions are "projections" (Quine, 1960, p. 220) and in some way not really true of the things to which we ascribe them. Unfortunately Dennett (nor Quine for that matter) does not spell out in any detail what the curious ontological status of intentional states is. As we saw above, Dennett is convinced that only physical states are 'real'. Perhaps what he wishes to claim is not that only physical states--the states ascribed in physics--are really real, but that the states ascribed by all the inorganic sciences, physiology, and neurology are 'real', but that psychological states, or any states with propositional content, are not. This more tolerant conception reflects the traditional physicalist belief that it is intentional states, or psychological states generally, which are problematic. But even on this more liberal reading, it is difficult to see what motivates drawing such a rigid line among the states ascribed by the various sciences.

Dennett's first argument is a non-starter, for it seems to be a presumption and not an argument. He repeatedly asserts that no conclusions about an entity are to be drawn from the fact that we

"adopt" the "Intentional Stance" in describing and explaining the behavior of that organism (Dennett, 1971). But this will not work. We do not just adopt the intentional stance, any more than we "adopt" the belief that there are physical objects; both beliefs are fully entrenched in ordinary and scientific praxis. Pretending that we could just drop the intentional stance does not mean we could. Some psychologists have tried to show that we could explain animate behavior in a non-intentional way, but it is now generally conceded that these attempts were failures. But at least they tried to show that we could dispense with the intentional stance because a viable alternative was (so they said) at hand. Dennett, on the other hand, wants to argue the point with no alternative in sight. And indeed, we shall see shortly that Dennett's arguments are (roughly) a priori.

But even if we accept Dennett's claim about the "adoption" of stances, we could by no means conclude from this alone that intentional accounts are fictional. For if intentional psychological theory is apparently true of certain things, the simplest assumption is that the theory works because it contains laws of states which the entity really has! (In fact, one wonders what better evidence there could be for an entity's having a property than that the property is ascribed by scientific law.) Unless Dennett wants to adopt an instrumentalist view of scientific theory generally, thereby casting doubt on the reality of physical as well as psychological states, it is difficult to see why he maintains that psychological states are not 'real'.

Dennett has another argument for the fictional status of psychological theory. The argument is latent in several articles, but is nowhere explicitly presented because it intertwines with his

claim that intentional psychology is question begging (see below, p. 81 ). The argument is really quite simple: a non-question-begging psychology is a psychology which does not ascribe intentional states; therefore, a psychology which does ascribe intentional states is fictional, for the states it ascribes are not the states ascribed by a non-question-begging psychology (See Dennett, 1974, pp. 17-18). But everything here depends on the question-begging, on how badly what question has been begged. Dennett complains that intentional psychology does not answer the crucial question "What makes for intelligence?" But at least on some interpretations, this failure does not entail anything about the fictional status of intentional psychology. Newton's theory of gravitational attraction does not answer the question "What makes for gravitational attraction?" in a clearly non-question-begging way. It does say that mass makes for gravitational attraction, but it does not say why it is or how it is that mass produces gravitational attraction. Indeed, this was one of the primary objections to its original acceptance, viz., its reliance on 'occult' qualities; perhaps a more modern philosopher would say that the theory begged the question (or some question). But I don't think that Dennett wants to claim that Newton's theory of gravitational attraction is fictional. Even if he does, he is not out of the woods; for if he does claim that acceptable physical theory may be fictional, it will (again) be hard to see what makes the (ex, hypot.) fictional status of some psychological theory problematic and a mark of differentiation from the physical sciences.

Dennett's most complete argument for the fictional status of psychology centers on the obscure claim that intentional explanations

are "normative" and not "empirical".

Deciding on the basis of available empirical evidence that something is a piece of copper or a lichen permits one to make predictions based on the empirical theories dealing with copper and lichens, but deciding on the basis of available evidence that something is (to be treated as) an Intentional system permits predictions having a normative or logical basis rather than an empirical one, . . . .

(Dennett, 1971, p. 97)

Now what does Dennett mean by 'normative'? He states

The assumption that something is an Intentional system is the assumption that it is rational; that is, one gets nowhere with the assumption that entity x has beliefs p, q, r . . . unless one also supposes that x believes what follows from p, q, r, . . . ; otherwise there is no way of ruling out the prediction that x will, in the face of its beliefs p, q, r, . . . do something utterly stupid, and, if we cannot rule out that prediction, we will have acquired no predictive power at all.

(Dennett, 1971, p. 95)

What I think he means, then, is this. If we ascribe beliefs p, q, and r to an intentional system, and s follows from the conjunction of p, q, and r, then we can say the intentional system ought to believe s. And we will have some reason to believe that the system does believe s: our previous ascription of p, q, and r and our assumption that the system is rational generate the prediction that the system believes s. The prediction is based on the logical point that s does follow from the conjunction of p, q, and r; and on the normative point that one who believes p, q, and r ought, other things being equal, to believe s. If it turns out that the system apparently does not believe s (he denies it, and acts consistently with this denial) we have two options: The first is to say that belief in s is part of the evidence for our believing that he believes p, q, and r;

and so perhaps stern denial of belief in *s* is reason to think he does not really believe *p*, *q*, and *r*. On this option, we juggle with the ascribed beliefs to make the system more consistent. The other option is to say that the system is, after all, "human"--it makes mistakes, there are gaps in its knowledge, it gets tired and has trouble following deductions, it never had the opportunity to wonder whether anything followed from the conjunction of *p*, *q* and *r*, it has inconsistent beliefs. There comes a point where both options begin to fail us. We might come across what we thought was an intentional system, but whose behavior was so primitive or disorganized that it no longer made sense to ascribe intentional states to it at all. If an entity never believed any of the consequences of any of its apparent beliefs, we would be hard pressed to say why we thought it had beliefs at all. The attribution of beliefs seems to make sense only if we can attribute some rationality to the 'handling' of those beliefs by the entity that is said to have them.

A good deal of this seems to be true, and obviously true. But I fail to see how it proves the two things Dennett wants it to prove, that "intentional explanation and prediction cannot accommodate either to breakdown or to less than optimal design" (Dennett, 1971, p. 104) and that intentional psychology is fictional. I fail to see how the fact that a modicum of rationality is necessary if a system is to be an intentional system entails that it must have "perfect rationality" (Ibid); nor how the failure of most things to be perfectly rational implies that the intentional states we ascribe to them are not real. Dennett (again) does not spell out his argument. Perhaps his point is just that the existence of difficult or indeterminate cases calls into question the genuineness of intentional states in general.

But certainly the existence of indeterminate cases does not prove that the determinate cases are fictional. That is just the old point about vagueness: some people are not clearly either bald or not-bald; and it does not follow from this that no one is clearly bald.

Dennett repeatedly says that the intentional stance requires that intentional systems have "perfect rationality". But this just seems to be false. When someone fails to be rational (in the sense of not believing a logical consequence of its beliefs), we have a plethora of intentional stance moves to cover the failure and keep the person within the family of intentional systems: 'he couldn't be expected to know', 'he feels very strongly about p', etc., etc.

Finally, there is a sense in which all predictions are normative. If I see something which I take to be, say, a typical comet, and hence possessed of properties P, Q, and R; and if I derive from the laws of astrophysics the law-like generalization that comets with properties P, Q, and R will, when near the sun, manifest property S; I shall say, this comet ought to manifest property S when it gets near the sun. If it fails to manifest S, I might question whether it really had properties P, Q, and R; or I might look for another property T which affects the others; or I might question whether the sun is constant in the relevant respect. Now how is this normativeness and the sort of juggling we do when the particular case fails to meet the norm, different from the normativeness and juggling of the intentional stance?

Dennett seems to suggest that some special problem lies in the fact that many intentional systems speak, and can evaluate their own rationality. Dennett argues that the ascription of, for example, beliefs to an entity is problematic because such ascription

requires (1) that the believer have reasons for the belief, (2) that more often than not beliefs are true, and (3) that a system's avowals are the best source of evidence about its beliefs; and these are not fully consistent.

Conflict arises, however, whenever a person falls short of perfect rationality, and avows beliefs that either are strongly disconfirmed by the available empirical evidence or are self-contradictory or contradict avowals he has made. If we lean on the myth that a man is perfectly rational, we must find his avowals less than authoritative ("You can't mean-- understand--what you're saying!"); if we lean on his "right" as a speaking Intentional system to have his word accepted, we grant him an irrational set of beliefs. Neither position provides a stable resting place; for, as we saw earlier, Intentional explanation and prediction cannot accommodate either to breakdown or to less than optimal design, so there is no coherent Intentional description of such an impasse.

(Dennett, 1971, p. 104)

Dennett's most visible argument is that intentional explanations and predictions are particularly unstable (and in fact true only of some perfectly rational system).

Now Dennett is surely correct that there must be some coherence of perception, belief, and action, else we will cease to attribute intentional states to an entity (see Shoemaker, 1974). Of course, when problems of consistency begin to arise, we can change our ascriptions to gain greater consistency. Part of Quine's and Davidson's objection to intentionality is precisely that too much of this sort of juggling is possible. But that is not Dennett's worry. He maintains that neither denial of avowal nor ascription of irrational belief provides a "stable resting place". He is not at all clear as to

what he means by "stable", but surely his criteria are too high. We can attribute beliefs even though there is (to our minds) little or negative evidence for a certain belief; someone's saying 'I believe because it is absurd' is in no obvious way evidence for his not being an intentional system. Belief is not so strictly tied to rational assessment of evidence as Dennett demands. It is a fact that people do often hold beliefs without evidence, believe demonstrably false things, and even have inconsistent beliefs; and act appropriately with such beliefs. Lacking adequate reason, we can attribute belief on the basis of need and consistency of putative belief and action. Our normal constraints on belief ascription are remarkably loose; an extremely high degree of instability is permissible.

Dennett's claim that an intentional system's avowals are best evidence for its beliefs is perhaps true as it stands. But Dennett takes it to mean that an intentional system's avowals are unimpeachable evidence, or are best evidence in all cases. And that is surely false. People claim regularly to believe in all manner of noble things and act utterly inconsistently with these sincere expressions of belief. We do not thereby question the viability of describing their actions in intentional terms. We can ascribe other beliefs, or ascribe desires or needs, or even inconsistent sets of beliefs in order to make sense of their behavior and speech. We do not assume that people are perfectly rational, nor do we assume that what they say is unimpeachable evidence for their beliefs. The impasse or conflict Dennett wants to find just does not exist.

### Question-Begging

Dennett's more substantive argument that intentional psychology is illegitimate is based on the claim that intentional psychology is question-begging in some important sense. He argues any use of intentional terms in the formulation of psychological theory posits "little men" who understand or interpret or "read" the propositional contents attributed to functional parts of the mind; these functional parts are, then, in some respects, replicas of persons; and thus their presence explains intentional states no more than Descartes' little men explained vision. The use of the intentional idiom is to be made unnecessary by

finding and analysing away these readers or comprehenders; for, failing this, the theory will have among its elements unanalyzed man-analogs endowed with enough intelligence to read the signals, etc., and thus the theory will postpone answering the major question: what makes for intelligence?

Intentional theory is vacuous as psychology because it presupposes and does not explain rationality or intelligence.

Skinner is right in recognizing that Intentionality can be no foundation for psychology, and right also to look for purely mechanistic regularities in the activities of his subjects, but there is little reason to suppose they will lie on the surface in gross behavior. . . . Rather, we will find whatever mechanistic regularities there are in the functioning of internal systems whose design approaches the optimal (relative to some ends). In seeking knowledge of internal design our most promising tactic is to take out intelligence loans, endow peripheral and internal events with content, and then look for mechanisms that will function appropriately with such "messages" so we can pay back the loans.

(Dennett, 1971, p. 99)

Now these are strong and serious claims. I hope it is not too pedantic to begin by asking precisely what question has been begged. The error of begging the question is usually said to be a property that pertains to arguments, not to whole sciences. An argument begs the question when it suffers from what is traditionally called the fallacy of petitio principii, namely including (perhaps tacitly) the conclusion one wanted to prove as a premise in the argument for that conclusion. Naturally enough, there is a broader meaning to 'beg the question', as when someone tries to give us a relatively uninformative answer, a 'dormative virtue' sort of answer. But one doesn't usually hear it said that a special science or the theories which apparently constitute the very heart of that science are question-begging. We have to appreciate how strong a claim Dennett is making in saying that intentional theories "postpone answering the major question" of psychology, or is saying that such theories are "vacuous". For Dennett is not just saying that it would be good or interesting if some theory or special science were able to explain the activity of mind without appealing to intelligence or meaning. He is saying that any psychological theory which does not do this is question-begging in a serious and fundamental way.

Unfortunately Dennett never spells out precisely in what sense intentional psychology is vacuous. He does not tell us in sufficient detail what question has been begged. Apparently he will accept as non-question-begging only those theories which give (vaguely) reductive accounts of apparently intelligent behavior. He refers with approval to the behaviorists' supposed success in showing

that what appears to be intelligent behavior on the part of a pigeon can be shown to consist of nothing but a sequence of conditioned responses. Fully aware that such reductive accounts are notoriously unsuccessful, Dennett argues that there must be internal mechanisms that account for the plasticity and rationality of behavior. Some of Dennett's worry is now clear. If one wants a mechanistic or reductive account of animate behavior, behaviorism is a plausible first step. Its failures motivate the view that an understanding of internal processing is necessary even for an account of overt behavior. Internal processing may at first be characterized in functional or even intentional terms; but ultimately (Dennett claims), these functional parts of the mind (or nervous system) must have reductive analyses into mechanistic systems. Otherwise the question "what makes for intelligence?" has not yet been answered.

The trouble I see in Dennett's portrait of the status of intentionality in psychology is that it makes sense if one wants a reductive account of intentionality; but it does not tell us why we should want that, or accept only such an account as non-question-begging. Consider the question "What makes for gravitational attraction?" It is certainly not obvious that the only non-vacuous answer to this question is an answer which makes no appeal to the notion of gravitational attraction. The answer 'The natural proclivity of things to move downwards' is vacuous, but Newton's theory of gravitational attraction does not seem to be. If it is considered question-begging (because it does not explain what makes for gravitational attraction), that only shows that a question-begging theory can be a good theory. It is not obvious that intentional psychology must be like 'natural proclivity' and not like Newton's theory. After all we want

psychology to provide not just mechanistic explanations of intelligent behavior, but an explication of what intelligence is. In good part this is what psychology has often done. There is a rich history, particularly in Gestalt psychology, of attempts to describe and define experimentally such notions as intelligence in a non-vacuous way (See, for example, Kohler's Psychology of Apes). In fact, it seems that psychology comes up with more interesting and more reliable results the less it fetters itself with a reductionistic methodology. Considerable articulation of mental phenomena is possible--in fact desirable--without appeal to mechanistic accounts. Intentional psychology is not, then, question-begging the way an argument may be; or rather, if it is question-begging, the consequences to be drawn from this fact are not so clear or devastating as in the case of an argument. And similarly, intentional psychology is not obviously vacuous in the way 'natural proclivity' or 'dormative virtue' theories are.

Surely, then, the concern that Dennett must want to express is not that all intentional psychology is seriously vacuous, nor that intentional psychology cannot provide any explication of intelligence. These targets are entirely too broad. I think his concern is rather that even with explications of the notion of intelligence we still explain behavior (or functions of the nervous system) in intentional terms. I think we can see Dennett's point better if we make a short detour. In the philosophy of action, the notion of a basic act is now a commonplace: I warn the burglar by turning on the light; I turn on the light by flipping the switch; I flip the switch by moving my arm and fingers; I move my fingers by \_\_\_\_\_-- and here one does not have anything to say. When one reaches this point where there is apparently nothing to be said as to how

or by doing what X one did Y, one has found a basic act. Dennett's concern is, I think, with what we can call basic mental acts. I recognize a certain painting as a Cezanne by recognizing a certain style. I recognize the style by noticing qualities like the color properties of the palette, the nature of the brush stroke, etc. I recognize the palette by noticing say, certain colors. But how does one recognize a color? There is nothing one does (other than looking, which at this point is not an illuminating answer) to recognize a color. Psycho-physical research might be able to push the questions further: one recognizes a color when one sees a color and takes it to be similar to one represented in the memory. Perhaps someday we will be able to go further. But I think we have gone far enough to see what Dennett's objection is: no matter how far we go, we have the person, or some functional part of the person, noticing, remembering, recognizing, believing--being in some intentional state or other. Dennett's claim is that this continued presence of intentional states makes psychological description and explanation question-begging. When he says, then, that psychology fails to explicate the notion of intelligence, he means not that there is no intentional explication at all of intelligence, but that psychological (and even psycho-physical) explications still attribute intentional states to the functional parts of the cognitive apparatus. We may find natural constraints on what a person or animal can think or perceive, but we still explain much of the behavior in terms of beliefs, desires, and perceptions we feel safe in attributing to the creature. And we will surely find such constraints for functional parts of the nervous system. But still, these animals or functional parts perceive, believe or desire.

And it is precisely this that Dennett finds question-begging. According to Dennett, a non-question-begging psychology would be one in which the basic mental acts were not intentional, but purely mechanistic. (Dennett is never fully explicit about what he means by 'mechanistic'. As an approximation he suggests that a process is mechanistic if it is Turing computable.) Dennett's point, then, is that there must be some basic psychological theory which is not intentional; and that such a basic theory is non-question-begging only if it is non-intentional.

But must there be a non-intentional psychology? The claim that non-intentional basic mental processes exist is entirely hypothetical. It could well turn out to be false. Dennett thinks it must be true, but his argument is not a good one

If we relegate vitalist and interactionist hypotheses to the limbo of last, desperate resorts, and proceed on the assumption that human and animal behavioral control systems are only very complicated denizens of the universe, it follows that events within them, characterized extensionally in the terms of physics or physiology, should be susceptible to explanation and prediction without any recourse to content, meaning, or Intentionality. There should be possible some scientific story about synapses, electrical potentials, and so forth that would explain, describe and predict all that goes on in the nervous system.

(Dennett, 1969, p. 78)

It is certainly true that we presume that physics and physiology will explain all the physical or physiological events in human bodies. But it doesn't follow from this that physics or physiology gives us an explanation of all mental event types. If every event is a physical event, physics will presumably give an explanation of all events in the body and in the brain. But it does not follow from this that the

laws of psychology are derivable from the laws of physics, nor that we could give a physical account of psychological law or the states and processes that constitute mental life. Event token identity and the thesis that all events have an explanation in terms of the laws of physics guarantees that intentional systems made from matter have some kind of reductive analyses. But this kind of reductive analysis, in which no psychological laws are derivable, hardly guarantees that there will be an account of intelligence in physics. Compositional reduction says nothing about the kinds or levels of explanation the special sciences will devise. Dennett seems to have made the traditional mistake of confusing the ontological and epistemological theses of physicalism (see my "Physicalism and Reduction"). The thesis that all events are physical events does not give one a reductive relation of the sciences.

The same points just made about physics and psychology can be reiterated for psychology and any other special science a physicalist might appeal to in a reduction claim--biology, neurology, neurophysiology. If inter-theoretic reduction does not obtain, there is no reason at all to assume that the laws of psychology have a perspicuous representation in any reducing science. Every event token of a psychological state may have an explanation in the reducing science; but that does not show that the laws of mental life have a reductive form in some reducing science. And even less does it show that there are mechanistic forms of these laws in psychology itself. If the nomic states of psychology are not defined in neurology, a move from psychological to neurological law simply leaves out of account those states and laws which define the subject matter of psychology, unless there is inter-theoretic reduction. The virtue of the traditional inter-theoretic reduction

model is precisely that it gives one a way of explaining mental events in non-mental terms without laying oneself open to the charge of changing the subject in changing to a different theory. In the traditional model, every nomic property of psychology is shown to be at least coextensive with a neurological property. But if inter-theoretic reduction does not obtain, the move to the neurological level will lose us the very things we wish to explain. And it might be precisely this that would happen when we move from intentional to non-intentional explanations of mental phenomena. Dennett's compositional analysis guarantees that every mental event has a mechanistic explanation; but it is by no means obvious that this explanation is part of psychology.

Footnotes

1. See among others, Merleau-Ponty, 1942; Chisholm, 1957; Taylor, 1964; Fodor, 1968.
2. See Putnam, 1960; Putnam, 1966; and Block & Fodor, 1972.
3. I am by no means convinced that such a view of the states ascribed in the special sciences is correct. Dennett does not even present a thesis or argument; I am merely trying to fill in some of the gaps in his presentation.)
4. This model is germinally present in Content and Consciousness:

the centralist makes his initial characterization Intentional, describing the events to be related in law-like ways using either ordinary, semi-ordinary, or even entirely artificial Intentional expressions. He then hopes that an adequate physical basis can be found among the internal states and events of the organism so that 'reductions' of Intentional sentences of the theory to extensional sentences of the theory is possible.

(Dennett, 1969, pp. 41-42)

Applying this hierarchy to neurology, we can see now what Dennett was trying to say with his "upgrading". Neurology may very well have to begin by giving intentional characterizations of brain states. Then, the hope is, these can be D-reduced to complex design stance theories; which, in turn, if anyone were interested, could be D-reduced to physical theories.

5. See my earlier paper "Physicalism and Reduction".

### Bibliography

(Block & Fodor, 1972)

Block, N. J. and Fodor, J. "What Psychological States Are Not." Philosophical Review, vol. 81, 1972, pp. 159-181.

(Chisholm, 1957)

Chisholm, Roderick U. Perceiving: A Philosophical Study, Cornell University Press, Ithaca, N. Y., 1957, pp. 168-188.

(Davidson, 1970)

Davidson, D. "Mental Events," in Experience and Theory, ed. by L. Foster and J. W. Swanson, University of Massachusetts Press, Amherst, 1970.

(Davidson, 1971)

Davidson, Donald. "The Material Mind," 4th International Congress For Logic, Methodology, and the Philosophy of Science (Bucharest, 1971), ed. by P. Suppes, et al., North Holland Publishing Co., Amsterdam.

(Dennett, 1969)

Dennett, D. C. Content and Consciousness, International Library of Philosophy and Scientific Method, Humanities Press, New York, 1969.

(Dennett, 1971)

Dennett, D. C. "Intentional Systems," The Journal of Philosophy, Vol. LXVIII, no. 4, Feb. 25, 1971.

(Dennett, 1974)

Dennett, D. C. "Why the Law of Effect Will Not Go Away," unpublished typescript. Presented to the Society for the Philosophy of Psychology, M. I. T., 1974.

(Fodor, 1968)

Fodor, Jerry A. Psychological Explanation: An Introduction to the Philosophy of Psychology, Random House, New York, 1968.

(Merleau-Ponty, 1942, 1963)

Merleau-Ponty, Maurice. The Structure of Behavior.  
Published under the title La Structure du Comportement,  
Press Universitaires de France, Paris. Translation by  
Alden L. Fisher, 1963, Beacon Press, Boston, 1963.

(Putnam, 1960)

Putnam, Hilary. "Minds and Machines" in Dimensions of Mind, ed. by Sidney Hook, New York University Press, 1960.

(Putnam, 1966)

Putnam, Hilary. "The Mental Life of Some Machines" in  
"Intentionality, Minds, and Perception," ed. Hector-Neri  
Castaneda, Wayne State University Press, Detroit, 1966.

(Quine, 1960)

Quine, Willard Van Orman. Word and Object, The M. I. T.  
Press, Cambridge, Mass., 1960.

(Shoemaker, 1974)

Shoemaker, Sidney. "The Problem of Other Minds" in  
Reason and Responsibility, ed. by Feinburg Dickenson  
Publishing Co., Encino, Ca., 1974.

(Taylor, 1964)

Taylor, Charles. The Explanation of Behavior, Routledge  
& Kegan Paul, London, 1964.

OTHER MINDS AND  
THE ARGUMENT FROM ANALOGY

## I

The other minds problem is often expressed by the questions "do we know and how do we know the thoughts and feelings of another?" (Wisdom, 1946). Of course, in a perfectly ordinary sense, we do sometimes know what the thoughts and feelings of others are. And the ordinary ways we come to know are many: by asking; by noticing all manner of behavioral facts--facial expression, gestures, actions, etc.; by determining the circumstances of the other's life, from the most immediate stimuli to general conditions such as health; and most commonly by some combination of these. But Wisdom's questions have a radical sense. For what usually concerns philosophers when they worry about other minds is how we justify, or how we could justify, our various assertions about the minds or mental states of others. Requests for justification of such an assertion, e. g., that Jones is depressed, will normally call forth an answer like 'because he has been a friend for years and I can tell his moods.' The skeptical concern, however, is that all these sorts of ordinary answers presuppose the more general claim that such answers are sufficient justification for the claim that you know that Jones is depressed. What the skeptic wants to know is what justifies our tacit assumption that our ordinary justifications are good justifications, or justifications at all.

The nature of evidence and justification is an issue throughout philosophy. But the other minds skeptic is worried that there is some special problem about our assertions about the minds of others, and about the justification of these assertions. Skeptics typically

express their worries or doubts in arguments designed to undermine our faith in the ordinary assertions and justifications. And these arguments are designed to show that there is some distinctive problem about our putative knowledge of the minds of others. The arguments seem intended (in part) to show that there is more of a problem, or a different sort of problem, with the justification of assertions about the minds or mental states of others than about the physical states of others. This greater problem of intersubjective availability is instantiated in claims like 'only Jones can know what mental state he is in', or in arguments to the effect that our ways of justifying our claims about Jones' mental states are all much weaker or epistemically less secure than Jones' way; whereas we are all in the same epistemic relation to the states of physical objects. We shall examine claims and arguments like this shortly. But it is worth lingering on this point, that skeptics-- and often those who answer them as well--have insisted that there is indeed some special problem about the minds or mental states of others.

The general sort of argument that philosophers have given to show that there is a problem about knowledge of the mental states of others which does not apply equally to knowledge of their bodies runs something like this.

(1) Only person P feels the pain he has.

This, or something like it, is often taken to be a necessary truth (Ayer, 1953; Plantinga, 1967). (1) is generalized to other mental states, e. g. ,

(1') Only person P has the depression P has.

From these it is inferred (perhaps illicitly) that

(2) Only P can observe the mental state he has

or

- (3) Only P has direct, non-inferential knowledge of his mental states

And from (2) or (3) it is inferred that

- (4) If I know about person P's mental states, it is not by observing them, unless I am P

And from this it is inferred that

- (5) Since I don't observe P's mental states (unless I am P), I can justify claims to knowledge about them only by inference from what I do observe

And from this it is concluded

- (6) There is a significant asymmetry between the justification of one's claims about one's own mental states and the justification of one's claims about the mental states of others.

This sort of argument is common to skepticism and to the most common response to it, the argument from analogy. Those who argue for the argument from analogy make claims like (2) through (6), just as skeptics do. They disagree about whether we do have warranted knowledge of the mental states of others, but they seem to agree in thinking that (2)-(6) are true. It is not their disagreement that interests me, but rather their apparent agreement about the skeptical portrait of our epistemic relation to the mental states of others as revealed in statements like (2) through (6). My primary concern in this paper will be to show that the argument from analogy (or at least some widely accepted versions of it) does accept this portrait, and to present reasons for thinking that this portrait is seriously in error. Its mistakes center

upon two complex and problematic notions, that of a mental state, and that of observation, which we shall examine carefully. I shall not try to show that there is no special problem about the epistemic status of the mental states of others. But I do hope to show that the standard arguments used by skeptics and analogists alike are not only unsound, but contribute seriously to the confusions which beset the philosophical understanding of minds and persons.

If I succeed at both my tasks, that of demonstrating the fundamental acceptance of the skeptical portrait by analogists and that of giving good reasons for thinking that portrait in error, I think we will have found the source, or a source, of the vague but strongly felt disaffection with the analogical position. For I will have shown that those who argue from analogy really accept debatable claims--in fact false claims--made by skeptics about other minds. The disaffection lies, then, with the skeptical portrait itself.

## II ANALOGY

I'm not going to discuss the argument from analogy, because there really is no such thing. That is, there is no one argument which is the argument from analogy; rather, there is a family of arguments that have certain features in common, or at least a family resemblance. I will discuss some of these general features. For what concerns me is not proper formulation, but why it is that various authors have thought that there is a need for an argument from analogy and why they think an argument from analogy answers their questions or resolves their doubts. I hope to get at the doubts that generate concern about knowledge of other minds by studying what so many have thought was a good or adequate answer to these doubts.

Let us begin with a simple formulation of the argument from analogy and see what conception of the problem (or a problem) of other minds manifests itself in it. John Stuart Mill argues:

I conclude that other human beings have feelings like me, because, first, they have bodies like me, which I know, in my own case, to be the antecedent condition of feelings; and because, secondly, they exhibit the acts, and other outward signs, which in my own case I know by experience to be caused by feelings. I am conscious in myself of a series of facts connected by an uniform sequence, of which the beginning is modifications of my body, the middle is feelings, the end is outward demeanor. In the case of other human beings I have the evidence of my senses for the first and last links of the series, but not for the intermediate link. I find, however, that the sequence between the first and last is as regular and constant in those other cases as it is in mine.

In my own case I know that the first link produces the last through the intermediate link, and could not produce it without. Experience, therefore, obliges me to conclude that there must be an intermediate link; which must either be the same in others as in myself, or a different one: I must either believe them to be alive, or to be automatons: and by believing them to be alive, that is, by supposing the link to be of the same nature as in the case of which I have experience, and which is in all other respects similar, I bring other human beings, as phenomena, under the same generalizations which I know by experience to be the true theory of my own existence.

(Mill, 1889)

1. One cannot but be struck by the simplicity of Mill's argument. One might well have thought that the problem of other minds and any solution to it would be rather subtle. The simplicity of Mill's argument from analogy arises, in part, from a good deal of abstraction that the reader must contend with unaided. For example, what counts as a relevant type of behavior must be very broad. If it is said that people in pain behave 'the way' I do when I am in pain, this 'way' must obviously cover a multitude of behaviors (and styles of expressing them), for obviously people in the same mental state often behave very differently indeed. Unless, on the other hand, there is some restriction on types of behavior, the argument could be literally analytic because there will always be some similarities between any two behaviors; but if that occurs, the argument will fail simply because I cannot correlate my pain state with yours, rather than with, say, your grief state. And obviously, the constraints on types must be precise enough that the argument is accurate. And it's not obvious that this can be done, or done easily. What distinguishes your pain from your grief

may be evident to me not from your immediate behavior (e. g. , crying) or even your immediate introspectable mental state as you report it to me, but from beliefs about the circumstances or causes of your behavior rather than the behavior itself, or even from the relationship of this grief-behavior to other behaviors and mental states.

2. Perhaps the most surprising (and philosophically disturbing) feature of Mill's argument is that all the different mental states we ascribe to others are compressed or reduced to feelings, or simply left out of account. The having of an intention or desire is surely a mental state, but it is only in an awkward and inaccurate sense that they could be called feelings. Discussions of the problem of other minds often use truncated accounts of mental states. The focus is always upon sensations, feelings, mental imagery, and the like--the sorts of mental states which have introspectable content, or qualia, or some sort.<sup>1</sup> The focus on mental states with introspectable content has several sources. It is sometimes said to be merely "pragmatic" (Aune, 1961), but we shall see that in fact many of the skeptical arguments will work, or even look like they work, only with those mental states which do have introspectable content. We shall see shortly just how problematic this emphasis can become.

Correlative with the selective attention to feelings comes a tacitly regimented notion of "outward demeanor". Mill clearly believes that the epistemic justification of 'I observed that S was in mental state M' is different from the justification of 'I observed that S's body was in physical state P'. The feelings of others are not, according to Mill, something for which "I have the evidence

of my senses." I do, according to Mill, have the evidence of my senses for the "outward demeanor" of others (what we would call behavior) and for the "outward signs" of the feelings I do not sense. But now obviously these "outward signs", these "modifications of the body", must be such that their presence is not sufficient for me to infer analytically that another person feels or thinks, else there will be no problem. Writhing in pain, for example, cannot be one of the observable modifications of the body, because one can infer from 'S is writhing in pain' to 'S is in the mental state of pain'. Writhing itself will presumably be observable, it being an outward sign of pain. But can all mental states be so nicely divided up into mental state proper and outward sign? What about something like being depressed? It is not at all obvious that 'Jones is depressed' is an assertion about Jones' conscious state or feelings. To say that someone is depressed is not to ascribe some particular state of consciousness, for people feel or experience many different things when they are depressed. Naturally if someone is overtly and consciously depressed, we think they must be in some introspectable mental state or other; but this feeling alone is not what defines depression. It seems likely that some of the behavioral features of depression are not "outward signs" of some inner state, but the state itself. The expressions of the face, the listless gestures, the withdrawal from one's common activities and projects do not seem to be mere outward signs of depression in the way the writhing is for pain. Many of our descriptions of persons do not fall easily into either the category of feeling or the category of outward sign of a feeling. Feeling may not be all that there is to mental states; and many of the things we

consider descriptions of behavior may not be mentality neutral. As we shall see, the skeptical-analogical insistence on a separation of inner states of consciousness and outward states of behavior gives us not only an incorrect account of mental states, but of behavior as well.

Another trap lies in the conflating claims about what is inward and what is outward with a claim about the causal status of mental states. If based on a causal claim alone, the argument from analogy is to be an argument to the effect that we know what causes the behavior of others because these causes can justifiably be said to be similar to what causes the similar behavior in our own case. Since statements of causal relations presumably express the relationship of distinct entities or events, a causal account of mental states and behavior is committed to the view that mental events and behavioral events are distinct sorts of things. This causal account of mental states is inconsistent with at least some forms of behaviorism and therefore motivates some worry about the epistemic justification of other minds claims, but it is not by itself sufficient to justify the extreme skepticism some find plausible. The importance of distinguishing the causal claim from the observation claim lies in the fact that if they are compounded, one gets the claim that mental states are the unobserved causes of behavior (unobserved except by the person in the mental state, that is). This permits one to use the dualistic language of inner states and "outward demeanor" without committing oneself to a dualism of substances. But at the same time it burdens the causal claim with a claim about the special un-observability of the mental states of others. The claims are not the same, nor do they, without argument,

stand in any logical relation. In many causal relations we do observe both the cause and the effect, though we do not observe the causal relation itself: the first billiard ball causes the second to move; Jones' argumentativeness spoiled the dinner party; the cold rain forced the bathers to seek shelter. Sometimes the effect is unobservable (in some sense) when the cause is observable: the injection killed the germs. Of course, there is a variety of cases in which the cause is not observed but the effect is, and we reason analogically to the presence of the cause: the presence of another planet caused the perturbations in the orbit of Uranus. Now all of these cases need refinement around the notion of observation. Nonetheless there is no reason to assume a priori that because mental states cause behavior they are unobservable in some important sense. Too many have made this assumption.

## III

## A

Now that we have a rough idea of the skeptical view and the analogical response, it is time to see how the confusions and problems I have hinted at manifest themselves in some typical skeptical and analogical arguments.

Claims about the problematic status of our knowledge of other minds typically begin in one of two ways, either by contrasting the entailment relations of statements about our own minds with those about the minds of others, or by articulating the unobservability of the minds of others. Of course, the two ways are often mixed, but it is easy enough to separate them.

The first method relies on the apparently significant difference between, on the one hand

(7) If I believe I am in mental state M, I am in mental state M  
and

(8) If I am in mental state M, I believe (or know) I am in  
mental state M,

and, on the other hand

(9) If I believe you are in mental state M, you are in mental  
state M

and

(10) If you are in mental state M, I believe you are.

Now (9) and (10) are grotesquely false, whereas (7) and (8) are close to the truth, and this difference is used to show that we have privileged access to our own minds.<sup>2</sup> But while (7) and (8) may

appear to be true at first glance, they are not. Both are surely false if we consider the full range of mental states. I may believe I am happy that Jones got a promotion, but in fact I may be sad, jealous, or resentful. I may sincerely insist that I am not angry about something, but in fact be very angry. And someone else may see very well that I am angry even though I am unaware of it.

It is only with sensations and feelings that (7) and (8) may be true. That is, it is certainly arguable that

(11) If I believe I am in pain, I am in pain

and

(12) If I am in pain, I believe I am in pain

are true. I am not entirely convinced that they are. I am not sure that the notion of repressed pain--as in the psychosomatic theory of Ida Rolf--is incoherent or nonsensical. Or, to take a more familiar case, one can, in athletic competition, suddenly 'feel the pain that was there all along' when the competition ends. And I am not sure how to explain cases where, under stress, one (thinks one) feels pain, but as soon as the general tension subsides, the feeling of pain disappears. These are subtle and important cases, and I do not know how those who argue for this epistemic privacy will explain such obvious cases. But let us not press this point here. For another issue, frequently overlooked, arises. Even if (11) and (12) are true, it does not follow from this that (7) and (8) are (as many have thought, Mill, 1889 ; Ayer, 1953). One could not claim that other minds are problematic on the basis of (11) and (12) alone, all they show is the problematic status of sensations and feelings. But it surely would be curious to hear that there was no problem with all the other sorts of mental states--intentions, thoughts, hopes, desires, wants, beliefs, moods, emotions, et al. Many formulations

of skepticism and the argument from analogy speak explicitly about a general problem with the full range of mental states (Ayer, 1953; Wisdom, 1946; Planginga, 1967). Yet the crux of most arguments is around states with introspectable content, such as sensations. As we have just seen, the entailment relations of (7), (8), (11), and (12) do not support the claim that all mental states are problematically private. How could such a glaring mistake be so repeatedly made? The answer lies, I think, in a confusion about what the terms 'mental state' or 'mental event' are to refer to. We use them to refer both to mental states proper, and to the feelings one might typically have while in a certain mental state. (I shall use 'mental state' the first way.) The role of experiential content in mental states is a complex one. There are many mental states one can have with no introspectable content at all-- intentions for example. Hope is an interesting case because one can hope with no feelings whatsoever (as in 'I hope they are on time for dinner') or with very strong feelings indeed ('I hope the plane doesn't crash'). These sorts of considerations can easily be extended to the other sorts of mental states. It becomes rather obvious that whatever the role of introspectable content may be, mental states will not generally be defined by their introspectable content. Hope, for example, would probably be defined in terms of a positive attitude towards a propositional object (what one hopes for). How a 'positive attitude' is to be understood is a mystery at this point, but it will surely not be understood as some introspectable feeling. The particular feelings that commonly go with hope in our culture might well be considered irrelevant to the definition of hope or the identification of particular hopes, just as muscle twitches in the throat, or one's awareness of these twitches, is not considered

part of the definition of thought or thinking, as several late 19th-early 20th century psychologists believed.

And now the question returns, is the other minds problem a problem about mental states, or is it a problem about the introspectable contents of mental states? The plausible historical hypothesis is that it began as the former and has drifted towards the latter. The modern skeptic or analogist can easily claim that he is concerned with the problem of introspectable content, and not with the old problem of mental states. This move is certainly a workable one, but only at a considerable cost. For the scope of the problem becomes immensely restricted. The problem of other minds, it turns out, does not concern the privacy of mental states, but the apparent privacy of the introspectable contents of some mental states. (The temptation is to try to slide from the restricted problem to the general one. Ayer takes the full trip: He insists there "must remain a sense in which it is necessarily true that experiences are private and necessarily true also that one cannot know the thoughts and feelings of another. . ." (1953, p. 352). Building a problem of other minds from privacy considerations is, then, a difficult task at best. The standard notion of privacy will tolerate only sensations and feelings as plausible candidates for private mental states, and even with these the case is problematic.

## B

Consideration of the distinction between mental state and content of introspection is required with the second battery of skeptical arguments intended to cast doubt on the security of our beliefs about the mental states of others. These arguments deal with the problematic notion of observation. Plantinga claims

- (P1) We cannot observe the thoughts and feelings of another; so we cannot determine by observation that another is in pain
- (P2) While a person can observe another's behavior and circumstances, he cannot perceive another's mental states.
- (P3) 'S determines by observation that S\* is in pain only if S is the same person as S\*' expresses a necessary truth.

(Plantinga, 1967, pp. 188-189)

We must begin with a point of clarification. Plantinga discusses, on the one hand 'determine by observation that', 'observe that', and 'see that'; and, on the other hand, 'observe \_\_\_\_\_' where the fillers are non-sentential. It is one thing to claim that we cannot observe the thoughts and feelings of another person and something else to claim that we cannot observe that another person is in some mental state. As it stands (P1) is a glaring non-sequitur, for we surely can observe that someone is in pain.

It is safe to assume, then, that what Plantinga wants to do is to differentiate 'observe \_\_\_\_\_' from 'observe that' in its ordinary uses, and both of these from 'observe that' (and its synonym 'determine by observation') in the technical sense he wishes to create. This technical sense will be such that statements

like 'I observed that he was in pain' will necessarily be false because I cannot observe his pain. It is easy enough to see why any philosopher might want to regiment the uses of 'observe that', for it is notoriously broad in ordinary language, being used to introduce both the results of simple observation, as in 'I observed that he had red hair', and the results of inductive reasoning as in 'I observed that he was a Frenchman from his distinctive accent'. Plantinga's technical sense for 'observe that' is intended to restrict its usage to those cases such that an 'observe that' statement is true only if the corresponding 'observe \_\_\_\_\_' statement is true. For example, 'I observed that he had red hair' will be permissible because 'I observed his red hair' is true. 'I observed that he was a Frenchman from his distinctive accent' will not be permissible because 'I observed his French-ness' cannot be true. And similarly, 'I observed that he was in pain' will not be true because 'I observed his pain' is necessarily false (or meaningless). The obvious intent of this legislated sense of 'observe that' is to enforce a putative distinction between observation and inference which 'observe that' tends to obscure. Plantinga wants to argue then, that 'I observed that he was in pain' is justified by inference, and is not a genuine observation sentence because 'I observed his pain' is either necessarily false or meaningless.<sup>3</sup>

It would seem to be part of Plantinga's position that a rigid distinction of observation and inference can be made, and that it can be placed in such a way that one's knowledge of one's own mental states is observational (or at least can be observational), while knowledge of the mental states of others is inferential. Claims like these pervade work on other minds. But we know

that the first claim is dubious indeed: it attempts to make a distinction that cannot (so far as we know) be made. Observation and inferential knowledge do not seem to be mutually exclusive categories. It seems fairly clear that our knowledge of things we do observe is in part inferential. I can certainly be said to see the drinking glass in front of me. But to see that something is a glass is to see that it would probably break under certain typical conditions. Such knowledge of the properties of glasses is surely inferential. A good deal of our knowledge of the things we commonly see is likewise inferential. If one wants some rigid dichotomy of observation and inference one must, apparently, claim that what we commonly think we observe--such as glasses--we do not, in the strict sense, observe, for the very reason that our knowledge of such things is inferential. If true observation comes unpolluted by inferential knowledge, we must observe very little indeed. It seems unlikely--given the history of failures--that this search for what is purely observable will produce usable results. And it seems obvious that this genuine observation will be so rudimentary a perceptual act that the observation/inference dichotomy will not place the physical states of others among the things we observe and the mental states of others among the things whose existence and nature we infer but do not observe. For knowledge of physical objects and their physical properties is, as we just saw, inferential.

Presumably then Plantinga wants to claim, not that observation and inference are mutually exclusive categories, but only that our knowledge of the mental states of others is inferential, whereas knowledge of our own is not (or, not always). This

seems to be the import of several of Plantinga's examples. Consider one case that Plantinga refers to, where a physicist says 'I saw that mu-mesons had entered the cloud chamber' and 'I saw the mu-mesons in the cloud chamber'. Both of these would seem to be acceptable usage. Nonetheless, it is evident that knowing that a mu-meson has entered the cloud chamber is justified in part by certain inferences. It would be preposterous to say that the physicist knows about the existence of mu-mesons from casual observation. At some point in time mesons were hypothetical entities whose existence was postulated in the context of a certain theory; and still a considerable part of what it is to be a mu-meson is revealed only by theory. Observing a cloud chamber will not allow us to see mesons unless we know what to look for (or, if we see them, we do not know we see them). Once the theory is known, however, we may justify the claim that a mu-meson is present by appeal to the simple claim that 'I saw it' or the like. But even when we observe a particular thing, be it a meson or anything else, we justify our claim that it is the thing we take it to be by appealing to previous experience, theory, and inference. Our seeing is informed or knowledgeable and in so far as it is, we can be asked to justify the knowledge claim implicit in an act of seeing. That it was a mu-meson that I saw begs for justification that will surely rely on inference, despite the fact that I truly see the meson.

Plantinga's point, then, is that just as our knowledge of mesons is mediated by a theory of physics, so our knowledge of the mental states of others is mediated by a theory which justifies our knowledge claims about the mental states of others

inferentially (whether or not we observe these mental states).

But, he seems to argue, we can know about our own mental states without inference. We need inference for knowledge of the mental states of others, but not in our own.

Now is it true that our knowledge of our own mental states is non-inferential? Obviously I may infer from my own behavior that I am, say, jealous or depressed. What Plantinga is asserting though, is that in addition to inferential knowledge of our own mental states, we have non-inferential knowledge as well. But what is this non-inferential knowledge? It is true that one does not need to infer that one is in pain to be in pain, but it does not follow from this that one knows that it is pain, or knows what pain is, without inference. The case is similar to the case of seeing what is in fact a mu-meson without knowing that it is a mu-meson. Infants must, for example, learn that certain states of discomfort are called 'pain' and others 'need', 'desire', 'fear', etc.; and presumably they do this by making certain inferences (among other things). I may feel intensely angry; but it seems likely that I would justify the assertion that it is anger I feel with inferences about my previous mental states and behavior and from what others have said and done when I have been angry before. I may indeed genuinely observe the presence of my own mental states, and be subjectively most certain of their presence, but that does not show that my knowledge of them is non-inferential.

Perhaps Plantinga would say that we have misunderstood what he means by non-inferential knowledge. Perhaps he would say that all he means by non-inferential knowledge is that privacy claims like

(11) If I believe I am in pain, I am in pain  
and

(12) If I am in pain, I believe I am in pain  
are true, whereas

(9\*) If I believe you are in pain, you are in pain  
and

(10\*) If you are in pain, I believe you are in pain  
are false. But we have been here before. These are part of the  
typical skeptical-analogical claims about privacy. We were told,  
however, that there were other sorts of claims and arguments--  
built upon the notion of observation--which supported skeptical  
claims about the mental states of others. These arguments  
cannot collapse into reiteration of the entailment relations which  
generate privacy worries, else the whole supposed problem of  
observation is lost.

At this point the skeptic or analogist may well want to with-  
draw the claim that a distinction between observation and inference  
bifurcates one's own mental states from those of others. Appeal  
might be made instead to ordinary speech and experience to show  
that these reveal at least a rough and ready distinction of observed  
and not-observed, and thereby reveal something problematic  
about the epistemic status of our beliefs about the mental states of  
others. This is perhaps what Plantinga intends when he asserts  
(P3), which entails that we do not observe the mental states of  
others, but do observe our own. Plantinga thinks this latter claim  
obviously true, but I do not. I find it difficult to determine in just  
what sense the mental states of others are supposed to be unobser-  
vable. It is certainly true that I can observe that someone is in a

certain mental state. And it is not at all obvious that

(13) I observed his mental state

is necessarily false, whereas

(14) I observed my own mental state

may be true. Now (13) is certainly false if it implies that when you feel pain, I feel pain, too. But it doesn't. (13) contains the problematic notion of observation, but does not say anything about telepathy or concomitant experience. There certainly are mental states for which (13) is true: In an ordinary sense of 'observe', it does seem correct to say 'I observed the cat's intentions', 'I observed the children's delight when they saw Santa Claus', and the like. In an ordinary sense of 'observe', we do observe intentions, moods, feelings, all sorts of mental states. When we observe and speak in this way, we do not mean that we are reporting the presence of some independently existing entity, say depression, that has settled on your mind. Mental states are states; they are not the sorts of things that exist independently of persons: they cannot exist (and ipso facto cannot be observed) without there being something one takes to be the animate creature having them. We do not see the depression separate from the person and the complex of behavior in which it is embodied. But mental states are not alone in this; the same is true of physical dimensions, smiles, laps, and other features of persons. But it certainly doesn't follow from this dependency of mental states that we cannot be said to observe the mental states of others, for it is certainly true that we can see the other's smile. So it certainly shouldn't follow from this existence dependency that, say, being depressed or being in pain are private mental states which

only our unfortunate friend can observe. Yet this is exactly what Plantinga seems to imply when he compares one's observing one's own pain and observing someone else's red hair (p. 188). He seems to argue 'I observe that S is in pain' can be known to be true by observation only if 'I observe S's pain' can be known to be true by observation; and that will be the case only if we can observe the pain without observing S (as we can observe S's hair on the floor of the barber shop).

The skeptical trick is to get us to worry about 'observe S's pain'. We are supposed to think that this is some metaphysical oddity, barely conceivable. But the same supposed problem can be generated with 'observe S's posture' or 'observe S's smile' or 'observe S's rotundity'. The relevant notion is not, however, 'observe S's pain', but 'observe S's being in pain'. Our concern is the observability of states of persons, not the qualia which may be present to the person in that state. The skeptic gets us to think about these introspectable qualia as though only the apprehension of qualia could count for 'observing S's being in pain'. The reason the skeptic so easily confuses us (and himself) is that we are seduced into interpreting 'observe S's pain' on the model of 'feel S's pain', which we know only works when we are S. Once one has accepted the notion that 'observing S's pain' is to be analyzed as 'feeling S's pain', the skeptical case is well on the road to success.

Now we can see from a new angle how dangerous it is to accept the idea that mental states are defined by their introspectable contents. For once that is accepted, it is easy to conclude that only apprehending that content as the other does counts as observing the other's being in the mental state which

contains that qualia. At this point, it becomes plausible to imagine mental states as mysterious sorts of things, curiously like ordinary things in that they are observable, but very unlike normal things in that the mental states of S are observable only by S. Now if the mental states of others are thought of as objects only he can observe, then the mental states of others surely would present awesome epistemological problems. But we don't have to imagine they are like this, for they aren't. This intuition--that the mental states of others are private objects, objects only he can observe--seems to lie at the heart of the skeptical-analogical portrait of mental states.<sup>4</sup>

## IV

My attack on the skeptical-analogical portrait of our epistemic relation to the mental states of others has focused first on its implicit conception of what mental states are, and then on its problematic claims about observation. It is easy enough to see how the two strands of my critique can be woven together.

Skeptics and analogists typically discuss feelings and sensations and not the full range of mental states. If confronted with the problem of the lack of (or irrelevance of) introspective content or the problem of the apparent observability-in-others of the other types of mental states, the standard move is to try to separate the feeling of or the feeling in the mental state, and run the arguments on this feeling, rather than on the mental state itself. That is, the skeptic can argue that we say 'I observed his feeling depressed' and not 'I observed his depressed feeling' and use the difference to argue that we do not really observe the depression, that we only observe the "outward signs" of the depressed feeling, which, because it is the other's experience, we cannot experience ourselves. Feelings have the same virtue sensations have, namely that there is often a relatively definite and relatively uniform introspective content. And where there is introspective content, it is easy to say that this content is the mental state and to infer from this that anything other than experiencing this content will only an "outward sign" of the having of the content. No one today would make the error of thinking that all mental states could be defined as introspectable items. (We should remember that

the other minds problem seems to have arisen at the very time when psychology was thought to be the science of experiential states and when introspection was considered the fundamental method, if not the only method, of psychology.) But it is also an error to assume that because I cannot observe your introspectable items (when such there are), I cannot be said to observe the mental state, the experience of which for you is in part that introspectable item. I can't observe your pain if that means observing the qualia you sense; but the relevant notion is that of observing someone's being in pain, not observing the pain. The assumption that it is underlies the common slide from the point that I can't feel your pain to the claim that I can't observe your being in pain. An appeal to a (tacit) introspective definition of mental states seems to be the source of Plantinga's claim that

(P2) While a person can observe another's behavior and circumstances, he cannot perceive another's mental states.

(p. 188)

But Plantinga plays with a stacked deck. For him, 'observe a mental state' means 'to come into contact with an introspectable item'. Hence 'observe the mental state of another' means 'come into contact with the introspectable item of another' which (barring magical telepathy) is impossible. If knowledge of other minds requires observing the introspectable qualia of the other's consciousness, the problem of other minds does look insoluble. But there is every reason to think that observational knowledge of the mental states of others does not require feeling the introspectable contents of the mental states of others.

Now of course the introspectionist is going to ask, 'In what does the observing the mental state of another consist, if not in the having of (duplicates of) their feelings?' The answer is just that in observing the person--the behavior, the face, the flow of expressions and actions--we observe the mental states. As long as we are beguiled by introspective definitions we keep looking for a special process or special mode of access to the minds of others. But if we don't demand introspective definitions, we don't need any. We don't need a special process to observe the smile, the muscular tone, the rhythm of action, and in observing these, we observe the mental state itself. The animate body is an appearance of the mental state. The feeling (if such there be) is how it usually appears to the person who has the mental state, and the behavior is how it appears to others. In both of these the mental state itself is revealed, just as in observing two very different sides of an object, two persons see the object itself.

This view is what Strawson seems to be asserting when he writes:

. . . we speak of behaving in a depressed way (of depressed behavior) and also of feeling depressed (of a feeling of depression). One is inclined to argue that feelings can be felt, but not observed, and behavior can be observed, but not felt, and that therefore there must be room here to drive in a logical wedge. But the concept of depression spans the place where one wants to drive it. We might say, in order for there to be such a concept as that of X's depression, the depression which X has, the concept must cover both what is felt, but not observed, by X and what may be observed, but not felt, by others than X (for all values of X). But it is perhaps better to say: X's depression is something, one and the same thing, which is felt but not observed by X and observed but not felt by others than X.

(Strawson, 1959, p. 105)

I take it that Strawson is suggesting just what I suggested, namely that a mental state token is intersubjectively available in that it, the mental state token itself, is felt by the person who has it and observed by those who do not feel it. A mental state token is, then, an intersubjectively available entity. The nature of its intersubjective availability is different from the intersubjective availability of, say, tables and chairs, because of the frequent assymetry of feeling and observing from first to third person; but that in itself does not seem sufficient to prove that it is not intersubjectively available. This issue would seem to be the central issue of the other minds problem. For if 'X's depression' refers to an entity which is available (observable or feelable) only to X, then some form of the problem of other minds could always be generated. But if 'X's depression' refers to an entity which is intersubjectively available, then the other minds problem, as originally formulated, cannot get a foothold.

It is worth noting why it is that Strawson (and I) are driven to the point of saying odd-sounding things like "X's depression is something, one and the same thing, which is felt but not observed by X and observed but not felt by others than X", or "X's depression' refers to an entity which is intersubjectively available. . . ."

I have just argued that mental states should not be thought of as independently existing things to which persons stand in some special relationship of ownership. When then appeal here to locutions which apparently reify mental states? The answer is that skepticism has played such havoc with the natural locutions like 'I observed that he was depressed'. Strawson would, I think, like to say simply that it is X's being depressed or the fact that X is depressed which is, in a non-technical sense, observable. But

skeptics typically complain that these locutions, and the experiences they express, obscure the fact that mental states are unobservable in an epistemologically significant sense. The skeptic may allow that X's being depressed is, in the ordinary sense, observable; but he will assert also that the depression itself is hidden from view. Therefore we are driven to assert that mental states are themselves not unobservable in the way skeptics claim they are. The point of our reifying usage is not to make it out that mental states are independently existing things like tables and chairs, but only that they are genuinely observable features of persons. Reifying locutions are always odd with features or properties, be those properties mental or not (e. g., 'I observed his rotundity'); but it is necessary to use them to make it clear that mental states are observable features of sentient creatures. The skeptic attacks our more neutral locutions, such as 'I observed that he was depressed'. The skeptic has cast what is an illegitimate doubt on our normal 'observe-that' expressions, and thus we use somewhat reifying expressions to say that even allowing doubt about 'observe-that' expressions, the case for the observability of mental states can still be made.

The Strawsonian view can easily be misunderstood. Many will interpret the claim that one can observe the mental states of others in their behavior as an obscure way of putting some logical behaviorist claim to the effect that sentences ascribing mental states are synonymous with or entailed by sentences ascribing behavioral or dispositional states. And Strawson gives credence to such interpretations by saying dangerously vague things like

What I have said is that one ascribes P-predicates to others on the strength of observation of their behavior; and that the behavior-criteria one goes on are not just signs of the presence of what is meant by the P-predicate, but are criteria of a logically adequate kind for the ascription of the P-predicate.

(Strawson, 1959, pp. 102-103)

One could interpret this as making a behaviorist claim. And we know that logical behaviorism is false. So then those sympathetic to skepticism will say that Strawson's view about the observability of the mental states of others depends on a demonstrably false claim. So we are forced, says the skeptic, to choose between logical behaviorism and the skeptical view that we cannot observe the mental states of others; since logical behaviorism is false, the choice is obvious.

I don't believe that we have to make this bitter choice, nor that the behavioristic reading of Strawson's theory is correct. What we need, and what I think the skeptic is right to demand, is some idea of what a non-behavioristic interpretation of the theory looks like. That will be the task of the next section, where I hope to show that our ordinary experience and description of behavior is considerably more mentalistic than behaviorists constraints can permit; and once it is understood that relatively mentalistic features of behavior are observable, the notion of observing the mental states of others becomes considerably less obscure. I hope it is clear now, though, why this is worth pursuing: It is generally assumed that logical behaviorist accounts of mental states are false. It is also generally accepted that in seeing an aspect of a physical object, one sees the object itself; and this is held to be true even though object statements are not synonymous

with sense-datum statements. But it is also commonly accepted (by philosophers) that in observing another's behavior, one does not observe the mental state of another. A good many philosophers have assumed that this asymmetry is guaranteed. But given the failure of the various arguments used in effect to demonstrate this very point, it would seem to be a wiser course to try to articulate (if possible) the way in which the mental states of others are observable. Theories of perception are no longer thought to be caught in the dilemma of sense-datum theory or thing-in-itself theory. There is no reason to assume that theories about the mental states of others are caught in the analogous dilemma of behaviorism or skepticism.

## V

Current use of the term 'behavior' is very broad indeed. Facial expressions, the use of language, manifestations of neurological disorders, biologically innate reflexes, intentional actions, et al., are all said to be 'behavior'. Understanding this imprecise and sweeping usage may be more of a sociological task than a philosophical one. But there is, I think, a philosophical confusion lurking within this strange fascination with 'behavior'. When philosophers speak of behavior in the context of the problem of other minds, they seem to mean 'behavior-and-thus-not-mental-state'. Skeptics often say, in effect, that all behavior is observable, whereas what is truly a mental state is not. I have objected to this sort of view on several grounds. Now I wish to focus on the (often tacit) claim that all the behavioral states of persons can be considered physical states (as opposed to mental states) and as such observable. The skeptical-analogical portrait of mental states seems to say that the problem of intersubjective observation begins abruptly with the specially mental states, and that all other features of persons can be considered observable because they are physical features. That is, I have previously objected to the skeptical analogical view because it falsified the nature of mental states. Now I wish to argue that the very same separation of features into mental and physical obscures or falsifies the nature of all those features of persons which are lumped together under the title of 'behavior'.

To begin somewhere close to the beginning again, let us remember that the other minds skeptic asserts that there is an epistemic problem about the mental states of others which does

not apply equally to their physical states. Skeptics now make it fairly clear that by 'mental state' they have in mind something quite special. But they are not at all careful when they consider the notion of physical state. The first problem is that it is simply not at all clear what counts as a physical state of a person. Certain features certainly are, such as physical dimensions or the color of the skin. As a general rule, one might say that those properties which are truly predictable of thoroughly inanimate objects as well as persons are clearly physical properties. Skeptics imply that all those features they consider genuinely behavioral or simply behavioral are safely physical. (Certain states which are undeniably behavioral are also undeniably mentalistic. For example, 'S is writhing in pain' describes S's behavior; but the statement surely entails that S feels pain. The skeptical view seems to be that such a state is not genuinely or not simply behavioral, because it entails the presence of a mental state. The skeptic will surely argue that this state is, like the predicate which ascribes it, a compound one, composed of behavioral and mental parts. 'S is writhing in pain' is true if and only if 'S is writhing' and 'S is in pain' are true. The skeptic will assert that only '. . . is writing' is a genuine or simple behavioral predicate. This position is eminently plausible in the case described, and in similar etiological ones like the case of 'jump for joy'. We shall see, however, that the skeptic tries, in effect, to extend this analysis to states and predicates which are not obviously compound, like 'smiling' or 'pointing'. It is with these that serious problems arise for the skeptical view.)

The skeptic says (roughly) that there is no relevant epistemic problem with the ordinary physical features, no problem with the

genuinely behavioral features, but that there is a very definite problem with the mental features or states. It is, however, not at all obvious that the behavioral states are as transparent as skeptics present them; that, for example, it is only with mental states that problems of observation begin. In the previous section I tried to argue that it does make sense to say that we can observe the mental states of others. The skeptic questions this. But now I want to pose the obverse question for the skeptic. If we accept his arguments about observation, can it really be said that we observe the behavior of others?

We must be careful here to speak of properties or states or predicates, and not events. A person's behaving in a certain way can be truly described in a variety of ways, and there appear to be important differences among some of these descriptions. It is certainly a very plausible thesis that one and the same event may be described as 'eating' or as 'flexing muscles A, B, and C'; or that one behavioral episode may be described as 'writhing in pain' or simply as 'writhing'. Every behavioral episode has, presumably, a physiological description. But this is hardly sufficient to show that behavior is physical in the strong sense implied in the skeptical-analogical view. For what is obviously meant there is that the relevant predicates (or properties) of behavior are physical predicates (or properties). And this is a difficult claim. I shall attempt to show that it is false.

Certain states of persons are clearly physical states, e. g., having a broken leg. Other states are clearly mental, e. g., believing that p. But between these two extremes lie the vast number of states that characterize what people (or their bodies)

do. It is not at all obvious at all that these states are physical ones. For example, when we say that someone is waiting, are we ascribing a physical state or a mental state (or neither)? One is tempted to say that waiting entails expecting, and that expecting is a mental state. Yet we normally think we can observe that someone or something is waiting. Should we conclude that we can observe that something is expecting that p? The skeptic will certainly say No, and try to drive a wedge between the mentalistic and physical parts of these states. It could easily be argued that we observe the cat's sitting in front of the mouse hole, but not the intentions, desires, or beliefs that make such an act a waiting for something. But the problem with this is that what remains observable in behavior is considerably less than we commonly think and considerably less than what skeptics suggest when they claim that behavior is observable, but mental states not.

Some behavioral predicates (and, one assumes, the states they describe) are mentalistic: 'reach for', 'look for', 'point to', 'wait for', etc. are all intentional, just as 'desire that p' or 'hope that p' are. With intentional mental states, like desire, the skeptical case is at its weakest, for introspection is irrelevant and privacy claims are false for such states. But now intentional behavioral states present the obverse problem for the skeptic: if they are considered physical, one wonders how they could be intentional; if they are mental, one wonders how they could be observable. To take the first course seems foolish, for one would have to drive the mental/physical wedge between, say 'expect' and 'wait'; and would still have to explain how it was that intentionality, which many have taken to be the mark of the mental, spans the gap.

The second option is more in the spirit of the skeptical position, but it too, is not easy. For, say, when you point to something, it certainly seems that we see or observe your pointing. But the skeptic should find this way of speaking false or deceptive, else he will have to admit that we can observe someone's being in an intentional state; and if he admits that, it is hard to see why we couldn't be said to see someone's being in a clearly mental intentional state, such as hope. It would seem that the skeptic will have to say that we do not really observe someone's pointing, but that we observe the muscular movement which has been done with a certain intention. We observe the muscular movement, but not the intention with which it is done.

The same considerations will apparently have to apply with non-intentional predicates and states as well. Things like smiles and frowns cannot really be observable either; for when someone frowns (in a genuine way, in the appropriate situation), we take it that they are displeased or distressed.<sup>5</sup> We certainly seem to see the person's mood in their frown; but the skeptic will have none of this. He might argue that just as someone from a foreign culture will surely observe something, yet not see the mood in the frown, so in epistemological reconstruction we must distinguish the frowning expression from the mood which causes it. He will want to say that the frown per se, the muscular motion is all we truly observe. Meaning, mood, or motive must be leached from what we can be said to observe. The skeptic should say something like "You think you see the animation of the body. But it could be a manniken or a robot. So all you are really entitled to say is that you observe 'frown-like expressions' and 'pointing-like gestures' and the like. The animation you think you see in the gestures,

expressions and actions is there only by dint of what your own mind contributes to what it truly does observe."

Thus what the skeptic really wants to say is not that we observe behavior (and not mental state) but rather that we observe behavior if we take a restricted set of descriptions of behavior as our observation predicates. It is not just mental states which are unobservable, but also the animation of behavior itself. The skeptical-analogical portrait of the mental is not just a theory about mental states, but also a theory about behavior as well. Not only does it give us an odd view of mental states, it gives us a correlatively odd view of behavior as well. Our experience of animation is denied, and only a limited set of behavioral predicates is permitted. Apparently innocuous statements like 'His face was tense' must be denied observational status. Thus when skeptic or analogist says we observe "outward demeanor" or behavior, he must mean it in a severely restricted sense. When the skeptic says 'behavior' he does not mean behavior as we ordinarily experience and describe it, but behavior in a much more narrow sense indeed, one in which only physical, physiological or anatomical descriptions count as observational.

But now a rather interesting connection can be brought to light. If the skeptic or analogist will accept only such weak predicates of behavior as observation predicates, then this constraint on the technical sense of 'observe' is just another way of putting the traditional behaviorist constraint on the genuine, unempathic description of behavior. The skeptical observation predicates correspond to the 'genuinely scientific' predicates of psychological theory. Skinner states:

The independent variables must . . . be described in physical terms. An effort is often made to avoid the labor of analysing a physical situation by guessing what it 'means' to an organism or by distinguishing between the physical world and a psychological world of 'experience'. The events affecting an organism must be capable of description in the language of physical science.

(Skinner, 1953, p. 36)

Hull suggested

An ideally adequate theory even of so-called purposive behavior ought, therefore, to begin with colorless movement and mere receptor impulses as such, and from these build up step by step both adaptive and mal-adaptive behavior.

(Hull, 1943, p. 25)

These goals require that we eschew our ordinary predicates of behavior in favor of predicates, which, like the ones needed by Mill and Plantinga, are mentality neutral in a strict sense. Under the condition of using a stern behavioristic language, in which the only permissible observation predicates are Hull's terms of "colorless movement", one would not be able to tell by observation that another creature was in pain, nor that it was waiting, nor even that it was smiling.

Skepticism, the analogical position, and behaviorism all suppose or argue that the mentalistic descriptions of animate behavior are not among the genuine or objective descriptions of behavior. What is truly observable is bodily movement or "outward demeanor" in a narrow sense. Now I am not claiming that all skeptics and analogists are behaviorists, nor am I trying simply to discredit their arguments by linking them to behaviorist ones. The arguments familiar in other minds discussion are generally

different from those employed by behaviorists.<sup>6</sup> The moral I hope to draw is a larger one, that behaviorists and skeptics are barking up the wrong tree, albeit from different sides.<sup>7</sup>

Skeptical and behaviorist arguments revolve around, and attempt to refine, the vaguely formed distinction between mental and physical phenomena. Our ordinary experience and descriptions of behavior pose a problem for both programs; for animate behavior resists easy classification in either category. The behaviorists attempted to deflate animate behavior, to show that its differences from inanimate motion are theoretically insignificant, through the rigidly physicalistic analysis of mentalistic predicates and stimulus-response accounts of behavior itself. The skeptic or analogist has to be equally hard with animate behavior. Arguments about the epistemologically rigorous use of 'observe' and 'observe that' are directed to this same end, to showing that what we are entitled to claim about behavior and about the animation of behavior is much less than our ordinary experience of other persons leads us to believe. Behaviorism and skepticism may disagree about the status to be accorded to conscious experience, but they agree in claiming that what we can truly be said to know about behavior is what can be said in physicalistic terms. Both behaviorism and skepticism err, I suggest, in so thinking. Their common error is their belief that such descriptions are forthcoming and hence that animate behavior will cease to be resistant to the mental/physical distinction.

The skeptic attempts to force his claim that we cannot observe the mental states of others by insisting that we only see the behavior of others. He tries to use this distinction, for example, to force the Strawsonian view (p. 118, above) to choose between logical behaviorism and skepticism about the mental

states of others. But the skeptic has trapped himself instead. He has forced himself to adopt behavioristic constraints on the permissible descriptions of behavior. If you take 'behavior' in this sense, of course you don't observe any mentality at all. But if we don't take 'behavior' in this sense, we can say that we certainly do observe mentality, or animation, in behavior. There is nothing mysterious in the idea that we truly observe animate behavior. The skeptic wanted us to say we only observe the behavior, not the animation. But he must say we do not even observe the behavior in the sense in which we ordinarily do. The behavior/mental state choice can be forced upon us only by forcing on us first a strange and dubious doctrine about behavior itself, and about the observation of behavior of others. In trying to show we cannot observe the mental states of other persons the skeptic seems to require that we cannot observe that the behavior of others is animate. There is so little reason for accepting this last claim, I think we can put a dilemma to the skeptic: if we don't observe the mental states of others, we don't observe their behavior (in the normal sense of the term) either.

### Footnotes

1. It is sometimes said that such mental states are more properly called 'mental events', 'states' being reserved for things like intentions which one may have while asleep, or when not thinking about them. I wish, however, to use the term 'mental state' in a broader way, such that it is a cover term for all the events, states or processes we call 'mental'.
2. Plantinga uses (7), whereas Don Locke (1968) uses (8) and explicitly denies (7). Ayer (1953) seems to believe both (7) and (8).
3. That this is Plantinga's position is not immediately apparent from his text. He does not so much argue for the acceptance of this technical sense of 'observe that' as present cases for our intuitive consideration:
 

. . . in a perfectly ordinary sense of 'see that' one may properly claim to see that someone else is in pain. In the same or a similar use, one can see that a child has measles, that a pipe will give a sweet smoke, and that electrons of a certain sort are sporting in the cloud chamber. . . . In a similar use of the term, the theist, impressed by the harmony and beauty of the universe or the profundity of the scripture may justifiably (even if mistakenly) claim to see that God exists.

(Plantinga, p. 189)
4. Ryle suggested this long ago. Cf. (Ryle, 1949), pp. 201-220.
5. The fact that we are sometimes wrong shows very little indeed. We are often wrong about physical objects and their properties. And it is surely an error to infer from this that we don't ever see them.
6. There is occasionally overlap. Plantinga, for example, may have fallen into the trap of thinking that the ascription of mentalistic predicates compromises the claim that all events are physical events.

7. As is well known, behaviorism first arose at a time when various forms of introspectionism were considered valid methods of psychology, and sometimes considered the only valid methods in psychology. The behaviorists wanted, or can be interpreted as having wanted an intersubjective method for obtaining objective and intersubjectively valid knowledge about the mind. Such objectivity is impossible if mental states are knowable only by the person who has them. Or rather: if there is no intersubjective way of studying mental states, there will be as many psychologies as there are persons--or more precisely, we will have no way of determining if the results of my introspection match with yours so that mental state types can be known to apply to different persons. In response to this problem the behaviorists insisted that the only way to achieve objectivity in psychology was to study behavior in a rigidly physicalistic way. (This interpretation of behaviorism as a worry over methodological solipsism, and its relevance to the other minds problem is not new. Kohler explicitly states it in his Gestalt Psychology, Chapter I.) If behaviorism was, in part, an attempt to find an intersubjective basis for the study of mind, the historical irony is that the insistence on physical or physiological descriptions of behavior reinstated the introspectionist idea that a mental state was an introspectable item of consciousness, radically different from the behavioral manifestations.

### Bibliography

(Aune, 1961)

Aune, Bruce. "The Problem of Other Minds," Philosophical Review, LXX (1961), pp. 320-339.

(Ayer, 1953)

Ayer, A. J. "One's Knowledge of Other Minds," Theoria XIX (1953). Reprinted in Essays in Philosophical Psychology, ed. by Donald F. Gustafson, Anchor Books, Doubleday & Co., New York, 1964.

(Skinner, 1953)

Skinner, B. F. Science and Human Behavior, New York, MacMillan, 1953.

(Locke, 1968)

Locke, Don. Myself and Others, Clarendon Press, Oxford, 1968.

(Malcolm, 1958)

Malcolm, Norman. "Knowledge of Other Minds," The Journal of Philosophy LV (1958), pp. 969-978.

(Mill, 1889)

Mill, John Stuart. An Examination of Sir William Hamilton's Philosophy, 6th ed. (London, 1889), pp. 243-244. Quoted by Malcolm in 1958.

(Plantinga, 1967)

Plantinga, Alvin. God and Other Minds: A Study of the Rational Justification of Belief in God, Cornell University Press, 1967, Ithaca, New York.

(Ryle, 1949)

Ryle, Gilbert. The Concept of Mind, Hutchinson and Co., Ltd., London, 1949. Published in the United States by Barnes and Noble, Inc., New York, 1949.

(Skinner, 1953)

Skinner, B. F. Science and Human Behavior, New York, MacMillan, 1953.

(Strawson, 1959)

Strawson, P. F. Individuals: An Essay in Descriptive Metaphysics, Methuen and Co., Ltd. (1959), Anchor Books, Doubleday and Co., Inc., Garden City, New York (1963).

(Wisdom, 1946)

Wisdom, John. "Other Minds," Aristotelian Society Supplementary Volume, XX (1946), pp. 122-147.

## BIOGRAPHICAL SKETCH

Stephen Prior was born on January 21, 1946, in New York City. He attended Columbia College in the City of New York from 1964 till 1968, majoring in philosophy. He received his B.A. in 1968 and was initiated in ØBK. From 1968 to 1971 he studied philosophy in the graduate program at the Massachusetts Institute of Technology, receiving an NDEA fellowship. His article, "The Importance of Metaethics," is to be published in The Journal of Value Inquiry this year. In 1972 and 1973 he travelled in the Americas and Europe. He married Patricia Ann White in May of 1976.