This is to certify that the

dissertation entitled

Folk Psychology and Eliminative Materialism

presented by

Stanley Charles Mortel

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Philosophy

Richard J. Hall
Major professor

Date August 12, 1988

FOLK PSYCHOLOGY AND ELIMINATIVE MATERIALISM

By

Stanley Charles Mortel

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Philosophy

1988

ABSTRACT

FOLK PSYCHOLOGY AND ELIMINATIVE MATERIALISM

By

Stanley Charles Mortel

A brief examination of some of the traditional dualistic views on the relationship of the mind and body reveals significant problems, especially regarding causal interaction. Our common sense, conceptual framework concerning the explanation and prediction of human behavior is construed as a theory, termed folk psychology. This theory is evaluated relative to a proposed competing theory which comes from the neurosciences, according to a variety of well accepted criteria, and found to be deficient. The relationship of folk psychology to dualism is also discussed.

Reductive materialism is shown to consist of two distinct components, intertheoretic reduction from folk psychology to a physical theory and an ontological view called the identity theory. Its relationship to folk psychology is explicated by the drawing of a distinction between weak and strong ontic commitments. Weak ontic commitments involve the types-of-things countenanced by some theory. Strong ontic commitments involve such ultimate types as mental or physical. Both psychoneural and psychofunctional reductivism are examined. Their similarities and differences are brought out by contrasting the type/type and the token/token versions of the identity theory.

Eliminative materialism is presented as an alternative to reductivism. This view is examined in relationship to the network theory of meaning, to the view that perception is theory-laden, and to the identity theory. The major problem facing the eliminative materialist is to account for the qualitative character of perceptual experience. It is maintained that both a dualistic folk psychology and eliminative materialism are best construed as paradigms. These paradigms are summarized and used to resolve some of the contemporary disputes between materialists and proponents of folk psychology.

It is concluded that there is very little substantive support for the strong ontic commitments of either dualism or materialism, regardless of the strengths or weaknesses of neuroscientific or folk psychological theories of human behavior. Thus, the proposed alternative to the theory of folk psychology is better termed `eliminative neuroscience'.

Dedicated
to my wife Susan
who made this work possible
and to my mother
who aroused my curiosity

ACKNOWLEDGMENTS

I am most deeply indebted to Professor Richard J. Hall for his careful and repeated reading of this dissertation, for the hours spent discussing philosophical issues and for all he has taught me over the years. His comments and guidance have been valuabe to me beyond measure. I would further like to thank Professors Theodore Johnson, Joseph Hanna, Winston Wilkinson and James Zacks for their time and effort spent reviewing this work.

A further debt of gratitude is due Susan Mortel for her willingness to discuss philosophy over dinner and for asking the obvious and difficult questions so often obscurred by the clouds of dust raised by philosophers. My thanks also go out to Michael Goodman for several years of camaraderie and long informative talks.

TABLE OF CONTENTS

LIST OF FIGURES

Chapter 1

Dualism

Introduction

Consider a mature speaker of the English language who knows the meaning of such words as `mind', `body', `mental event', `brain event', etc. Without getting into the technicalities of reference, such a person can ask several philosophically interesting questions. Does the term `mind' refer to anything (real)? Does `physical body' refer to anything (real)? If they both refer to something real which does indeed exist, then do they refer to one and the same thing or to two different things? If they refer to two different things then what is the nature of this difference? Are they different in kind? That is, do they differ in their basic natures, as do footballs and even numbers?[1] Finally, if `mind' and `body' do refer to two different (kinds of) things, then are the two related, and if so in what ways? In particular, is there a causal relationship which exists between them such that a physical event, e.g. dropping a bowling ball on one's foot, can cause a mental event, e.g. excruciating pain?

Now, someone familiar with the argumentative maneuverings of philosophers may well be cautious in responding to such inquiries. Especially when the questions are formulated using words such as `refer' and `different in kind'. Still, there seem to be two quite different common responses by the clear headed non-philosopher to the above queries. They go something like this. 1.) The mind just is the

brain, or an active, normal brain anyway. Sure, physical events cause mental events because mental events just are brain events. Hence, if one's foot is injured by a bowling ball then certain neural events occur which are the painful sensations. (This is the view which philosophers call the identity theory. It will be the topic of a later chapter.) 2.) Yes, there are minds and bodies. All (or at least most) people have both. Yes they are different, probably different in kind. Minds are mental and bodies are physical. And yes they interact. We all know that damage to one's body causes mental suffering, and mental activities can certainly cause bodily reactions, such as sadness causing crying.

This latter view, that there are minds and bodies, and that they are somehow different in their basic natures, is the essence of the philosophical position of dualism. There are different versions of dualism. One might maintain a substance, event and/or property dualism, depending on one's ontological inclinations. In any case, if there are both mental and physical substances, events or properties, then we must explicate the relationship between them, particularly as regards causation. The view that there exists a two-way causal relationship is the essence of interactionism. A one-way causal relationship, physical to mental, characterizes epiphenomenalism. Versions of parallelism posit no (direct) causal interaction. There is, then, a matrix of dualistic views, as expressed in Figure 1.1 below.

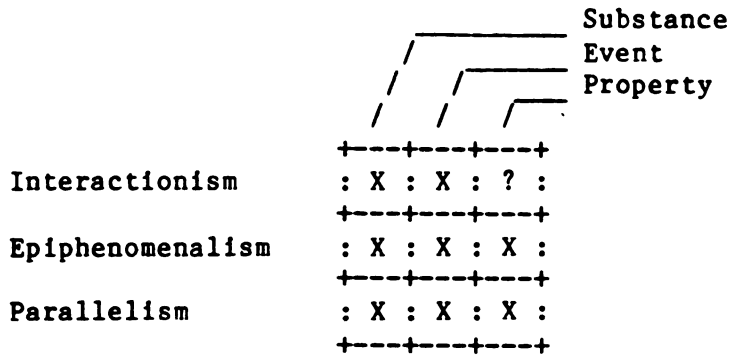

Figure 1.1 : Versions of Dualism

(Note that the `?' above is due to the oddness of properties causally interacting.)

The one question which has most troubled dualism concerns the causal interaction between the mental and the physical. How can something non-physical have a causal influence in the physical universe, or vice versa? Attempts by philosophers to give detailed answers to these and similar questions have made clear the intense conceptual difficulties that are involved. These problems have been so intractable that many philosophers have been led to abandon the dualistic outlook altogether. Still, dualism is the primary alternative to the physicalist theories which have grown up in response. It is important to understand dualism, and to appreciate its limitations, if one is going to be open minded while evaluating the physicalistic theories of later chapters.

* * * * *

Versions of Dualism

- substance/event dualism and interactionism[2]

The first dualistic view to be considered is a substance/event form of interactionism. One might maintain an ontic commitment only to events. If one accepted the existence of only substances then causal interaction would be rather difficult to account for, since it is tied to causal relations between events. Since this view is intended to parallel the second common-sense response to the mind/body problem, as indicated above, it will be construed as a substance/event dualism.

As indicated in Figure 1.1, there are two main parts to this view. It is maintained that there are two basic (kinds of) substances, inherently different in their very natures. Our bodies are physical, being composed of bone, muscle, skin, water, calcium, carbon, etc. They are always located somewhere or other in time and in three-dimensional space. They have mass, weight, shape, size, color, texture, etc. In and of themselves, they are not very different from computers, cars and rocks, except perhaps for details of composition and complexity. They are also publicly observable.

Our minds, on the other hand, are located in time but not really in space, though they somehow seem to be behind our eyes and between our ears.[3] There are many things that seem to be part of a mind. Thoughts, desires, feelings, sensations, fears, and attitudes are just some of them. Minds do not have mass, shape, color, texture, weight, or any of the properties so characteristic of our bodies. Minds are

the source of our free-will. Minds, or perhaps people insofar as they have minds, are purposive. They have a sense of morality. They feel sorrow, pain and regret. None of these things really seem to characterize a physical object, no matter how complex and sophisticated it may be. Minds and bodies are just plain different. Their defining characteristics overlap very little.[4]

The second feature of this mind-body position has to do with the causal interaction existing between minds and bodies. Our bodies are thought to be in causal contact with the rest of the physical world. Carelessly wielded hammers cause tissue damage. But our bodies are also causally related to our minds. Tissue damage causes pain, regret and anger. Anger often results in such things as swearing, and the throwing of hammers across rooms. Hence, there is a two-way causal connection posited between minds and bodies, as illustrated in Figure 1.2 below.

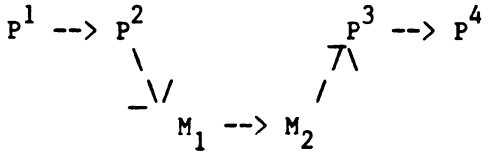

Figure 1.2 : Interactionism

In the above figure, $P^{1, 2, 3 \ldots}$ represent physical events 1, 2, 3, etc. $M_{1, 2, 3 \ldots}$ represent mental events 1, 2, 3, etc.

(The arrows are causal.)

This is the classic view of Descartes. According to this view the mind is more than a mere "pilot in the ship", since the causal connection goes both ways. It is an easy view to live with, for a while. We are both physical and mental beings. Our physicality places us in the physical world along with everything else. But our existence as mental beings places us above most, or all, of the rest of the world. Our conscious apprehension makes us special. Our minds receive a vast array of sensory experiences from our bodies, generally informing us of the state of our physical selves and of the ambient environment. We, as mental beings, can then direct many of the physical activities of our bodies. We are in control.[5]

The possibility of the continued existence of some crucial aspect of the person after bodily death, and the existence of human free will and moral responsibility, seem to follow naturally from the above view. It should be noted, however, that one could accept the view and still be a determinist. The mental --> physical causation is only a necessary condition for the existence of free will, not a sufficient one.

Comforting as all this is, the view is not without its drawbacks. Upon close analysis, the position of interactionistic substance dualism begins to break down, or at least show signs of serious structural flaws. There are arguments against both substance dualism and its interactionistic elements.[6] Many philosophers have criticized substance dualism as an unnecessary compounding of our ontology, a theory having ontological commitment to only one type of substance

being simpler. This "simplicity" argument is used against substance dualism generally, not just the interactionistic version.

A second problem concerns the notion of causal interaction between two substances which are inherently different in their very natures. Causation is a tricky business. It is quite mysterious, even when considered solely within the physical realm. Trying to characterize mental-physical causation has proven to be an exasperating experience. There may well be a good "in principle" argument against a causal interaction between any two things as different as minds and bodies. There certainly has never been a good explanation of such interaction.

From the scientific point of view, if a non-physical mind can be causally efficacious when it comes to the physical behavior of a human body, then there is a substantial amount of miraculous causation around. The physical universe is thought to be a closed system. The net amount of matter and energy remains constant. If one assumes that energy is required in a cause/effect transaction, then a mind/body causal interaction results in either a gain or a loss of matter/energy by the physical system. If the mind simply is not part of the physical universe, and energy is transmitted to the "mental side" when our bodies cause ideas in our minds, then energy has been lost from the physical universe. In cases of willing, where our minds make a decision and then cause our bodies to act in a certain way, energy is gained by the physical system.[7]

Now there are replies to these objections. When dealing with something as mysterious as mind/body causation, one can well claim that energy is not required, hence none is lost or gained. Or one could accept that the physical universe is not a closed system and just live with the consequences of such a premise for physical science. Perhaps this could be done by maintaining that there is a larger and more encompassing system which includes mental substances and which is closed.[8] One might try to do such a thing by making a case for minds being a form of energy which has not yet been recognized by science. This would not be a pure substance dualism, since matter also is thought to be a form of energy. One might even maintain that minds and bodies are two different types of packets of energy, or packets of two different types of energy. Given our present state of knowledge about these matters, the above replies must be considered pure metaphysical speculation.

- epiphenomenalism

Epiphenomenalism can be formulated as a substance, event and/or a property dualism. Typically it is formulated as a property dualism, where thoughts, feelings, desires, beliefs, etc. are viewed as mental properties of functioning neural tissue. Thoughts are viewed as emergent properties of a sufficiently complex physical structure, e.g. the human brain. Mental experiences are properties of a kind inherently different from physical properties. What we call mental experiences are just non-physical properties of our brains. Mental

experiences are "epiphenomena." Usually it is maintained that there will be (and can be) no reduction of mental properties to physical properties of the brain.[9]

A substance/event ontology posits a one-way causal influence, from the physical to the mental. On a property version, the causal relationship is replaced by the relationship between a functioning brain and its mental and physical properties.

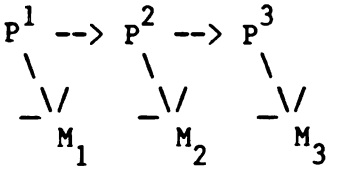

Figure 1.3 : Epiphenomenalism
$P^{1, 2, 3 \ldots}$ represent physical events 1, 2, 3, etc. $M_{1, 2, 3 \ldots}$ represent mental events 1, 2, 3, etc. (The arrows are causal or indicate the relation between the brain and its properties.)

There is a great deal to be said for epiphenomenalism. Since there is no mental --> physical causation, there are no "miraculous" causes in the physical world. There is no causal interference from outside the physical system to interrupt the physical causal chain. This is a definite plus for those who hold the physical laws dear to their hearts. Still, epiphenomenalism does allow that there is something to the mental side of humanity. There are mental experiences, at least experiences qua properties. We genuinely have

thoughts, feelings, desires and so on. We are not forced to construe these experiences as somehow purely physical, a construal which seems strongly counterintuitive to many people.

As one might expect, epiphenomenalism is not without its problems. It must either account for physical --> mental causation or make explicit what mental properties are and how they can be attributes of a physical brain. Whether one construes the mental as a type of substance, event or property, a physical substance is responsible for (the creation of) something non-physical. If the mind is an emergent phenomenon of a functioning brain, it would be most interesting to discover something about this mode of creation. It is not at all clear how one would even begin to tackle such a problem.

Another undesirable consequence of epiphenomenalism lies in the implied non-existence of human free will. One of the appealing features of interactionism is the causal control our minds can have over our bodies. If I want to go swimming (and I know how to swim, and the pool is not frozen, etc.) then I darn well can go swimming. And I often do go for a swim BECAUSE I want to. The notion of human willing is an important one. If there is no mental --> physical causation then it is very difficult to see how my physical body jumps in a pool because I will it to do so. There may be an invariable connection between certain types of mental events and certain types of physical events, but it seems that this must be a non-causal form of invariable connection; much as day following night following day is a non-causal invariable connection.

There is a partial reply to this difficulty of no free will.[10] In a "covariation" sense, one's willing to swim is both causally and logically related to one's body diving in the water. If one refers back to Figure 1.3 above, it can be seen that if both M_1 and P^2 are caused by P^1, and only by P^1, then the existence of M_1 entails P^2. Let's say that the hot sunshine causes my body temperature to rise, and this causes some particular neural event (P^1) to occur. This neural event in turn causes both my decision to go swimming (M_1) and my jumping into the pool (P^2). There is at least a logical relationship between my decision to swim and my jumping into the pool. "If M_1 then P^2" is implied by the joint premises that M_1 is caused by, and only by P^1 (if M_1 then P^1), and that P^1 also causes P^2 (if P^1 then P^2). The existence of my decision to swim logically entails the existence of the neural event which caused it. The existence of the neural event P^1 entails the existence of P^2, since physical causation involves an element of necessity. There is thus both a logical and an indirect causal connection between my decision to swim and my jumping into the pool. Admittedly this is not sufficient for anything like a normal conception of human free will. However, it does at least argue that our decisions are not totally unrelated to our actions. Under the appropriate circumstances, given a particular decision, a specific bodily action could be predicted.

A second point to be made concerning this lack of free-will objection has to do with the theory of the self which is implied in the objection. It is very typical of dualistic theories to equate the self

with the mental or conscious part of humans. The "I" which exhibits free will is the mental me. This is the sense in which I control my body. There is good reason to believe that this is at best a naive view of the nature of a self, qua decision maker and behavior controller. There is significant evidence which indicates that much of the control of our behavior is generated from a non-conscious source. There are, for instance, many examples of humans engaging in simultaneous multi-tasking, where conscious attention is devoted to one task, e.g. thinking about tomorrow's lecture, while one physically carries out a very different task, such as driving home on "auto-pilot." It also appears that there are many precognitive factors influencing, if not determining, our decisions. For example, the phenomenon of sub-liminal suggestion is well known.[11] The main point is that the objection to epiphenomenalism due to an implied lack of free will is based on a view of the self which is suspect to say the least.

One might make something of an evolutionary objection to the existence of the mind if, as per the epiphenomenalists, the mind were totally non-efficacious in terms of our behavior. If this were the way of things then having a mind could have no survival value. Traits without a positive survival value do not generate a positive selective pressure during evolution and hence do not tend to spread through the species. On this view, minds, as epiphenomenal properties, should not exist except perhaps as infrequent instances of random mutation. This objection need not cause much concern however, as it rests on an

inadequate view of the evolutionary process. Traits with positive survival value will be selected for, those with negative survival value will be selected against. However, traits which are neutral in terms of survival value may become common or they may disappear. There is no selective pressure one way or the other. Indeed, if mindedness happened to be genetically tied to some other biological trait which did convey a selective advantage, then minds might propagate rapidly. Hence, minds might well be very common among animals even though they were mere epiphenomena.[12]

It is interesting to note that we will be at a loss if we try to find a way to verify that the epiphenomenalists have it right and the interactionists have it wrong, or vise versa. Improved technology will not help here. The inherently subjective nature of the mental lies at the very heart of the problem. This is a problem of dualism generally, a detailed discussion of which will be postponed until later.

- epiphysicalism

It is possible to specify a position which might be termed "epiphysicalism" and would mirror epiphenomenalism. On this view a mind would have both mental and physical properties. That which we currently describe as brain activity would be viewed as a set of physical properties of our minds. Here there would be a one-way causal relationship from minds to bodies. Minds would be causally efficacious relative to our physical bodies, with no causal input in the other direction.

Describing our bodies, and their activities, as physical properties of a mind would not be an easy task. Free-will might be rather easy to account for, but perception of the physical world would need some work, reversing the problems faced by epiphenomenalism. Whereas epiphenomenalism posits a physical body sensing and reacting to a physical world without the aid or intervention of a mind, so epiphysicalism would posit a mind experiencing a world without the intervention of the physical organs of sense. The closest analog to such a view probably comes from eastern mysticism, where direct apprehension of the true nature of reality through meditation would seem to bypass bodily sensory systems.

One interesting point to be made here concerns the analogous move of epiphenomenalism. The claim that mental events are emergent phenomena (or properties) of a sufficiently complex physical system might seem to be a reasonable statement worthy of consideration. The claim that physical (neural) properties are emergent phenomena of a sufficiently complex mind seems much less sensible. It may be that epiphenomenalism seems more reasonable because it is embedded in a web of physical realism. As physics leads us away from a mechanistic/corpuscular view of the universe, our intuitions here may change.

The second point that might be stressed has to do more with ontology than causation. Might the activity of a sufficiently complex mind create neural events (or substances)? This approach might be termed "creationism."[13] This would be the view that we (as mental

beings) create the physical (neural) events in our brains. In its substance form it might be claimed that we actually create our brains, our bodies, and perhaps the rest of the physical universe as well. This view, though a possible one within dualism, runs so contrary to the views of contemporary society that it is difficult even to conceive of it clearly. As a purely speculative note, it might be that quantum theory will lead us in exactly this direction.

- parallelism

The term `parallelism´ can be viewed as a general descriptive term for a variety of different possible positions. Whether construed as a substance, event or property dualism, the common element or defining characteristic of parallelism is that there is no (direct) causal connection between the mental and the physical.

$$P^1 \dashrightarrow P^2 \dashrightarrow P^3$$

$$M_1 \dashrightarrow M_2 \dashrightarrow M_3$$

Figure 1.4 : Parallelism

(The arrows may be causal and/or temporal.)

On this view, one physical event may cause (precede) a second physical event which causes (or precedes) a third one. Following this chain of physical events is a parallel chain of mental events. The one thing that does not happen is to have a physical substance/event causally affect a mental substance/event, or vise versa. A needle stuck in my arm may cause a nerve to fire, but the nerve´s firing has

nothing to do with my felt pain. Nor does my decision to yell have anything to do with my yelling.

Here, mental --> physical or physical --> mental causation need not be explained because it does not exist. However, a new difficulty arises, namely how and why there is a paralleling of mental and physical events. If pinpricks do not cause pain, then why do we feel pain when stuck? There may be many ways of handling this type of question. Only two will be dealt with here.

Occasionalism posited a supreme being as a constant, causal intermediary between minds and bodies. If my arm is injured, God causes an associated experience of pain in my mind. If I decide to yell at someone, God initiates the appropriate physical chain of events. Although not relevant for the current discussion of the relationship between minds and bodies, occasionalists also viewed God as the causal agent in all cause/effect relationships, even purely physical ones. On this view, the arrows in Fig. 1.4 would be temporal arrows rather than causal arrows, since God is taken to be the one cause of all events. If there is any causal relationship between our minds and bodies, it is at best an indirect one with God as an intermediary.

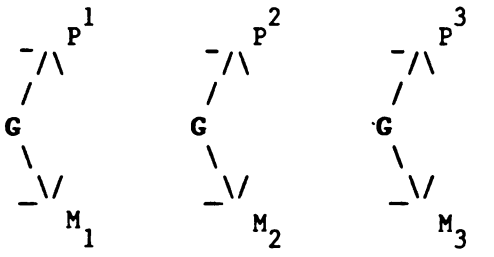

Figure 1.5 Occasionalism

From a contemporary point of view, the limitations of a view like this are not difficult to see. It offers very few advantages relative to the problems it raises. It introduces theistic notions into the philosophy of mind, and ends up with a very busy God at that. It requires a causal theory concerning the interaction between minds and God, and God and bodies. Free-will would only be apparent. Occasionalism is an historically interesting view, but it is best to dismiss it and move on.

Pre-established Harmony maintains that God set up the physical/mental universe in the beginning in such a way that the two realms would follow one another perfectly. Again there are no real causal connections between minds and bodies. The apparent connection comes about due to the perfection of the synchronization of the mental and the physical.

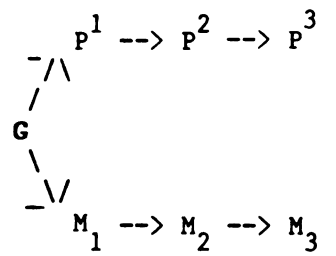

Figure 1.6 Pre-established Harmony

As with occasionalism, the problems with pre-established harmony far outweigh the advantages. It represents a logically possible relationship between minds and bodies, and an historically interesting view, but not one to be dealt with here, except in passing.

* * * * *

Problems with dualism

There are many problems associated with accepting a dualistic ontology. First of all, if one accepts either interactionism or epiphenomenalism, then one must account for the causal interaction between two things as different as minds and bodies. Parallelism avoids these causal problems, but it pays a price since it cannot explain free-will or perception. The second, and in the present context the more important, problem has to do with the nature of the mind. The mental is purported to be inherently private, subjective, and not objectively accessible. This privacy may go a long way toward explaining why the mental/physical relationship has remained such a mystery. It also underlies the difficulty one faces in an attempt to

argue for interactionism over epiphenomenalism or parallelism. If mental events are inherently subjective, accessible only to the one individual who has them, then they are ruled out of bounds for science, for any form of objective and empirical investigation. That is a problem!

Consider for a moment the question of energy transfer during mental/physical interaction. Imagine that a particularly sophisticated neurophysiologist were to be measuring the total amount of matter/energy in a normally functioning human brain. After the most careful testing, our experimenter finds that the the brain under investigation seems to be losing (or gaining) matter/energy from some unaccountable source. Would this prove the interactionist's point? Hardly! One would only need posit that the experimental technique was not up to the task. There must be some, as yet unaccounted for, source of energy transfer between the brain in question and the rest of the PHYSICAL world. The one, and probably only, way that our investigator could conclusively support interactionism would be to measure the amount of energy transferred out of the brain and into the mind. But this of course is just what cannot be done. The mind is not accessible and measurable in a way that would be required to "prove" that interactionism were right (and that epiphenomenalism were wrong, etc.).

Despite its problems, the advantages of dualism should not be overlooked. There does seem to be something thoroughly subjective about us. `Mental event' and `perceptual experience' do seem to refer to something that is different from physical objects. Talk of mental

experiences should not be given up too quickly. Dualism does seem to be getting at something that is self-evidently real and existent. It would be hasty to claim it meaningless to ponder such things as the inverted spectrum thought experiment.[14] It is not jibberish to maintain that it is like something to be human, or to be a bat.[15] The task now is to save what seems to be right without having to accept what is clearly wrong.

Chapter 2

Folk Psychology

Introduction

"...we observe very clearly...only thought alone; and consequently this notion of thought precedes that of all corporeal things and is the most certain; since we still doubt whether there are any other things in the world, while we already perceive that we think."

"...it is evident by the natural light...that we know a thing or substance so much the better the more properties we observe in it. And we certainly observe many more qualities in our mind than in any other thing...if I persuade myself that there is an earth because I touch or see it, by that very same fact, and by a yet stronger reason, I should be persuaded that my thought exists..."

René Descartes, 1644.[1]

"Our common-sense terms for mental states are the **theoretical terms** of a theoretical framework (folk psychology) embedded in our common-sense understanding, and the meanings of those terms are fixed in the same way as are the meanings of theoretical terms in general. Specifically, their meaning is fixed by the set of laws/principles/generalizations in which they figure."

Paul M. Churchland, 1984.[2]

For Descartes, that thought (cogitatio) existed was indubitable. He took the EXISTENCE of one's thoughts, and of one's self as a thinking thing, to be evident by "the natural light", known clearly and distinctly. This provided him with a philosophical foundation, something secure and certain from which to work. If the mind, and its contents are taken to be directly observable, then their existence is highly certain.

On this view, when one uses a mentalistic term in a specific instance sincerely to report an event occurring in oneself, there will

be a sense in which one is certain that there really is something going on to which the term refers. It is this certainty, this purported epistemic primacy of mental events, which is thought to justify ontic commitments to the mental, and hence to support the primary thesis of mentalism. We have indubitable knowledge concerning the existence of that to which (the first-person uses of) our mental terms refer, i.e., our own mind. Hence, we know of the existence of minds generally, since we know of the existence of at least one, namely our own. (That is not to say that we can be certain of the existence of OTHER minds, only that there is at least one.)

Dualism maintains both mentalism, that minds exist, and physicalism, that the physical world exists. However, it is typically maintained that one cannot know that the world exists in the way that one can know that one's own mind exists. In order to accept the physical world as real one must argue for its existence, as Descartes did in his Meditations. The claim here is that knowledge of the physical world is inherently theoretical, hence its existence is not certain, as is the existence of the mind.

If we cannot even be certain of the existence of the physical world, it is clear that we cannot have immediate, pre-theoretical knowledge of its attributes, i.e., of the nature of any part of that to which our physical terms refer. On the other hand, even if we can be certain that there is something to which our mental terms refer, there is still a further question concerning whether or not we have immediate and indubitable knowledge of the NATURE of that something. There is a

very real sense in which judgments are going to be involved when one attempts to make a determination concerning the character, nature or attributes of that something, the existence of which we are certain.

There are different levels of judgement possible here, beyond the mere knowledge that there is something about which we must judge. The most global judgement is it is something mental. There will be further judgements concerning the correct way to categorize that thing. We may, for example, judge that it is an event of a certain type occurring in our mind. In particular, we may judge that it is a pain.

It is very tempting to maintain that both types of judgements are indeed certain. How can I be wrong about it when I judge myself to be in excruciating pain? However, a fairly convincing case can be made to support the claim that judgements can only be made within the context of a theory. If this is so, then the judgement can only be as good as the theory in which it is made. Hence, these judgements may not be certain and indubitable.

This theory-ladenness of judgements is central to some of the contemporary approaches to the mind-body problem which claim that `mind' and `thought' are theoretical terms (even `thought' in the cogitatio sense used by Descartes).[3] There are two sides to this claim. The first is that the judgements concerning the nature of the mental are all theoretical. The second is that the basic judgement that it is a mind which I know to exist is also theoretical. This two-fold construal of mentalism as a theory has far-reaching implications. Since (in principle) theories can be false, inadequate and/or

misleading, if we view mentalism as a theory then it is at least possible to abandon it if it is found to be deficient. Doubt is thus cast upon the search for a first philosophy, for a secure foundation from which to philosophize.[4] Since `mind' and `thought' are taken to be theoretical terms, propositions about minds and thoughts cannot provide a secure basis for philosophy.

As long as our knowledge of the existence of our minds is given pre-theoretical status and as long as minds themselves are directly observable, it is difficult to see how one could abandon mentalistic ontologies. If we can be sure of the existence of the referent of `mind', then mentalism has at least some claim to certainty. Further, if the truth of such statements as "I am now in pain" can be known in an immediate, pre-theoretical way, i.e. if pains are directly observable, then the mentalistic view has a very strong claim to indubitability. On the other hand, since any theoretical judgement might be given up as mistaken if the theory itself is dropped, if mentalism is a theory, then at least specific claims about the nature of mental states are not going to be indubitable regardless of the certainty of the existence of the referent of mental terms generally. If mentalism is a theory, then it is open to assessment just like any other theory. If this is the case then the best way to approach the mind/body problem will involve the determination of the adequacy of competing theories.

* * * * *

Theories: Structure, Function and Evaluation

- structure

Theories are human constructs. Ideally they are models of reality. They are composed of interrelated terms and statements. They involve both law-like universal generalizations and singular (observation) statements. Hence, theories are closely related to language. The relationship between language and theory is both intricate and complex. Theories depend upon language for their formulation and expression. Languages grow and change in response to theoretical development. There is a reciprocity between languages and theories. Further, our way of viewing the world changes as our theories of it change, and vice versa. As our world views and theories change, our language changes in relevant ways to reflect the new ways of conceiving of things. Most of our concepts come from communication with others. They originate from, and thus depend upon, our language.

Rudolf Carnap proposed that the universal generalizations of a theory may be divided into two types.[5] First are the empirical generalizations involving observation terms. These are relatively "low-level" law-like statements which result from or express the observed regularities within the world. Such statements as "all copper wires conduct electricity" and "all iron expands when heated" are examples of what Carnap called empirical laws. These laws deal generally with macroscopic observable events/objects and are testable in a relatively direct way. The second type of generalizations are the

purely theoretical laws which do not address observables nor contain observation terms. These are "higher-order", more abstract laws which deal mainly with microscopic events. They concern such things as molecules and electromagnetic fields of force. These laws are not themselves directly testable through observation since they do not contain observation terms and thus do not say anything directly about observable phenomena. However, theoretical laws do imply many empirical generalizations. They explain such things as why copper wires conduct electricity. Hence, they are indirectly testable insofar as the implied empirical laws can be confirmed or falsified. A full-blown theory will contain both types of laws together with the singular statements.

The above construal is somewhat arbitrary, in that it relies on a clear-cut distinction between observation terms and theoretical terms. The legitimacy of such a distinction is not a debate to be gone into here. However, even if one does not accept this distinction, it may still be granted that Carnap's division is a useful one. It illustrates the spectrum of laws which constitute a theory. There is a continuum from those which are closely related to experience to those which are far removed from it. Carnap's correspondence rules, which relate the laws with no observation terms to those laws with only observation terms, seem to lie somewhere in the middle ground of the continuum. Following Quine, one might describe the observation terms, singular observation statements and the empirical generalizations as lying closer to the periphery of a theory and as constituting its

attachment to the world. Alternatively, one might view the meaning or content of empirical laws and singular observation statements as being tied into perceptual experience. Either way, Carnap's distinction between theoretical terms/statements and observational terms/statements is useful in the discussion of theories. As long as it is not viewed as precise distinction, but as a way of talking about different parts of a continuum, no serious problems should arise.

- function

A second key feature of a theory is its function. A theory is linked up with the world, and to experience, through its use in explaining and predicting observable phenomena. (This is one of the ways that the domain of a theory is specified.) This connection to observation may be either direct or indirect. Some of the work that goes on within a theory involves the working out of the logical implications which exist within the theory, e.g., the relationships between theoretical laws and empirical laws. As an example, consider the following argument.

1. Ice (solid water) is less dense than (liquid) water.
2. A solid which is less dense than a liquid will float on the surface of that liquid.

---

3. Therefore, ice will float on the surface of liquid water.

Here we have an empirical law concerning all ice and all water being derived from the more general (theoretical) statements concerning such things as density, floating and the different states of water.

Theories are also linked with observation and with singular observation (factual) statements. For example, the above general statements can then be used in conjunction with certain of our observations as follows:

4. This solid is ice.
5. This liquid is water.

---

6. Therefore, this solid will float on the surface of this liquid.

The prediction in number 6 above makes use of some of the theory's general statements (the hypotheses in numbers 1, 2 and 3) and several singular observation statements (numbers 4 and 5), stating the relevant initial conditions. A singular statement (number 6) is then deduced which addresses the phenomenon to be explained or predicted. An explanatory hypothesis (or set of hypotheses) from the theory is used in conjunction with one or more observation statements deductively to derive a further observation statement.

- evaluation

There are numerous evaluative criteria which have been presented for use when judging the adequacy of theories.[6] Since theory

evaluation has come to be an important part of the mind/body debate, it will be worthwhile to take a quick look at several of these proposed criteria.

1.) Explanatory power and Predictive success: The ability to explain and predict is a criterion for something's being a good theory. Predictions, expressed in terms of observation statements, can be deduced from the general statements (hypotheses) and the singular observation statements of a theory, in a more or less straightforward manner through the use of standard logical procedure. According to the hypothetico-deductive (H-D) method of justification, (the acceptance of) a theory is justified to the extent that the derived observation statements come out true. A statement expressing the predicted observation is given the value "truth" just when the observation which it expresses or describes actually occurs in the relevant context.[7] Theories are given a positive evaluation to the extent that their predictions succeed.

There is a sense in which explanation and prediction might be viewed as two sides of the same coin. For a theory to provide a genuine explanation of some observed phenomenon it must (at least) be the case that, given the conditions prior to that event, the theory would have predicted the event that actually occurred. On this view, explanation is just postdiction. If a theory seems to explain yet does not generate predictions, then it might legitimately be maintained that there is only an appearance of explanation, since a crucial element is missing.

There is another way of viewing explanation which extends beyond mere postdiction. Here the notion of a genuine causal explanation is tied to the ontological commitments of the theory. For a theory to give a real explanation of some phenomenon, it must be the case that the theory "has it right". There is an element of realism here, ruling out genuine explanation if the theory does not describe the world the way that it really is, even though it may work as a "useful fiction."

2.) Testability and Falsifiability: These two criteria are closely related to each other and are tied to explanation and prediction. They have been used both to delineate scientific theories from non-scientific ones, and to evaluate one theory relative to another. In order for a theory to be a scientific one it must be testable. It must generate predictions in the form of singular observation statements and there must be some possible outcome of those tests which would be inconsistent with the theory, making the theory falsifiable, at least in principle. A theory can be falsifiable in this way despite the fact that it may not be at all clear just when a theory has actually been falsified, due to the possibility of changes in ceteris paribus clauses, etc.

This testability/falsifiability criterion also addresses the richness of the predictions generated by the theory. If all of the singular statements predicted by two theories come out true, the theory which generates the most predictions will be the better theory. Further, the more predictions made the more falsifiable is the theory.

Thus, a theory is a good theory insofar as it is highly testable and highly falsifiable.

3.) Precision: There are degrees of predictive success, testability and falsifiability. In general, the more precise a theory the more testable and hence the more falsifiable it is. The more precisely it can predict the relevant phenomena the better. Theories can be vague and imprecise to the point that they are actually unfalsifiable. If it is not at all clear just what a theory predicts then it is difficult to find an observation that is inconsistent with it. One way that this difficulty can arise is if the key theoretical terms/concepts are vague and the criteria for their application are imprecise.

4.) Theoretical integration and synthesis: Mathematics, physics and chemistry integrate very well with one another and each lends support to the others. Theories which fail to integrate well with other theories which are accepted are evaluated downward. Two theories might be incompatible in this way if they imply inconsistent ontologies.

5.) Simplicity: This has also been put forward as an asset of a theory. This may be in terms of the basic ontology implied by the theory, the complexity of the formulae required by the theory, the extent to which ad hoc modifications must be relied on to account for observations, etc.

6.) Generality, range or scope: The wider the application of the theory the better it is thought to be. This explains why the reduction of one theory to another is viewed as a good thing. Such reduction often results in an "explanatory unification" [8] which involves both greater generality and theoretical integration.

7.) Conservativism and Modesty: These features have been presented by Quine and Ullian[9] as virtues of a theory. The better theory would be the one which conflicts least with our background beliefs or the one which makes weaker and more humdrum claims. These criteria are thus closely related to theoretical integration and synthesis. In certain contexts these may not be virtues. If progress occurs by the corroboration of bold new hypotheses, i.e., those hypotheses unlikely relative to the background knowledge, then presentation of a (new) hypothesis which conflicts very little with what is already accepted would be viewed as a less than momentus event.

8.) Progress: One of the key features of a good theory is evidence of advancement. A theory should be progressive. It should grow, e.g., predict and explain more phenomena and become more precise. If a theory does not grow, then either it is mature and relatively complete (correct?) or it will be viewed as degenerating. (It is useful to discuss evaluation of a theory based on progress by placing the theory in the larger context of something such as a research programme (Lakatos), a paradigm or a disciplinary matrix (Kuhn). If a theory is an integral part of such a program and the program does not progress, then this reflects badly upon the theory.)

* * * * *

Ontic Commitments

There is much to be said concerning the relationship between theory and ontology. From a realist's point of view anyway, theories do imply ontologies. The very notion of an ontic commitment suggests some form of realism. Realism can take many forms, but the main thesis is always along the lines that the statements of theories can be either true or false and that the terms of a (true) theory really do refer to existing entities. Actually the two notions, truth and reference, are quite closely related. If a theory maintains that one amp of current will flow along a circuit with one volt applied across one ohm of resistance, then the truth of that theory depends on the real existence of electrical potential, current, and resistance and on these things being related according to Ohm's law. In the present work, emphasis is on reference, not truth, and in particular on the issue of genuinely successful reference.

The alternative to realism is instrumentalism. On this view theories are thought of as "useful fictions," as uninterpreted formal systems which relate observation statements to one another. Theories are acceptable insofar as they are useful instruments or calculating devices for relating observation statements in ways which are consistent with the corresponding observations. Since questions of ontic commitment are simply ruled out by the instrumentalist, the following discussion will assume some general form of realism.[10]

Theories have a specified, or specifiable, domain and hence carry with them, at least by implication, an ontology. The ontology implied by a theory may consist of almost anything. If a theory is formalized then the ontic commitments implied by it will be those things in the universe of discourse which may count as values for the variables of existential quantification. This view is expressed clearly by Quine.[12]

"...the objects we are to be understood to admit are precisely the objects which we reckon to the universe of values over which the bound variables of quantification are to be considered to range. Such is simply the intended sense of...`there is an object x such that'. The quantifiers are encapsulations of these specially selected, unequivocally referential idioms of ordinary language...quantification being a device for talking in general of objects."[13]

A great deal can be learned about the ontology of a theory through an examination of the indefinite singular terms (the predicates) of the language in which that theory is expressed.[11] A theory will imply an ontic commitment to those objects which are required to make the statements of that theory true. For example, a theory which contains the term `a molecule' will imply that molecules exist. Electrical theory implies an ontic commitment to electrical potential, current, and resistance. According to that theory, the world is such that certain selected parts of it are best described, categorized and related in those terms.

Theories divide the world up into types. They also relate things of one type to things of other types, e.g. potential, current and resistance. Further, there are hierarchies of types. When one is

offering an explanation of some phenomenon, one is often forced out of one theory and into another, since each theory is operative only at a certain level of description. For example, the amount of current flowing in a circuit can be explained in terms of resistance and potential, but it can also be explained in terms of electrons flowing down a wire. If the topic of discussion turns to electrons then one moves out of electrical theory and into atomic theory.

Theories thus imply the existence of things, but **as members of certain classes.** Quine is quite correct in specifying quantification as that which implies ontic commitments. However, quantification and predication go hand-in-hand. Ontic commitments involve both. Theories imply the existence of things of certain types, that is of **types-of-things**, although not necessarily of types as such (perhaps not even of things as such). One cannot sensibly speak of there being something without supplying some predicate or other with which to describe or specify that thing. We never say "there is some thing x" without adding "such that...", where `...´ is replaced by a predicate.

Electrical theory implies that there is electrical potential, current and resistance. However, within that theory proper, the ultimate or fundamental nature of such things is left underdetermined. One might attempt a further explanation and maintain that there is electric charge because there are electrons (and protons). Once again though, one can leave the ultimate nature of an electron open, claiming only that there are such things in the world, whatever they are. As we shall see, the specification of such ultimate natures is a very

difficult task. The point to be made here is that theories imply ontic commitments to types-of-things as specified by their predicates.

The realization that theories imply ontic commitments to types-of-things is related to the evaluation and justification of theories. For example: a.) The simplicity of implied ontology is one desiderata of a theory. If a given phenomenon can be adequately explained within a theory which implies the existence of only one type-of-thing, then, everything else being equal, that theory will be better than any other one which implies two or more types-of-things. b.) The hypothetico-deductive method can be used to justify a theory and an ontic commitment to the types-of-things implied by that theory. If the predictions of the theory are accurate, then there is good reason for maintaining that the world does instantiate the types-of-things countenanced by that theory. c.) Two theories can be integrated insofar as the types-of-things specified by one theory can be related to the types-of-things countenanced by the other theory, e.g. current and electrons.

It is important to mention here that there may well be a significant difference between a.) the global judgements and ontic commitments to ultimate categories, characteristic of dualism and materialism, and b.) the judgements and ontic commitments to types-of-things, other than ultimate types, as countenanced by some theory or other. Consideration of global judgements and ultimate categories will be postponed until a bit later.

In the rest of this chapter, attention will be focused upon the theory of human behavior which Jerry Fodor calls our common-sense belief/desire psychology.[14] According to this theory, there are beliefs, desires, fears, wishes, wants, etc. and they are related to our behavior. Thus, there are ontic commitments to those types-of-things. For the time being, this theory will be characterized as a mentalistic theory, allowing for mental types-of-things. This is in keeping with standard practice. However, caution is urged here. It is the types-of-things other than the ultimate types which will be under consideration. Specifically, the intention is to remain neutral concerning the ultimate ontological nature of "the mental."

* * * * *

Folk Psychology

- viewed as a theory

It has been argued that our common-sense mentalistic views fit the description of a theory as presented above. This common-sense framework which we use to explain and predict the behavior of humans has been dubbed the "P-theory"[15], "folk psychology"[16], and the "autonomous inner man" theory[17]. In this work it will be referred to as Folk Psychology, or FP for short. The main idea is that FP consists of folk wisdom concerning the causes of human behavior. It is described as a pre-scientific view, whose roots are to be found in

antiquity and which has been used for centuries in our interpersonal relations. The following passage from Aristophanes' play "Wasps" (400 B.C.), depicting the power of a judge, is illustrative.[18]

> My own dominion, I maintain, is precisely Zeus' sort.
> On hearing our din, passers-by exclaim how thunderous,
> Zeus, the court!
> When my lightning I let fly and my thundering bellow,
> The rich and stately I terrify;
> They stain their garments yellow.
> You fear me, it's very clear,
> though you think you're clever.
> By Demeter I swear it's me you fear.
> But I fear you? Never!

FP can be viewed as a theory both in terms of its structure and its function. Consider the following argument:

1. People who believe in ghosts tend to avoid going into cemeteries after dark.
2. Jones has expressed his belief in ghosts many times.
3. People who repeatedly express a belief in ghosts generally do believe in ghosts.
4. Jones thus (most likely) believes in ghosts.
5. It is dark now.

---

6. Jones (most likely) will avoid going into a cemetery now.

As this example illustrates, FP consists of a set of generalizations concerning the behavior of humans. It contains observation terms, empirical laws, theoretical terms, and theoretical laws.[19] Hence, it has the structure of a theory. The generalizations (hypotheses), when conjoined with certain observation statements, can

be used to derive further observation statements about people's behavior. FP thus serves to explain and predict the behavior of our fellow humans, and hence functions as a theory. Finally, since it can be used to make predictions whose truth or falsity can be established through observation, its adequacy can be evaluated, at least relative to other theories. FP can be judged according to the evaluative criteria presented earlier.

- the adequacy of folk psychology

Characterizing our common-sense mentalistic framework as a theory has transformed the mind/body problem into a problem of theory evaluation. FP must be evaluated relative to other theories which attempt to explain and predict human behavior, e.g. behaviorism and neuroscience. Much of what has been written on the topic of the philosophy of mind in recent times concerns this very topic and most of the objections to FP can be summed up in terms of the evaluative criteria laid out earlier.[20]

As discussed above, theory evaluation is tied to the evaluation of ontic commitments. For example, the success of a theory justifies an ontology at least to some extent.[21] Hence, for the realist anyway, the ontologies of the alternative theories are also in competition. If neuroscience wins out then the ontic commitments to such types-of-things as c-fibers firing will replace commitments to such types-of-things as beliefs. An evaluation of a theory will thus also be an evaluation of its implied ontology, qua types-of-things which exist.

1.) Theoretical synthesis and integration: Folk psychology appears to stand apart from the physical sciences since its key theoretical terms are not to be found in scientific theories. One will not find `fear´, `anger´ or `belief´ in physics, chemistry, anatomy, physiology, neurobiology, etc. The few (social) scientific theories which do contain terms and concepts from FP also do not integrate very well with the rest of science. The extent to which FP terms are really separate from the terms of the theories of physical science is one topic of debate between the reductivist and the eliminativist, a topic which will be discussed in the next chapter.

2.) Progress: Folk psychology is a very old common-sense view. It has not improved much over the last several thousand years. It might well be deemed to be stagnant and degenerative. This shortcoming is especially noticeable when FP is compared to neuroscience, where progress has been rapid indeed.

3.) Precision, testability, explanatory power and predictive success: The predictive and postdictive power of FP is seriously compromised by its imprecision. The best that one can do to formulate the statements of FP is to come up with such generalizations as "People who believe in ghosts tend to avoid going into cemeteries after dark." The imprecision of hypotheses of this "tend to" construction makes prediction all but impossible, except in the most general terms.

The best that one could do to test a folk psychological claim, such as the one above, is to try to establish the probability of

someone's going into a cemetery after dark. Even this approach seems doomed to failure however, since there are such difficulties with the key theoretical terms of FP. A moment's reflection is all that is required to see that `belief' is a rather vague term. It is not always clear when to apply it, even in one's own case. The difficulties surrounding the criteria for application of the key theoretical terms rule out even a probabilistic analysis of the "tends to" construction of the FP claims. The main point is that, insofar as FP is to be evaluated as a theory which serves to explain and predict human behavior, the vagueness of its key terms places it at a disadvantage relative to its more precise competitors, e.g. neuroscientific theories.

4.) Falsifiability: FP could easily explain (postdict) why Jones would not enter the cemetery after dark. On the other hand, if Jones were to enter the cemetery at night, FP could equally well explain that. It is because he macho and was trying to impress Sally, etc. To a very large extent, and within the relevant explanatory parameters, FP can explain everything. Regardless of how one behaves, FP has an explanation for it. Unfortunately, as pointed out by Popper, this is far from a virtue of a theory. Rather, it makes the theory unfalsifiable, since regardless of the outcome of a test, FP has an explanation for it.

This unfalsifiability claim merely states the flip-side of the problems with prediction. The reason why FP can explain Jones'

entering or avoiding a cemetery after dark is because it does not predict one behavior or the other. The reason it does not predict one behavior rather than another is tied to its imprecision. Imprecision leads to untestability, which accounts for the explanatory omnipotence and the predictive impotence. This untestability makes FP unfalsifiable. (Of further importance here is the unfalsifiability of the main thesis embeded within FP, that mental states are causally efficacious relative to human behavior. This leads directly to problems with the justification of the mentalistic ontology which is implied by FP. Since this involves the justification of ultimate ontological categories, nothing more will be said here.)

* * * * *

Conclusion

The proponents of the materialistic school have made clear their views concerning the adequacy of FP. It is a degenerative, pre-scientific, antiquated, out-moded theory whose time is about up. If one accepts this materialistic view of our mentalistic FP as unacceptable, then one faces a significant methodological problem. How are we to make the transition from FP to a scientific, physicalistic theory? There appear to be three possible procedures to follow.

First of all, one might attempt a translation of the elements of FP into elements of some materialistic theory. This translation would entail an intertheoretic, type-type reduction. The mentalistic terms

of FP would be translated, via the use of bridge laws, into physicalistic terms. This is the sort of move made by the psychobehavioral and the psychoneural reductivist. A second possibility is exemplified by most versions of functionalism. A token-token identity relation is posited. Each mental event is accepted as being token-identical to some neural event or other. What is denied is a type-type reduction from mentalistic theory to a physical theory. These two views will be addressed in the next chapter. The third alternative view involves the outright elimination of FP. This is the disappearance theory of mind of Rorty[22] or the eliminative materialism of Churchland[23] and will be the topic of chapter 4.

Chapter 3

Reductivism

Introduction

The scientist concerned with human behavior is interested in formulating a theory which is precise, which is testable, which explains and predicts well, and which integrates with other scientific theories. This suggests to many that folk psychology (FP) must be replaced by a scientific theory, since FP is imprecise (lowering its predictive success and testability) and idiosyncratic relative to other, scientific, theories. The method of replacing FP under consideration here is reduction, and the associated philosophical position to be examined is reductive materialism.

There are two main components of reductive materialism. First, it proposes an intertheoretic reduction, hence it is a meta-theory concerning the relationship between two object level theories, viz., two explanatory models of human behavior. In particular, it maintains that a theoretical reduction can (and will?) occur from FP to a scientific theory. There are different versions of reductivism; each of which envisions a reduction of FP to a different type of theory. For example, psychobehavioral reductivism maintains that FP will be reduced to a behavioral theory. This view has encountered fairly serious criticism over the last few decades and will not be considered here. Psychofunctional reductivism argues for a functional analysis of FP, and psychoneural reductivism envisions a reduction to a

(completed) neuroscientific theory (NS). `Reductive Materialism' has generally been used as a proper name for psychoneural reductivism.

Second, reductive materialism involves an ontological view which denies the reality of the mental, qua ultimate type-of-thing distinct from the physical. This is, of course, one of the motivations for proposing reduction in the first place. It is also why reductive materialism has generally been conflated with the identity theory, which maintains, for example, that minds just are brains.

The general notion of intertheoretic reduction is a very complex one and cannot be explored fully here. One of the reasons that reduction is such a difficult topic is that it is closely tied to semantics and to ontological matters. Any adequate analysis of reduction would have to address theories of meaning, translation and synonymy, and also theories of reference relating to coextensionality. Even though that is far beyond the scope of the present work, a few summary remarks must be made before examining psychoneural and psychofunctional reductivism. To that end, the next section in this chapter will deal with the relationship between theories and language, focusing on the distinction between reduction and translation.

The following sections shall deal with psychoneural and psychofunctional reductivism. Since the ontological views of the materialist are distinct from any claims about intertheoretic reduction, the two topics shall be examined separately whenever possible. Some of the advantages and disadvantages of the reductivist's approach will be laid out. Two versions of the identity

theory will be examined: type/type and token/token. Along the way, a distinction will be drawn between the (weak) ontic commitments implied by theories and the (strong) ontic commitments of the materialist.

* * * * *

Theories versus Language

There is a distinction to be made between a theory and the language used to express that theory. A theory is generally characterized by its conceptual categories. Naturally theories are also characterized by their laws. However, insofar as the laws of a theory serve to define relationships between its conceptual categories, the laws constitute an explication of those conceptual categories. (The imprecision of FP is rooted in its imprecise conceptual categories, as evidenced by its vague and imprecise laws.) These categories are expressed by the indefinite singular terms of the language in which the theory is expressed. These terms denote the types-of-things in the world countenanced by the theory. Hence, these terms are the linguistic carriers of the ontic commitments of the theory. One can speak of a theory implying ontic commitments to types-of-things and of words in a language, i.e. the indefinite singular terms, as expressing those commitments. The following diagram illustrates the proposed view.

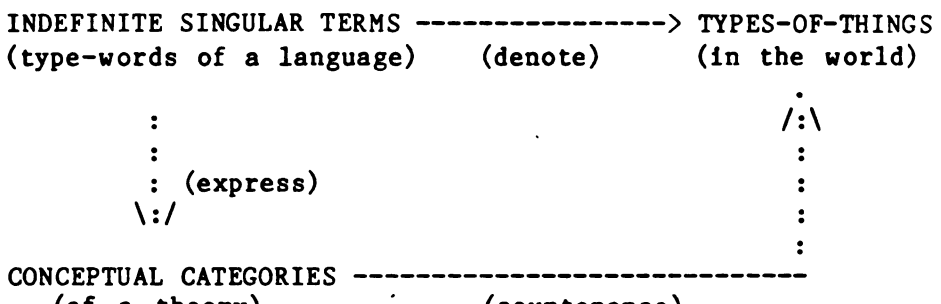

Figure 3.1 : Theories, Language and Types-of-Things

Given the linguistic turn in philosophy, one might object to this move. However, there are some advantages to making it. Failing to make this distinction would cause problems with the very notion of intertheoretic reduction. If the frameworks of the conceptual categories of FP and NS are isomorphic, and if the extensions of their respective type-words are identical, then in what sense do we have two theories rather than one theory with two separate languages? This is an important question. If it is maintained that there is only one theory, then intertheoretic reduction would be a misnomer at best. The transition from FP to NS would be no more an intertheoretic reduction than would be translating FP from one language to another. Thus, without the distinction it will be difficult to differentiate between the case where FP is translated from English to German and the case where FP is reduced to NS. But, intuitively there seems to be a big difference here. It would thus be wise to allow for a genuine distinction.

Further, distinguishing between theories and languages permits one to remain neutral concerning a syntactic or a semantic theory of

theories. Theories can be viewed as more than mere sets of sentences. They can be viewed as having cognitive content, as being meaningful, as expressing concepts and the relations between them. This allows for the construal of the conceptual categories of a theory along traditional lines. They can be thought of as relating words to the world and thus as specifying extensions. They may even be viewed as (part of) the source of intentional meaning.

Making this distinction also allows one to address intertheoretic reduction as a separate issue from the translation of one language to another, and it permits one to remain neutral concerning theories of meaning. Since the type-words express conceptual categories, the meanings of those terms will be explicated by the laws of the theory, e.g., within Newtonian mechanics, "Force = Mass times Acceleration". The EXTENT to which one views these laws as specifying the meaning of theoretical terms will depend upon the semantic views one accepts.

- reduction and translation

Since type-words express the conceptual categories of a theory, reduction and translation are intimately related. Intertheoretic reduction concerns the mapping of the conceptual categories, and hence the laws, of one theory onto those of another. A successful reduction of FP to NS will require (at least) that every theoretical term in FP be coextensional with some term in NS. Translation, on the other hand, involves establishing sameness of meaning, i.e. synonymy, between terms in the two theoretical languages. Coextensionality is a weaker

requirement than synonymy. Hence, reduction may be viewed as weaker than translation. (However, reduction also involves the laws of the two theories. If FP is to be reduced to NS, then the laws of FP must substantially map onto the laws of NS. In this sense, reduction may be stronger than translation, at least insofar as translation can be merely term-by-term.)

The extent to which translation differs from reduction will depend quite heavily upon one's semantic views. As an example, consider the situation if one assumes the main tenets of the network theory of meaning, i.e., that the entire meaning of a term is fixed by the role it plays in the theoretical framework of which it is a part; by the laws and statements within which it is contained. On this view, the distinction between reduction and translation will be minimized and perhaps disappear altogether. Both the character of the conceptual categories and the meanings of the terms expressing those categories will be determined by the general structure of the theory. Just as the conceptual categories of the theory are fully explicated by the law-like relationships between them, the meanings of the terms will be specified entirely by their interrelationships.

Given the network theory, both reduction and translation will assume an (approximate) isomorphism of FP and NS since a.) it is the type-words of FP that are to be translated into type-words of NS, b.) the type-words of a theoretical language express the conceptual categories of the theory, c.) the nature of these categories and the meaning of the relevant terms are both specified by the relationships

posited by the theory, hence, d.) if the theories are not isomorphic, i.e., if the structural relationships of the conceptual categories of FP and NS do not match-up, then reduction will fail since the laws will be different, and e.) if there is no isomorphism then the terms of FP and NS must express different conceptual categories, hence they cannot have the same meaning and translation will fail. So, translation and reduction will succeed or fail together.

However, given a different theory of meaning, it is not at all clear that reduction and translation will succeed or fail together. For example, given an extensional theory, where the meaning of a term is determined by its reference or by the entities contained in the extension of the term, translation could succeed while reduction failed. FP terms may be coextensional with NS terms, yet their laws may not coincide. (Hence, coextensionality would be a necessary but not sufficient for reduction.)

There are, thus, several advantages to distinguishing between theories and languages. If it turns out that such a distinction is unworkable, e.g., if the network theory of meaning is correct or if the semantic theory of theories is unacceptable, then one can simply conflate the levels of language and theory and the only harm done is temporary redundancy. If we make this distinction then we can at least begin to talk about intertheoretic reduction without getting bogged down in the additional problems associated with translation.

* * * * *

Psychoneural Reductivism

- intertheoretic reduction

The major claim of the psychoneural reductivist is that FP is reducible to NS. Patricia Churchland characterizes reduction

"...[as] first and foremost a relation between theories. Most simply, one theory, the **reduced** theory T_R, stands in a certain relation ... to another more basic theory T_B. Statements that a phenomenon P_R reduces to another phenomenon P_B are derivative upon the more basic claim that the **theory** that characterizes the first reduces to the **theory** that characterizes the second ... For example, the claim that light has been reduced to electromagnetic radiation means...that the theory of optics has been reduced to the theory of electromagnetic radiation..."[1]

Advantages: There are several advantages to reducing FP to NS. Psychoneural reductivism allows one to attain theoretical integration with science, greater precision, testability, falsifiability etc. by moving from FP to NS. This is certainly a primary concern for the scientist. Second, one could use the conceptual categories of FP as a starting point for research. One might, for example, go into the brain and look for pain centers. Further, by accepting conceptual categories which are at least coextensional with those of FP one allows some legitimacy for the old world view. This has the advantage of saving much of what humans have learned about themselves over the last two millennium. Finally, psychoneural reductivism has a significant advantage over such views as psychobehavioral reductivism in that its reducing theory deals with the internal (neurological) mechanisms which underlie behavior rather than being restricted to the relatively molar approach of behaviorism where the human organism is investigated as a black box.

Disadvantages: On the other hand, the proposed coextensionality of the terms expressing the conceptual categories of FP and NS, and the mapability of the laws of FP onto those of NS, presumes a rather positive evaluation of FP as a theory. The types-of-things countenanced by FP must be substantially the same as those countenanced by NS. The two theories must have a great deal in common at some very basic level where things in the world are divided up into types. It is not necessary to imply that FP is as correct as NS, nor in any way to be committed to a strong notion of truth (Truth with a capital `T') or to the view that either FP or NS is True. What must be accepted is the view that the framework of conceptual categories of FP is close enough to being isomorphic with that of NS for reduction to be feasible.

Many arguments have been presented against the likelihood of reduction based on this consideration.[2] It is indeed a very optimistic view of things; this assumption that the framework of FP concepts and laws will be isomorphic with that of NS. This is especially clear in the case of propositional attitudes.[3] From the reductivist's point of view, the neuroscientist should go into the brain looking for neural structures and functions which correspond to such things as beliefs and anger. But in fact, it may well be unduly restrictive to force such a top-down approach upon neuroscientific research.

This is further supported by the fact that there are such significant shortcomings on the part FP when it comes to explaining and predicting human behavior. Recall the discussion of the imprecision and vagueness of FP as presented above in chapter two. The

explanations and predictions of FP are significantly out of line with those that come out of contemporary NS. The prospects for reduction are arguably poor. The one bright spot here for the reductivist is that FP may simply reduce to a very small part of NS, which might go a long way towards accounting for the shortcomings of FP.

- type/type identity theory

The reductive materialist maintains that the type-words of FP are coextensional with type-words of NS. (For example, `pain´ is thought to refer to exactly the same things as `c-fibers firing´.) This is the main thesis of the type/type identity theory. This theory has been given many formulations. It has been maintained that "consciousness is a process in the brain"[4], that "sensations are brain processes"[5], and that mental states are nothing but the causes of certain sorts of behavior and hence can be identified with purely physical states of the central nervous system.[6] The key idea here is that the mind just is the brain. Mental events are (numerically) identical with brain events. What is proposed is a strict identity involving more than just spatial and temporal continuity. A sensation is taken to be a brain process just as the number seven is taken to be the smallest prime number greater than five.[7]

The type/type identity theory is thus an ontological view. Rather than making the reductive claim that FP can be reduced to NS, or the semantic claim that `mind´ can be translated as `brain´, this theory makes the ontological claim that minds just are brains. It is

important to keep the ontological claim of identity distinct from a reductive or translational claim. These claims, though closely related, are distinct. For example, J.J.C. Smart says that "...the thesis that sensations are brain processes...is not the thesis that, for example, `after-image' or `ache' means the same as `brain process of sort X'..."[8] What is claimed is that, insofar as a statement about a sensation is about anything it is really about a brain process. This is one way of claiming that `sensation' and `brain process' have the same extension, that they refer to the same type of feature in the world, if they refer at all.

The central claim here is that the (mental) types-of-things countenanced by FP can be identified with the (purely physical) types-of-things countenanced by NS. However, one of the necessary conditions for a successful psychoneural reduction is that minds really be brains. It is not possible to accept psychoneural reductivism without also accepting this type/type identity. Therefore, in what follows, it will generally be assumed that this ontological view is included as a part of psychoneural reductive materialism.

* * * * *

Weak and Strong Ontic Commitments

The reductive materialist rejects the dualistic ontic commitments of FP, but accepts the conceptual categories of FP and thus the types-

of-things which it countenances. Hence, according to the reductive materialist there is a sense in which FP and NS imply the same ontic commitments since they posit the same types-of-things. However, there is also a sense in which they imply different ontic commitments. Accordingly, one might view the identity theory associated with reductive materialism as positing one or both of two very different ontological claims, both of which concern types-of-things. These two forms of identification involve what will be called weak and strong ontological commitments.

As a weak ontological claim, the type/type identity theory maintains that the FP and NS posit the same types-of-things in the world. They divide up the world in the same way. This is the sense in which FP and NS must be isomorphic for the reductive strategy to succeed. Weak ontic commitments are tied to theories.

These intratheoretical, weak ontic commitments are closely related to what Carnap called "internal questions" of existence,[9] and to what Quine calls "sub-class" questions.[10] Consider the following two questions, a.) "is there electrical current?" and b.) "are potential, current and resistance related according to Ohm's law?" If these are formulated within electrical theory, then for Carnap if b.) is true then it must be empirically true, since it has to be established through testing. On the other hand, a.) is logically true within electrical theory, since the reality of a thing is implied by a theory that contains a term which purports to refer to that thing **as a member of a certain class.** In essence, if a theory contains a specific type-

word, then, insofar as one is working within that theory (i.e., accepts that theory) it makes very little sense to question the existence of the types-of-things to which that term refers.

Consider the weak ontic claim that is being made within the context of electrical theory. There is something (electrical current) such that it is related to other things (potential and resistance) in a certain specifiable way (Ohm's law). This is very much like what U.T. Place called the "is of predication".[11] Current is the kind of thing that is related to resistance and potential as indicated in Ohm's law. It might be formalized as I=E/R, E=IR, or R=E/I. One could not say however, again following Place, that current is the kind of thing related to potential and resistance according to Ohm's law, AND NOTHING ELSE. Current is also the kind of thing that passes through filaments in light bulbs, is either alternating or direct, etc. What the weak sense of ontic commitment picks out are the types-of-things whose relationships are described through the use of predicates in the theory. There is something (whose basic nature is at this point left undetermined) such that it is related to other things thus and so.[12]

Clearly a theory depends upon its associated ontology, qua weak ontic commitments, for its identity since the ontology is implied by the the conceptual categories which characterize that theory. Can one change the indefinite singular terms of a theory without changing the implied ontic commitments and hence the theory itself? If one changes the indefinite singular terms of the theory then clearly something about the theory changes. However, the change may be trivial. For

example, merely substituting `glub' for `electron' will not really change anything significant. The change may also be non-trivial. Try substituting `demon' for `charge', while retaining the mythical meaning of `demon'.

The mythical meaning of `demon' points to the fact that there is a stronger sense in which one might be committed ontologically. One might claim that there are electrons, and make a further claim that electrons are physical substances. Or one might claim that there are two basic kinds of things in the world, mental substances (or events) and physical substances (or events.) It is this strong sense of ontic commitment which serves to characterize dualism, materialistic monism and the inherent differences between them. The defining characteristic of a strong ontic commitment is the specification of the nature of the basic constituents of the members of one's ontology, of the ultimate ontological categories, or of the most basic and fundamental level of description.

The strong sense of ontic commitment is more like what Place termed the "is of composition". Just what exactly is current? Well, it is electrons in motion along a copper wire, for example. Given that answer, one could next ask "just what are electrons?" The strong sense of ontic commitment addresses such questions as "of what are electrons (or protons, or neutrons, or quarks) composed?" Even if we found the ultimate constituents of matter, the true atoms in the original sense of the term, the question could be asked "just exactly what are these atoms made of?" or "what kind of stuff are atoms?" To say that these

atoms are physical, corporeal, or material substance is to make a strong ontological claim.

Strong ontic commitments do involve types-of-things, yet they are quite different from weak ontic commitments. Strong ontic commitments are not implied by object-level theories such as NS and FP. One will not, for example, find the terms `physical substance´ playing a crucial role in NS. There is no claim that "everything is physical" in NS. Questions of strong ontic commitments are more akin to Carnap´s "external questions"; they are thoroughly philosophical and non-empirical. Weak ontic commitments are implied by (scientific) object-level theories, and our acceptance of those theories argues for the acceptance of the existence of the types-of-things which they countenance. Strong ontic commitments are best viewed as either being made by people or implied by second-order (non-scientific?) theories. Later, in chapter 5, they will be construed as part of a paradigm.

Strong ontic commitments are independent of weak ones. Regardless of one´s strong ontic commitments, one could either accept or reject the conceptual categories, and hence the weak ontic commitments, of a theory such as NS or FP. The reductive materialist rejects the (mentalistic) strong ontic commitments of FP while accepting its weak ones, qua explanatory model of human behavior. One could just as well view FP as a substantially incorrect theory, one which misconstrues reality, or fails to explain and predict human behavior, and still accept dualistic strong ontic commitments.

It is possible to specify two senses of strong ontic commitment, a negative one and a positive one. One might make a negative strong ontic commitment by rejecting the existence of a certain type of substance, e.g., mental things or events. Thus, one could be a non-mentalistic monist. One could also make a positive strong ontic commitment, e.g., by maintaining a materialistic monism and claiming that there is only one type of stuff in the world and it is physical. One might even accept a non-committal strong ontic commitment, maintaining that there is only one basic kind of stuff without going so far as to attempt to characterize it as mental or physical.

The range of possible ontic commitments extends from the total non-commitment characteristic of instrumentalism to the positive strong ontic commitments of Cartesian dualism. One could make weak commitments without strong ones or negative strong ones without positive strong ones. The strong position that one takes is largely independent of the weak commitment to which one adheres. There may be some difficulties in adopting what are prima facia incompatible strong and weak positions, e.g. positive strong materialistic monism and the weak ontic commitments of FP. However, the weak position that one takes is essentially a decision concerning the adequacy of the conceptual categories of different theories. The following figure summarizes the possibilities.

STRONG ONTIC COMMITMENTS		WEAK ONTIC COMMITMENTS: None	NS	FP
None		X[1]	X	X
Negative	non-committal monism	X[2]	X	X
	non-mental monism	X	X	X[3]
	non-physical monism	X	X[3]	X
Positive	materialistic monism	X	X	X[3]
	mentalistic monism	X	X[3]	X
	dualism	X	X	X

Figure 3.2 : Weak and Strong Ontic Commitments

1. This is instrumentalism, neither weak nor strong commitments.

2. One could consistently adopt any strong view and still deny that either FP or NS were correct.

3. If one were to accept non-mentalistic monism and FP, or a non-physicalistic monism and NS, then one must reinterpret the prima facia mentalistic or physicalistic words expressing their respective categories. This is essentially the reductivistic strategy.

Even in the weak sense of ontic commitment, a theory might be thought of as attempting to get at the true nature of reality. Even if we do not know just exactly what electrons, protons, charge, potential, amperes and resistance are, knowing that they are something such that they are related according to Ohm's law is to know something about the world over and above the mere ability to account for our observations. We have at least a partial explanation (in the realist's sense) for why we observe what we do. Thus, rejecting the strong sense of ontic commitment does not necessarily lead to instrumentalism. There are thus two senses of realism, weak realism and strong realism, corresponding to the two senses of ontic commitments.

Psychoneural reductivism, as a view concerning the relationship between FP and NS, could be construed in a way which is neutral relative to strong ontic commitments. However, as traditionally presented it is a de facto physicalistic monism. When it is claimed that minds are identical with functioning brains, what is really meant is that there are only brains and brains are physical. (Technically, with an identity relation, one could say that a brain event is a mental event just as easily as one could say that a mental event is a brain event.) Proponents of reductive materialism make a strong ontological claim, one which involves the rejection of dualism and/or mentalism and the acceptance of physicalism or materialistic monism. The materialistic monist allows for only one ultimate type of stuff, maintaining that everything that exists is physical. When one conjoins psychoneural reductivism, the type/type identity theory and the strong

ontic commitments of materialistic monism, the full-fledged reductive materialism emerges.

- more objections to psychoneural reductivism

There are various objections to reductive materialism which stem from these ontological considerations. First of all, the reductive materialist seems to miss one key feature of such things as pains. Pains hurt! They have a certain feel to them. There is a fairly specific and identifiable phenomenal quality about them. (The technical term `quale´, or `qualia´ in the plural, has been widely used to denote such phenomenal qualities of experience.) Familiarity with the experience of pain seems to many to be an essential ingredient in the true understanding of the meaning of `pain´.[13] NS does not seem to recognize or allow for these phenomenal qualia. It might be claimed that the prima facia physicalistic ontology of NS misses the proverbial boat right here.

A second objection is that, at least in one´s own case, one knows very clearly when to apply the term `pain´. When it hurts! No brain state and no physical behavior need be observed. I certainly do not need to observe my own behavior in order to know when I am in pain. Hence, the conceptual categories of FP seem to capture something which NS misses. An organism is "aware of" its internal states in a way that does not seem to require observation; certainly not overt behavior and probably not internal neural behavior.[14]

Cases such as this illustrate the difference between weak and strong ontic commitments. Within the context of the strong dualistic ontic commitments of FP, it makes sense to ask what it is like to feel pain, and it does not make sense to ask how one knows one is in pain. What is being picked out here is the qualitative character of the perceptual experience of a pain, and that is something to which the subject of the pain has direct and immediate access.

If reductive materialism is correct, then the quale of excruciating pain just is something like c-fibers firing together with certain cortical activity. The claim of the reductive materialist is not merely that NS is a better theory than FP, in terms of the explanation and prediction, progress and theoretical integration. The claim is that the quale that is associated with being dipped in boiling hot water is nothing other than a neural event. What FP refers to with `pain' is exactly the same thing that NS refers to with `c-fibers firing'. And this is not just in the weak sense of a type-of-thing in the world, but also in the strong sense of mental versus physical. This really is the main point of contention between dualists and materialistic monists.

If `mind' or `mental event' is taken to denote what a dualist/mentalist wishes to denote by it, then the extensions of such terms are given an ontological status which may well violate some of the background assumptions of the physicalistic sciences. According to the mentalist, minds are a special sort of entity. They are inherently non-physical and private, and thus may well be, in principle, beyond

the bounds of the empirical investigation so essential to physical science.

Consider, for example, the free-will/determinism debate. Minds are typically viewed as being unconstrained by the laws of the physical universe. Within a dualistic FP there will be generalizations to the effect that minds are free to act independently of physical laws. This generalization will have to be dropped if one identifies minds with purely physical brains. Thus, the intended referent of `mind´ is closely tied to positive strong ontic commitments. Identification of minds and brains is ruled out by the mentalist since minds are free and brains are not. This will generate problems for a psychoneural reduction due to a failure of the two theories to be isomorphic and coextensional.

There is a further problem with the type/type identities associated with reductive materialism. Consider a claim such as "pains just are c-fibers firing." If pain is taken to be type identical to c-fibers firing, then if a creature does not have c-fibers that creature cannot be in pain. Thus, psychoneural reductivism rules out the possibility of any creature substantially different from ourselves having pains. The somewhat fanciful thought experiment of a visitation by extra-terrestrial creatures, who lack c-fibers, illustrates the absurdity of this view. Such creatures could not legitimately be said to be in pain, regardless of how they behaved after bodily injury. The point is a general one and encompasses the other conceptual categories of FP. For example, regardless of how sophisticated computers become

they can never have a belief. All in all, this is not a very desirable consequence and constitutes one of the traditional objections to psychoneural reductive materialism.

* * * * *

Psychofunctional Reductivism

The theory to be considered here is metaphysical functionalism (or the functional state identity theory) which characterizes mental states as functional states.[17] A type of mental state (e.g. pain) will be characterized in terms of its relationship to the (sensory) input, other mental states, and the (behavioral) output of the organism in question. Whatever internal state fulfills a given functional role in a particular creature will, by definition, be that mental state (e.g. pain).

In what follows, `functionalism´ will be used descriptively for psychofunctional reductivism, rather than as the proper name of a particular position. The main difficulty with using it as a proper name is that there does not seem to be one unique position going by that name. The risk of working with a composite view is that it may end up being a "straw man". To avoid that, an attempt will be made to present the strongest possible version of functionalism by incorporating the central theses of its various formulations.

According to functionalism, terms signifying mental states are extensionally equivalent to terms specifying functional states. Functional states are dispositional states. Thus, mental states are dispositional (functional) states.

Functionalism has much in common with psychoneural reductivism and with behaviorism. It focuses on the coextensionality of FP terms and terms which constitute functional state descriptions of purely physical systems. Operational definitions may come into play here, as they do in behaviorism. Functionalism stresses both the dispositional account of human behavior, as presented by the molar behaviorist, and the internal/structural account of NS as endorsed by the psychoneural reductive materialist. There are, however, significant differences.

- type/type identity

Functionalism accepts a special version of the type/type identity theory. It identifies mental types with functional types. It accepts the weak ontic commitments and the conceptual categories of FP. It accepts, for example, the claim that there are such things as pains and desires to stop pain, and that these states are related in specific ways. In this it is in agreement with the psychoneural reductive materialist. However, whereas the latter maintains that types of mental states are just types of neural states, functionalism maintains that the mental types of FP just are types of functional states. (Thus the name "functional state identity theory".)

One of the features of functionalism which distinguishes it from psychoneural reductivism is the claim that the mental events of a given type can be variously instantiated physically. Thus, under functionalism, "being in pain" might be correctly ascribed to humans and to other creatures very different from humans. Under psychoneural reductivism, if those creatures were such that they had no c-fibers to fire, then "pain" could not be ascribed to them. Psychofunctional reductivism avoids that problem.

It should be noted that functionalism does not merely propose a "species-specific" type/type identity theory. It does not, for example, merely claim that since all humans share the same type of physical system, "human pain" could be identified with human c-fibers firing, whereas "Martian pain" would have to be identified with some other type of physical event. Granted, this would overcome one difficulty with psychoneural reductivism, where pain is ruled out in the case of creatures without c-fibers.[18] However, human pains and pains in creatures without c-fibers have something in common, by virtue of which they are all pains. That which they all have in common is their functional roles in their respective physical systems. The claim of functionalism is that human pain, Martian pain or any other pain, is what it is because of the type of functional role it plays, or because of its causal role in the behavior of the individual in question.[19]

This type/type identity theory associated with functionalism characterizes a mental state in terms of a functional or a causal role. Such states may be instantiated in a particular physical system in a

particular way. But that is merely a contingent matter of fact, relative to the classification of the state, e.g. as being a pain. What is crucial, what makes the mental state what it is (e.g. a pain), has to do with its functional/causal role.[20] It is this functional/causal role that makes a particular state a pain. Any given mental state countenanced by FP just is some functional state or other. Mental states are functional states.

In principle, these functional states could be non-physically instantiated, so long as the correct functional/causal role in a system were served. This is consistent with the independence of weak and strong ontic commitments. However, functionalism generally maintains, as a de facto standard, that the only systems which exhibit such mental/functional states are in fact physical systems. Thus, each particular functional state is some particular state of a particular physical system. Hence, it is most reasonable to portray functionalism as a form of physicalism.

There is, then, a sense in which mental states, events or experiences are real under functionalism. They are tied to the world (causally) through the input/output of the organism which has (or instantiates) them. They are at least real in the weak ontic commitment sense, as types-of-things in the world as countenanced by the conceptual categories of FP.

Even though functionalism accepts the same weak ontic commitments as FP, there is a sense in which functionalism differs from psychoneural reductive materialism. According to functionalism, mental

states are second-order properties; they are types of types. In humans, one type of physical event (i.e. c-fibers firing) is a pain. In martians, a different type of physical event may be a pain. But both types of events are correctly classified under the (second-order) type pain. The following diagram illustrates the point.

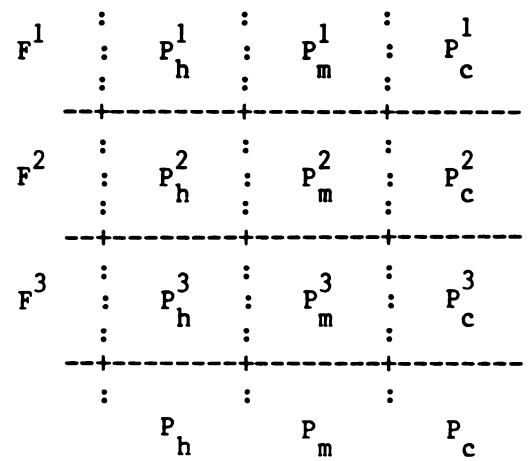

Figure 3.3 : Second-order Types of Functionalism

In Figure 3.3, each row F^i represents a type of functional/mental state: e.g., F^1 is pain, F^2 is joy, and F^3 is anger. Each column P_k represents a type of (physical) "organism", e.g. P_h for humans, P_m for martians, and P_c for computers. Each P_k^i represents an i-type of physical and functional state within a k-type of organism. (Each P_h^i thus represents a column as presented in Figure 3.4 below. Hence, P_h^1 might consist of c-fiber firings. Other diagrams, similar to Figure 3.4 could be used to represent the various types of states of the "organisms" P_m and P_c. See, for example, Figure 3.5 below.)

Given this scenario, each token thing (e.g., a token pain) which falls under a functional type (e.g., pain) also falls under a physical type (e.g., human c-fibers firing). However, mental states are functional states of physical "organisms," and hence are second-order physical properties. This is a main point of disagreement with the reductive materialist, who identifies a mental state with a first-order physical property (e.g., pains with c-fibers firing).[21]

The functionalist admits the reality of such things as beliefs and pains, thus accepting the weak ontic commitments of FP. In this he is in agreement both with the proponent of folk psychology and with the psychoneural reductivist. The functionalist's strong ontic commitments are materialistic, in contrast to FP but in agreement with the psychoneural reductivist. In order to bring out the points of disagreement between psychofunctional and psychoneural reductive materialism, a second version of the identity theory must be presented.

* * * * *

Token/Token Identity Theory

Until now, coextensionality has been discussed with regard to the type-words of FP and NS. Psychoneural reductivism maintains that the types-of-things referred to by FP terms are identical to the types-of-things referred to by NS terms. For example, `pain' and `c-fiber firing' pick out exactly the same type-of-thing, the same set of things in the world. There is, however, a sense in which one might maintain

an identity theory without accepting this coextensionality of type-words. One might maintain a token/token identity theory while rejecting type/type identities. On this view, each token pain, i.e., each particular pain, will be a token physical state. However, it may not be the case that all pains, qua physical events, are the type-of-thing denoted by `c-fiber firing'.

The type/type identity theory makes a significantly stronger claim than does token/token theory. Type identity implies token identity, though token identity certainly does not imply type identity. The possibility of token-identity without type-identity is illustrated below.

M_1	[a]	[b]	[c]
M_2	[d]	[e]	[f]
M_3	[g]	[h]	[i]
	P_1	P_2	P_3

Figure 3.4 : Token/Token Identities

In the above figure, each row represents a type of mental event as specified within FP, each column represents a type of physical brain event as specified within NS, and the letter in each box represents an individual or token event occurring in the world. None of the mental types correspond to any of the physical types. For example, the mental

type M_1 consists of the token events a, b and c, and does not correspond to any physical type. Still, each individual event falls under both a mental type and a physical type. For example, b is an M_1 type of mental event and a P_2 type of physical event.

- psychoneural reductivism

Generally speaking, the token/token identity theory has also been presented as a de facto physicalism which rejects mentalistic positive strong ontic commitments. Given the acceptance of this sort of token/token identity, one might either accept or reject the weak ontic commitments of FP and/or of NS. Further, this view allows one to accept the identity of mental events and brain events without thereby also accepting either a type/type identity or psychoneural reductivism. There are at least three possible variations of this position. One might:

1.) Accept the weak ontic commitments of both FP and NS: It might be maintained that no reduction can occur because the two theories are describing some very different aspects of that to which `mind´ and `brain´ refer. They might both be "right", in the sense that their respective conceptual categories accurately reflect the world, but the overlap is partial at best. There are different ways that this might work.

a.) It may be due to the fact that each theory is incomplete. However, if this were the situation then the failure to reduce might be only temporary since a completed neuroscience might well capture all of

the details of the brain, including those presently described only in mentalistic theory.

b.) It may be because FP and NS are describing the world at different levels. Consider the following analogous example. Some theory, say T_1, might have many dispositional terms such as `delicate' as type-words. Finely-blown glass vases, sand-dune ecosystems, fresh flowers and wheatstone bridge circuits might all be correctly described (within T_1) as delicate. Yet, T_1 might say nothing about molecular structures, electrical potential, current or resistance, or the susceptibility of the roots of certain grasses to mechanical damage. A second theory (T_2) might have no dispositional terms but address these latter issues instead. There will be no type-word in T_2 which will correspond to `delicate' in T_1. Yet, it may well be the case that for each thing in the world described in T_1 as delicate there is some type-word or other in T_2 to cover it. (This is very close to the view of the functionalist.)

2.) Reject the weak ontic commitments of either FP or NS: One theory might be thought to contain some serious conceptual errors. If one theory were to be significantly wrong and the other right,[15] then the reduction of either theory to the other would be doomed from the start. The "wrongness" of one theory might be due to a referential failure or it may be due to a poor showing along the lines of the criteria for evaluating theories presented in chapter 2. Again there are several ways that this might occur.

a.) There may be a failure of the type-words to refer to "legitimate" types of things in the world. The terms expressing the conceptual categories might not refer to anything at all. Or, they may refer to sets of things, the members of which do not have any, or at least not the intended, features in common. They may not refer to "natural kinds", if one accepts such a notion.

b.) On the other hand, the conceptual categories of one theory might simply be very poor, at least relative to some other theory and to some set of goals or other. The conceptual categories of a theory might not get us anywhere. This is a pragmatic consideration related to the evaluative criteria of theories. (This is generally the view of the eliminative materialist.)

3.) Reject the weak ontic commitments of both FP and NS: One could accept or reject the basic identity of the intended referents of FP terms and NS terms and yet reject both FP and NS as fundamental misconstruals of reality. Here reductivism would be a moot point at best. (This view may best describe the position that would be accepted by both the eastern mystic and by a proponent of the relativistic quantum model of the universe characteristic of contemporary western physics.[16])

- functionalism

Functionalism maintains that mental/functional types are instantiated in various (physical) systems. All types, whether first-order or second-order, are ultimately grounded as states in physical

systems. What MAKES something a pain is its being a second-order physical property. But the fact that there is something to be categorized at all (as a type) depends on there being some physical system or other in a specific physical state. Functionalism maintains that there are types of TYPES, and that pain is such a second-order type. There is thus, as described by N. Block[22], a metaphysical disagreement between functionalism and psychoneural reductivism. However, both agree that the only types of THINGS that there are are physical. Thus there is an ontological agreement, viz., first-order types and strong ontic commitments.[23]

Thus, it is legitimate for a proponent of functionalism to maintain that each particular pain "just is" a particular state of some physical system. They thus posit a token/token identity theory.[24] This can be presented independently of arguments over whether or not a pain is a first-order or a second-order type.

The following diagram illustrates the point.

[a]	[b]	[c]	[a′]	[b′]	[c′]	[a″]	[b″]	[c″]
[d]	[e]	[f]	[d′]	[e′]	[f′]	[d″]	[e″]	[f″]
[g]	[h]	[i]	[g′]	[h′]	[i′]	[g″]	[h″]	[i″]
P_1	P_2	P_3	P_1	P_2	P_3	P_1	P_2	P_3
(humans)			(Martians)			(computers)		

Figure 3.5 : More Token/Token Identities

In the restricted universe of Figure 3.5, the set of tokens which constitute all of the pains are represented by (a, d, g, a′, d′, g′, a″, d″, g″). Each column P_1 represents the species-specific physical type of event identical to pain in that species, as a type of physical "organism". The main point of disagreement between functionalism and reductive materialism turns on whether pain should be represented as one or more columns in Figure 3.5, or whether it is a (second-order) type to be represented as a row in Figure 3.3 presented earlier. What they do agree on is that, however represented, token-pains are instantiated only as physical states.

* * * * *

Conclusion

Functionalism, with the token/token identity theory, has certain advantages over psychoneural reductivism and the type/type identity theory. The proponent of the token/token identity theory can maintain that each pain is something physical and yet avoid the problem of pains being idiosyncratic to humans. Since functionalism denies a first order type/type identity, it need not support the claim that FP is reducible to NS. If there is to be any reduction of FP, it will be to the second-order physical types of functionalism.

One major problem with the token/token theory is related to the notion of the qualitative character of perceptual experience. In particular, the identification of (an individual case of) excruciating pain or a sensation of bright red with some physical and (in our case) neural state or other will require justification. The token/token identity theory is no better off here than is the type/type theory. The view of the functionalist, that any token pain just is a specific instantiation of a functional state within a particular physical organism, seems to miss the hurtfulness of pains just as badly as does the view that a pain is just a c-fiber firing.

It should be recognized that the force of the above objection comes from the (at least tacit) acceptance of the positive strong ontic commitments of FP. As long as one is caught up in the attempt to reduce FP, or to specify either type or token identities, one will be faced with such charges from the proponent of FP. Insofar as mentalistic or dualistic positive strong ontic commitments play a role in the meaning of key terms within FP, e.g. `pain', the reductivist/identity theorist is going to have to address the claim that there is something about, for example, pains which the (reductive) materialist misses.

One way that has been proposed to avoid all of these problems is to abandon the reductivist's approach altogether. If that is the approach taken, then one may not need even to address what the proponent of FP calls pain. The reductivist/identity theorist must

answer the question "what is pain". This is where the counterintuitive responses occur, e.g., a c-fiber firing or a functional state. Proponents of the position termed `eliminative materialism´ maintain that by dropping FP altogether, by simply eliminating the theory and its proposed ontology, we are free to claim that pain is nothing at all. It is thus to the view of eliminativism that we now turn.

Chapter 4

Eliminative Materialism

Introduction

The view to be discussed in this chapter is eliminative materialism (EM). It is a view which has come about in response to the failings of folk psychology (FP) and of the different versions of reductivism. EM is motivated by an analysis of FP as a theory which is deficient in terms of the evaluative criteria discussed in chapter 2. The desire for elimination rather than reduction concerns the problems with reductivism, discussed in chapter 3. EM does share the positive strong ontic commitments of reductive materialism and of functionalism, namely a physicalistic monism. The main contrast between eliminativism and reductivism concerns their respective views of the merits of the conceptual categories, and hence the weak ontic commitments, of FP and the proposed method of transition from FP to a physicalistic theory, see Figure 4.1 below. Although functionalism, and its cousin behaviorism, could perhaps be given eliminative formulations, in what follows, `EM´ will be used to stand for psychoneural eliminativism.[1]

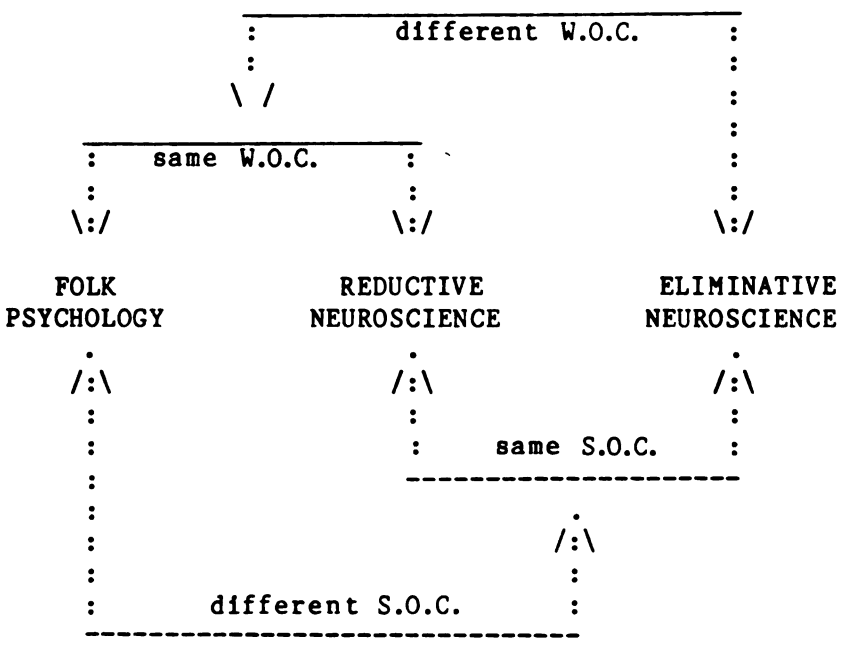

Figure 4.1 : Reductive and Eliminative Ontic Commitments

Note: "W.O.C." = "weak ontic commitments"
"S.O.C." = "strong ontic commitments"

There are three main lines of argumentation for elimination and against reduction, which have been presented by proponents of EM, e.g., by Paul M. Churchland.[2] One argument concerns scientific methodology and appeals to the desirability of a bottom-up approach in the neurosciences, an approach which is implied by eliminativism. The alternative is a top-down approach which is implied by reductivism. If the neuroscientist is a reductivist, then the correct procedure in the lab involves finding the neurological correlates for such things as beliefs, fears, and sensations of red. The eliminativist is free to abandon such, presumably ill-fated, endeavors. One can "start afresh"

so to speak and engage in brain research without the undue constraints of an old and outdated theory. One is free to develop the conceptual categories which seem appropriate as one's research progresses.

The reduction of one theory to another really is a rather demanding endeavor. In many respects, elimination is an easier approach. In this sense the eliminativist makes a weaker claim than does the reductivist. Even if the claim is only that the conceptual categories of FP are possibly wrong, there is still something to be said for starting over. If we end up with roughly the same conceptual categories, that will be fine. The claim here is that it is not the best approach to demand that NS mimic FP in the way that it divides up the world.

The second argument concerns the feasibility of reduction from FP to NS. As discussed earlier, if one is to reduce one theory (T_R) to another (more basic) theory (T_B), then if T_B is substantially correct, T_R must be basically right as well. Alternatively, if one wishes to avoid notions of correctness, this objection can be reformulated in terms of the isomorphism of theories. If FP is structurally so very different from what a completed NS will be like, then there is little chance of an isomorphism and reduction is quite unlikely. The idea here is that if FP is so fundamentally wrong or so very different from NS, then the prospects for reduction are minimal. If reduction is not feasible then an outright elimination of the former theory in favor of the latter is called for.

The third argument supports the first two and involves an historical perspective. There are many pre-scientific theories which proved to be so vastly incompatible with scientific views that they were simply dropped. Consider the many examples of theories which have been eliminated.[3] The caloric theory of heat was not reduced to the kinetic theory. The phlogiston theory was dropped and replaced by modern views of oxidation reactions. Theories of demonic possession and witches have now been replaced with theories of neurochemical imbalances. In these and similar cases the previous theory turned out to be, at least from the point of view of the later theory, so radically mistaken that reduction was simply out of the question.

The eliminative materialist maintains that the prima facia evidence is that folk psychology will suffer the same fate as the other pre-scientific "folk theories", namely elimination. This is an especially forceful argument when one considers the domain that the theory is meant to explain. The explanation and prediction of human behavior has turned out to be a complex affair indeed. Given the intricacies and complexity of the human organism, especially our brains, it is fairly obvious why our behavior is so difficult to understand. The likelihood that our ancestors would have gotten it right on this most difficult of topics when they were so wrong about virtually everything else is low indeed.[4] The view of the eliminative materialist is that FP is too wrong to be reduced to the theories of physical science. The type-type identities and the (approximate) isomorphism required for intertheoretic reduction do not exist, nor

should we spend our time trying to find them. On this view, "...our common-sense psychological framework is a false and radically misleading conception of the causes of human behavior and the nature of cognitive activity."[5]

At this point we have mainly negative arguments against FP and reductivism, and a suggestion as to a promising new approach. It is now time to consider some positive arguments in support of EM, mainly as presented by its strongest contemporary exponent, Paul M. Churchland. Two related views are involved in the support for EM. The first concerns a view of perception as a theory-laden activity. The second concerns the network theory of meaning. Both views are central to Churchland's arguments in favor of eliminativism. Both come into play in his attempts to anticipate key objections to EM. One problem, which will be examined soon, concerns the prima facia absurdity of the denial of the existence of mental states (in the mentalist/dualist positive strong ontic sense of "mental"). The discussion of this problem will center on an analysis of the qualitative character of perceptual experience, i.e., the phenomenal qualia, and on the relationship of EM to the token/token identity theory.

* * * * *

The Theory of Perception

"Perception consists in the conceptual exploitation of the natural information contained in our sensations or sensory states".[6] This is one of the most explicit statements of Churchland's theory of perception. The main questions here have to do first with the nature and source of this conceptual exploitation, and second with the extent to which it is efficacious in the perceptual process. His main claim is that this exploitation proceeds within a theoretical framework and that the framework that is currently in use, FP, is inadequate. The foundation for our conceptual exploitation of the information contained in the states of our sensory systems does not come directly from the world which we perceive, and certainly not from the phenomenal character of experience. Rather this foundation lies in the background theory within which we work.[7]

The first step in understanding Churchland's theory of perception is to examine what he calls the "objective intentionality" and the "subjective intentionality" of perception. The process of perceiving the properties exhibited by the external world begins with one's sensory equipment, one's physiological capacities. Herein are the biological origins of our sensations. Our physiological hardware partially determines the information we can acquire about the world. Our sensory equipment constitutes a limiting parameter for perception. The actual information naturally contained in our sensory states is called the "objective intentionality" of our sensations.[8] "The objective intentionality of a kind of sensation consists in its being a

reliable empirical indicator of the presence or value of some environmental feature or parameter..."[9] It is a relational and contingent matter. It depends on the anatomical structure and the physiological capacities of our sensory systems and upon what features in the external world happen to prompt the occurrence of the relevant sensory state.[10]

The second aspect of perception involves the theory through which we exploit this information contained in our sensory states.[11] The determination of what a sensation is a sensation of involves a non-inferential response or a perceptual judgement. Churchland calls this the "subjective intentionality" of a sensation.[12] When we perceive something in the world, what we perceive it as is a function of the subjective intentionality. For example, we see a tree as a tree and not as something else.[13] All perceptual judgements are thus theory-laden. "Any perceptual judgement involves the application of concepts...Any concept is a node in a network of contrasting concepts, and its meaning is fixed by its peculiar place within that network. Any network of concepts is a speculative assumption or theory...Therefore, any perceptual judgement presupposes a theory."[14]

Since they are theory based, these subjective intentionalities depend on the particular conceptual framework we are using in a given perceptual situation. They are thus only as reliable as the theory within which perception occurs. There is nothing in the act of perception itself that guarantees the propriety of the concepts that we apply. Changes in either the sensory apparatus or the theory can thus

result in a change in the perceptual judgements we make about the world.[15]

To sum up, Churchland's view of perception is as follows. Perception is theory laden. It is an active process which must take place within the context of some theory.[16] Our perceptual experience is only partially determined by the external environment interacting with the physiological hardware that we have. In its most full-blown sense, perception is also dependent upon the theory within which it occurs. Perception involves immediate, non-inferential judgements. These judgements are accomplished by means of concepts. These concepts come from our background theory (and both are closely tied to the language that we use). Hence, the conceptual exploitation of the information contained in our sensory states can only be accomplished within the context of some theoretical framework. Thus, perceptual experience cannot be taken to be theory neutral.

- the caloric example

The example that Churchland uses to help clarify the above discussion involves the posited substance caloric. In contrast to the theory of thermodynamics, the caloric theory maintains that there exists a substance within physical objects which is responsible for certain observations concerning their (temperature related) behavior, including their perceived hotness or coldness. Caloric was thought to be contained under different pressures, in physical bodies . The higher the caloric pressure the hotter the object would be perceived as

being. One thing would heat another thing through the exchange of caloric. By placing one's hand in hot water, one could "feel" the presence of a high caloric pressure, relative to one's hand. A proponent of the caloric theory might take himself to be making an immediate non-inferential judgement as part of the perceptual experience; that is, he would "perceive" the water as containing a high caloric pressure. He might even claim to perceive caloric directly. In this way the beliefs about caloric could be viewed as mere generalizations from direct experience.[17]

However, as Churchland points out, one's judgement about the water having a high caloric pressure, and hence the judgement about the existence of caloric, is not substantiated merely by the qualitative character of the perceptual experience. There is something else involved, namely the theoretical interpretation of the objective intentionality of the sensation resulting from placing one's hand in hot water. The intrinsic character of one's sensations is not sufficient to guide one in selecting between the caloric theory and the theory of thermodynamics.

It is worth noting here that there are three different levels at which one's theory may be operative in perception. The first involves those explicitly inferential judgements such as "there exists a substance called caloric". One might make such a statement, and thus be committed to the existence of caloric, for a variety of reasons. One of these could be one's perceptual experiences. The second level of theory involves the immediate, non-inferential judgements which are

part of perceptual experiences (i.e., subjective intentionalities). Here there is more than the mere feeling associated with placing one's hand in hot water. There is also the feeling of the water AS having a high caloric pressure. The third and deepest level involves the qualia themselves. It is possible that even these qualia may be (at least partially) determined by the theory that one holds. This depends on the extent to which perception is active rather than passive.

Hence, if one's theory changed three things might occur. First one's inferences and ontic commitments might change, given identical experiences. Second, one's immediate and non-inferential judgements might change, given the same objective intentionalities. Third, the phenomenal characteristics of one's perceptual experience might change, given the same physical input from the environment.

* * * * *

The Network Theory of Meaning

"Our common-sense terms for mental states are the **theoretical terms** of a theoretical framework (folk psychology) embedded in our common-sense understanding, and the meanings of those terms are fixed in the same way as are the meanings of theoretical terms in general. Specifically, their meaning is fixed by the set of laws/principles/generalizations in which they figure."[18] There are three distinct yet related points made in the preceding passage.

1.) The first has to do with the construal of FP as a theory (thus allowing for the construal of mentalistic terms as theoretical terms). This notion was discussed at length in chapter two. 2.) The second point involves a specific semantic view relative to the source of meaning for theoretical terms generally. Churchland supports the network theory of meaning, according to which theoretical terms acquire their meaning from the role that they play in a theory. "They are implicitly defined by the network of principles that embed them".[19] On this view, the meaning of theoretical terms is a relative and relational affair, each being thoroughly defined only by its relationships to the others. This is very much a holistic conception of meaning where the basic unit of meaning is ultimately the entire language.[20]

3.) The third point has two parts and is implicit in the above passage. (It will be brought out more explicitly below in the discussion of the "infrared people" example.) a.) The first part has to do with the debate over the distinction between observation terms and theoretical terms.[21] One of the corollaries of the network theory of meaning is a denial of the distinction between observation terms and theoretical terms. If all terms are such that their meaning comes from their role in a theory, then even those terms traditionally called observation terms acquire their meaning in just the same way as do theoretical terms. There is thus very little reason to draw a distinction between them.[22]

b.) The second part of this third point has to do with the fact that our common-sense terms for mental states can be roughly divided into two groups. First, there are those mentalistic terms which are used to report our perceptual experiences and their qualitative character (the qualia terms). Examples include the visual experience described as "seeing red" or a sensation described as "feeling pain." Second, there are those terms which are used to report such things as beliefs, fears, hopes, etc. Churchland classifies these latter terms as "propositional attitudes" since they express a certain attitude toward a proposition. For example, if I say that I believe that I am looking at a flower then I am expressing an attitude of assent toward certain propositions, such as "I am looking at a flower."[23]

The claim that propositional attitude terms are theoretical terms may be less problematic than the claim that terms denoting the qualitative character of perceptual experience are theoretical terms. When the functionalist argues that the mentalistic terms of FP are best analyzed along functional lines, the best attempts at refutation turn on the meaning of qualia terms. The charge made against functionalism is that it fails to address the inherent qualia of seeing red and feeling pain. However, for the propositional attitude terms there does not seem to be any specific qualia by which they can be identified. There is certainly no visual, auditory, gustatory or olfactory sensation associated with such things as beliefs. If I say that there is a feeling associated with belief, then either I speak metaphorically or I must admit, upon careful reflection, that what I feel is some

somatic sensation which is contingently related to that belief. I may, for example, believe that I am about to get mugged and have "a knot in the pit of my stomach." But the feeling of the knot is not the feeling of the belief, nor of the correlative fear. It is but a somatic sensation associated with whatever it is that happens in my mid-section when my autonomic nervous system is appropriately stimulated. At any rate, it seems to be less of a problem to maintain that propositional attitude terms are theoretical terms than it does to say that qualia terms are theoretical terms. Hence, in what follows, attention will be focused on the latter claim.

The truly controversial aspect of Churchland's position results from the conjunction of the network theory of meaning with the construal of the qualia terms as theoretical terms. There are two important issues here. The first concerns the debate between the network theory and other semantic views. The second concerns an ambiguity in the use of `observation terms'.

There are two alternate semantic views to be considered. The first, the extensional theory of meaning, makes the claim that at least some of the semantic properties of terms used to report mental states are determined by their relationship to things in the world, i.e., by their extensions. For example, part of the meaning of `water' will be tied to the existence and properties of water.

The second view, the sensation theory of meaning, maintains that at least some of the meaning of qualia terms has to do with the qualitative characteristics of the experience being reported, i.e., the

phenomenal qualia. Hence, the meaning of `pain´ will be tied to feeling pain. At least my understanding of the meaning of `pain´ will depend upon my having felt pain. If this is the correct view, then the claim "I am now in pain" will have an air of indubitability if sincerely uttered when I feel pain.

There is a clear sense in which, from the dualist's point of view, the sensation theory of meaning is just a branch of the extensional theory. Given that one is ontically committed to mental states one might conflate these two semantic views relative to terms denoting mental experiences. (Part of) the meaning of `pain´ would be its extension, which would be the qualitative character of pain. (A physicalist could perhaps also adhere to the sensation theory by identifying the reference of qualia terms with neural states and then accepting the extensional theory.)

The network theory should not be interpreted as allowing for part of the meaning of qualia terms to come from sensory experiences, nor from the extension of the term (if any). The network theory maintains that ALL of the meaning of EVERY term, including one used to report a psychological state, comes from its role in the theoretical framework which embeds it. The only role that Churchland is willing to allow for the actual qualia to play is in our decision to apply a sensation term. The phenomenal qualia may therefore "have an epistemological significance". If there is anything, qua qualia, to which the folk psychologist refers with `pain´, then it serves only to inform that

person about a particular bodily state. Even if qualia do exist, they have no "semantic significance."

One of the problems with the sensation theory of meaning is that the attribution of semantic significance to qualia leads directly to the problem of private languages and "semantic solipsism." If what you mean by `pain' is directly tied to the qualitative character of your pain, then I simply cannot know what you mean by that term. Even if only part of the meaning (for you) comes from your experience, then to that extent I do not know what you mean. This is a direct result of the view that our mental states are inherently private and subjective. The network theory avoids these potential problems if "...the dominant, and perhaps the only, source of meaning for psychological terms is the common-sense theoretical network in which they are embedded."[24]

There is a second important issue involved here which is related to the conflation of the extensional theory and the sensation theory by the dualist. The issue concerns an ambiguity about the denotation of observation terms. What an observation term might be thought to denote depends heavily upon the ontic commitments that one makes. a.) Both the dualist and the physicalist would allow for an observation term to denote something in the physical world that is being observed. For example, if I say that I am now seeing something red, or some red thing, then `red thing' can be an observation term which refers to something physical and "out there." b.) One might also use observation terms to report on something going on inside oneself as an observer. One might be reporting the subjective (or perhaps the

objective) intentionality of a perceptual experience. However, at this point there will be a disagreement between the dualist and the physicalist concerning this (introspective) use of qualia terms. Here there is a certain ambiguity concerning the reference of an observation term. The mentalist might interpret the referent to be something mental (in the strong ontic commitment sense), while the physicalist will interpret it as being something physical occurring in one's sensory system. This turns out to be important when observation terms are given the status of theoretical terms. This topic will be brought up again in the next chapter. The reader is urged to keep the above ambiguity in mind while reading the section below on the infrared people.

Thus, according to the network theory of meaning, the terms used in FP to describe mental states are theoretical terms. Theoretical terms are meaningful due to the role they play within a theory. They are implicitly defined by the way that they figure into the system of theoretical principles which embed them. (Hence, once again one can see that the language that one uses is intimately tied to one's background theory.) Concepts are acquired through the learning of a language. We learn the correct use of conceptual (theoretical) terms while learning the language of which they are a part. The terms used to express theoretical concepts are meaningful only within the context of that theory. Churchland's theory of perception is thus closely related to his (network) theory of meaning, since these concepts are then involved in our immediate, non-inferential perceptual judgements.

Hence, both the concepts used in the exploitation of sensory information, and the meanings of the terms used to express those concepts, come from one's theory.

- the infrared example

In an attempt to illustrate the network theory of meaning (and to defend it against the sensation theory) Churchland proposes the following thought experiment. We are asked to imagine a race of creatures much like ourselves except that their visual systems are keyed to temperatures rather than colors. They have retinas whose photoreceptors are sensitive to electromagnetic radiation in the far infrared. To them the world looks something like a black and white photograph (taken with infrared-sensitive film) would look to us. They "see temperatures in black and white." Hot (cold) things look to them the way white (black) things look to us. The primary linguistic difference is that they lack our color vocabulary. Since their visual systems do not have the appropriate range of response, they have no achromatic color words analogous to our `white', `gray' or `black', nor any chromatic color words for that matter.

Problems arise when we attempt to translate their language into ours. We can agree that fires are hot and ice cubes are cold. We feel it and they see it. Now, if the meaning of our observation terms, such as `white', comes from the qualitative character of our perceptual experience, then we must maintain that their terms, `hot', `warm' and `cold' must really mean `white', `gray' and `black' respectively.

After all, if the meaning of `white´ comes from our phenomenal experience of whiteness, and their visual sensation in the presence of an object vigorously radiating in the infrared range is the same sort of experience phenomenally, then their `hot´ means our `white´.

Churchland is quick to point out the difficulties with this "heterophonic sensation-guided translation." If their `hot´ really means the same thing as our `white´, as the sensation theory of meaning seems to indicate, then when they say "fires are hot" and "food keeps better in a cold place", we must interpret them as claiming that fires are white and food keeps better in a black place. This is clearly an unacceptable translation. From their point of view, when we say "snow is white" we are really saying that snow is hot, which any of them can plainly see is false. Clearly this approach to translation is inadequate. The lesson we are supposed to draw from this is that the meaning of observation terms has nothing to do with the phenomenal qualia of perceptual experience.

Although Churchland moves from this point directly to the claim that the network theory must be correct, it remains a possibility that the observation term `hot´ has the meaning it does because of its association with those physical objects radiating in the infrared due to a high mean molecular kinetic energy. In particular, if the extensional theory is correct, then observation terms might have an epistemic status different from that of theoretical terms even though meanings are not determined by sensations. Assuming our sensory hardware remains constant we can have a theory neutral source of

information about the world. We can observationally establish the correct application of an observation term and hence, in a derivative, theory-neutral way, pin down its meaning.[25]

Churchland embellishes the story of the infrared people a bit by adding the assumption that they also have tactile sensations of hot and cold. He uses the case of bimodal sensory experience as a further illustration of the claim that observation terms cannot get their meaning from sensation. Were the phenomenal qualia of sensory experience the source of meaning then, not only would the infrared people and we mean different things by `hot', but for them `hot' as seen would mean something different than `hot' as felt. Similarly, for us, `round' as seen would mean something different than `round' as felt.[26]

If we are to allow that an observation term, such as `round', refers to an objective property of the world, it must be (in principle) possible for that property to be detected by different sensory systems, either the same type of sensory system in different individuals or different systems in one individual. By whatever sense modality information of an objective property enters (objective intentionality), an inference is involved (subjective intentionality) to get from the sensory information to the objective property. If it is possible for other beings to share a common observation vocabulary with us, if it is possible to have the same observation term apply in cases of bimodal perception, then it must be the case that the meaning of an observation term is not given in sensory experience. Churchland goes so far as to

say that the notion of a phenomenal meaning might never have been conceived if all observational properties were at least bimodal for us.[27]

In a line of argumentation related to the contrast between epistemic and semantic significance, Churchland draws a distinction between understanding observation terms and the ability to apply these terms non-inferentially in cases of perception. It is possible for someone to know the meaning of an observation term even though they lack the appropriate sensory functions. A color blind person can use `red' correctly without the associated phenomenal experience of vision. They lack the sensory capacity to be able to apply the term `red' non-inferentially in response to a sensation. Conversely, the use of an observation term in connection with a particular kind of sensation does not provide that term with semantic identity. Thus, a color blind person who is familiar with the theory knows the meaning of `red' while a child can respond in a stimulus-response fashion to red things, but not understand what `red' means.

* * * * *

Implications

Several things follow from the joint theses of the theory-ladenness of perceptual experience and the network theory of meaning. These claims will be laid out here. However, it will be most expedient

to postpone their evaluation until later, after EM and FP have been construed as paradigms.

Perceptual judgements: Even if we do have sensory experiences with a particular intrinsic identity, the perceptual judgements that we make are based on our theories. Due to the role of conceptual exploitation in perception, i.e., the subjective intentionalities, "the intrinsic qualitative identity of one's sensations is irrelevant to what properties one can or does perceive the world as displaying. The meaning of a term (or the identity of a concept) is not determined by the intrinsic quality of whatever sensation happens to prompt its observational use, but by the network of assumptions/beliefs/principles in which it figures."[28] Depending on the theory that we hold and the sensory hardware that we have, any sort of experience could be conceptually tied to any sort of objective property in the world. Hence, our sensory experiences themselves cannot be thought to supply an indubitable basis for ontic claims. Such claims are judgements made within a theory and are necessarily tied to that theory. This point can be made, in the language presented in chapter two, by saying that our weak ontic commitments arise from the theory that we use, along with its indefinite singular terms.

Consider, as an example, how we apply observation terms such as `hot' or `white' as dictated by our theory. The use of such terms, even in response to what the theory describes as the experience of hot or white, does not ensure the existence of such things as heat, hot

things, whiteness or white things. As an illustrative example, recall the case of caloric. The validity of our perceptual judgements will be based on the theory within which they are made.[29] There is nothing guaranteed about the outcome of our conceptual exploitation of our hardware-generated sensory information. A perceptual judgement, as expressed in a statement using observation terms, derives no indubitability from the sensory experience which happened to prompt it. (It is even true, in a derivative way, that the theory itself does not attain indubitability from the sensory experiences. Rather, those experiences are what they are only as judged within the theory.)

Introspective judgements: Theory neutrality is not attained by shifting our focus to introspection. One might, for example, maintain that things in the world are not really hot, that `hot' refers to a secondary quality, that being hot is a quality of an object-as-perceived rather than an intrinsic characteristic of the thing independent of its being perceived. However, the use of `hot' as descriptive of the intrinsic character of the sensory experience does not carry any special epistemic clout. Introspective judgements involve theory just as do any other kind of perceptual judgement. Hence they are prone to the same pitfalls. "The propriety of our introspective judgements remains contingent on the adequacy of the general conception of those inner states that those judgements presuppose".[30]

Since all judgements are theory laden, if one were to change one's theory, then even one's introspective judgements might change. They

could even change radically. This might well lead to a change in one's ontic commitments concerning the nature of that upon which we introspect. Specifically, the experiences that we have and the judgements that we make within FP are no guarantee of the existence of such things as beliefs, desires, or sensations of red. Given the network theory, the qualia associated with sensations of red have no semantic significance for such folk psychological descriptions as "sensation of red." Hence, the perceptual experiences that we do have, whatever they are, constitute no obstacle to the abandonment of FP.

The analogy that Churchland uses in his discussion of introspective judgements is that of a measuring instrument. Consider an instrument which has a dial to indicate something about an objective feature of the world. There is no extra reliability gained when the instrument "introspectively" measures current flowing in the circuit which causes the dial to point where it does. (In fact, there is a greater possibility of error, since there is an extra step involved in this "introspective" measurement.) In both cases an interpretation function is required and this function is derived from the operative theory. The information output from the system is no more theory neutral in the "introspective" case.[31]

One of the consequences of all this is a prescription to drop any claims of indubitability concerning the existence or nature of our internal, mental states, e.g., seeing red, here, now. If we simply drop (rather than reduce) the common sense conceptual framework, we would be free to adopt the more powerful NS. All of the phenomena

which we currently interpret from within FP as mental experience might then be interpreted physicalistically, e.g., as C-fibers firing rather than pain.

The elimination of experience: Since our judgements concerning the intrinsic qualitative identity of our experiences depend on our theories, the intrinsic quality of those experiences (as judged by us) might change given a change of theory. In fact, since the actual qualitative character of sensory experience is incidental to the act of perceiving the world, not only might one type of qualia do as well as another, but they may well be unnecessary as long as the relevant causal connections remained intact. These sensory experiences, as "causal middle-men," might simply be eliminated. One could still learn and theorize about the world, make the same judgements concerning the nature of the properties "out there."[32] (This illustrates one of the important differences between the qualitative character of perceptual experience and the judgements we make about the world.)

Theory selection: Since perception must be laden with theory, we should use the best theory that we have available, which at present is a scientific one. Since folk psychology is a theory and since our sensory experiences per se do not support any particular theory, the way is open for a total elimination of FP and the rejection of all its ontic commitments. The decision as to which theory to adopt must be governed by the relative merits of the contending theories relative to the evaluative criteria laid out in chapter 2. The ontic commitments that we make will follow from the theory which we accept.

* * * * *

Eliminative Materialism and the Identity Theory

This section will examine the ontological implications of EM. Given that EM rejects a type/type identity theory, there are several possible alternate views to be considered. The first allows for the integration of EM and a token/token identity theory. The second view accepts a very general and ill-defined identification of the extensions of mental and neural terms, although it rejects token/token identities. On this latter view, the actual referents of FP terms are a nonsensical potpourri of sub-tokens of NS tokens.

A third possible view denies any legitimate form of mental/physical identification. This might involve a positive claim that mentalistic terms have no reference, much as `unicorn´ has no reference. It might also involve the acceptance of only such hypothetical formulations as "IF the (mentalistic) terms of folk psychology (FP) do refer to anything, THEN what they must in fact refer to is something physical. There is an element of ontic neutrality in this third view, which comes about as a result of increased emphasis on the notion of a theoretical description. This claim is also tied most closely to the network theory of meaning, since the failure of FP terms to refer is due to their meaninglessness, which results from the abandonment of FP. The claim here is that, insofar as FP is actually describing anything, it is (poorly) describing something that is better

described within the context of NS. In any case, under any of the above views, EM allows for only physical tokens and/or types.

1.) Eliminative materialism and the token/token identity theory: Let us first consider the conjoining of EM and token/token identity theory. There are some advantages to such a view. Even though there will be no type/type identities to be found between FP and NS, still there can be a significant relationship between them. The types of FP may make no sense from the point of view of NS, but when someone who is working within FP describes himself as being in a state of pain, he can be viewed as at least making a sensible claim. There really is something going on that he is describing. It is simply that NS can better describe it as, for example, a token of a particular neural state. This would also allow FP to provide a partial starting point for neuroscientific research. If one accepts the token/token identity theory, then one would seem to respect at least some of the traditional wisdom engendered by FP. The conjunction of EM and the token/token identity theory thus makes some intuitive sense, for whatever that is worth. It will also give a certain credence to such philosophical approaches as sense-data theory and phenomenology, though in a very restricted way to be sure. Claiming that such things as "having the visual experience of red-here-now" provides an epistemic foundation would have some merit. One could at least admit that those who proposed such views were addressing something that really existed.

There are several places in the literature espousing EM where a token/token identity seems to be implied. Consider the following:

"Sensations are just **causal** middle-men in the process of perception, and one kind will serve as well as another so long as it enjoys the right causal connections. (So far then, in principle they might even be **dispensed** with, so far as the business of learning and theorizing about the world is concerned...)"[33]

This passage occurs in a context of arguing for the possibility of the elimination of experience. Here, `sensations' is being used in the folk psychological sense of a mental event. If what is meant is a physiological sensation then the last sentence would not make much sense, as we could hardly be thought to be able to learn about the world without our sensory systems. Given this mentalistic interpretation of `sensation', a sensation could be a causal middle-man for a materialist only if it really existed, and it could be said to exist only if it were actually a physical event in the causal chain involved in perception (physicalistically conceived). Thus, the most natural interpretation of the above passage seems to be that, what FP would describe as an occurrence or token of a mental sensation of red really is something, namely a token brain event. Thus the implied token/token identity. (There is, though, a certain carelessness evident here, since sensations qua brain events would not likely be eliminable without loss to the organism.)

A second indication of the acceptance of a token/token identity of mental events and brain events can be found in Churchland's treatment of the problem of self-consciousness.[34] The view expressed here is that self-consciousness consists in the apprehension of "mental states" within a theoretical framework such that discriminatory judgements might be made about them.

"Self-consciousness, it seems, is a kind of continuous apprehension of an inner reality, the reality of one's mental states and activities." "...one's introspective consciousness of oneself appears very similar to one's perceptual consciousness of the external world. The difference is that, in the former case, whatever mechanisms of discrimination are at work are keyed to internal circumstances instead of to external ones." "Our faculty of judgement is in...systematic causal contact with the rest of the internal domain of which it is a part. Who will express surprise that one kind of brain activity enjoys rich causal connections with other kinds of brain activity?"[35]

The above passages should not be interpreted as indicating a positive strong ontic commitment to mental states. However, there is a pretty clear weak ontic commitment to the existence of mental states as type-of-things, qua brain states. What FP would call an occurrence of "introspective self-consciousness", NS would perhaps describe as a token brain state which involves the assessment of other tokens of brain activity. If I can "introspect" on a token pain$_r$ then that token pain had better be a token brain state.

- problems

a.) Despite these passages and the earlier mentioned advantages, there are several problems associated with conjoining EM and a token/token identity theory. The first difficulty involves the fact that accepting token/token identifications would be inconsistent with at least some of the reasons presented in favor of accepting EM. It was argued that the search for type/type reductions placed certain unnecessary top-down constraints upon the neuroscientist.[36] This line of reasoning can be extended to argue against accepting even a token/token identification. Consider, for example, the case of a pain.

If the search for a neurophysiological type to which the FP type `pain´ can be reduced presents an unnecessary top-down constraint upon the neuroscientist, then the search for a specific neural event to be identified with a specific instance of pain will also be constraining. `Pain´ does not represent a legitimate neurophysiological concept. For the eliminative materialist there is no reason to suppose that the neuroscientist, or anyone else, should be concerned about finding either types or tokens of neural events which are pains. Trying to identify a particular pain in the nervous system is really not much better than trying to identify a type of neural event with pains generally.

Note that this is contrary to the advantage of the token/token identity theory stated earlier, namely that it provides a starting point for NS. Thus, whether the acceptance of token/token identities is something positive or something negative depends on one´s point of view about the advantages of a top-down approach. Churchland seems ambivalent about this. Consider the following:

> "The `facts´, as currently conceived and observed by us, form the starting place for theoretical inquiry, but its successful pursuit may well reveal that we should vacate that starting place as hastily as possible. Large-scale intellectual progress will involve the wholesale rejection of old explananda as frequently as it involves the wholesale introduction of new explanantia."[37]

b.) A second, perhaps deeper problem with the integration of EM and the token/token identity theory turns on the realization that the characterization of a token as a token requires the (implicit)

recognition of the relevant type. For example, in order to specify a token pain one must somehow recognize the type pain, else there would be nothing to guide one in picking out just this feature of the world as a particular token. `Token' should be viewed as a two-place predicate. One would say that "x is a token of y", where `y' refers to some type.

One of the key features of EM is its rejection of the types of FP. Under EM, there are no legitimate grounds for accepting the type pain. If one were to accept such a type then one would have to be more sympathetic towards reductivism. All of the versions of reductivism do recognize the types of FP and hence can talk about the relationship of the types and the tokens of FP to those of NS. According to EM, FP is a theory which consists of vastly inferior conceptual categories (types). If the types which are countenanced by FP are so radically misconstrued, then the tokens of those types should be as well. Hence, EM should reject the claim that each token of the type pain just is a token brain event since `token of the type pain' will not be well defined from the EM point of view.

Thus, for EM the tokens of FP should be just as unacceptable as its types. When one is contrasting the psychoneural and psychofunctional versions of reductivism/identity theory, the distinction between types and tokens is useful since it brings out a major difference between these two views. However, both of these positions recognize and accept the types, the conceptual categories and the weak ontic commitments, of FP. Thus, for them, the notion of a

token of the type pain does make sense. EM, on the other hand, totally rejects the FP types, and thus should reject its tokens. It presents NS as a theory which is not (reductively) related to FP.

c.) A related problem with the conjunction of EM and the token/token identity theory turns on the fact that `pain´, as used within a dualistic FP, purportedly denotes a mental experience. Each pain is a token of the type pain, but since pain is one sub-type of a larger type, namely the ultimate type "mental events," each token pain is also a token mental event. A token pain is thus not only an instance of (and identified by membership in the class of) the type pain, but it is also an instance of the type mental event. Thus, `a token pain´, as used within FP implies not only the type pain (qua weak ontic commitment) but also the type mental event (qua strong ontic commitment). Hence, the recognition of a token pain may well allow for mentalism to slip in the back door, so to speak. This is clearly something which EM rejects.

Given these problems, it would seem that the best, and perhaps the only coherent, route for EM to take is to reject a token/token identification altogether. There are two alternative approaches one might take here, one more extreme than the other. The less radical view accepts a non-token/token identity and maintains that although tokens of the indefinite singular terms of FP do not legitimately refer to any physical tokens, as specified under NS, they do refer to something, and that something is indeed physical. The more extreme view maintains that mentalistic words fail to refer altogether. There

is simply nothing that `pain' actually picks out, there is just no reference for FP terms. Each of these views will be examined in turn.

2.) Eliminative materialism and a non-token/token identity: The problem with FP tokens, on the less radical of the two remaining construals of EM, is analogous to the problem with FP types from the point of view of the functionalist. Here, the failure of FP is that, even at the level of tokens it does not match up referentially with NS. Perhaps the terms of FP refer to part of one physical token and part of another one. Hence, although there is some sort of a reference, it is so confused, from the point of view of NS, that it is to be discarded.

As an illustration of this situation, consider the following:

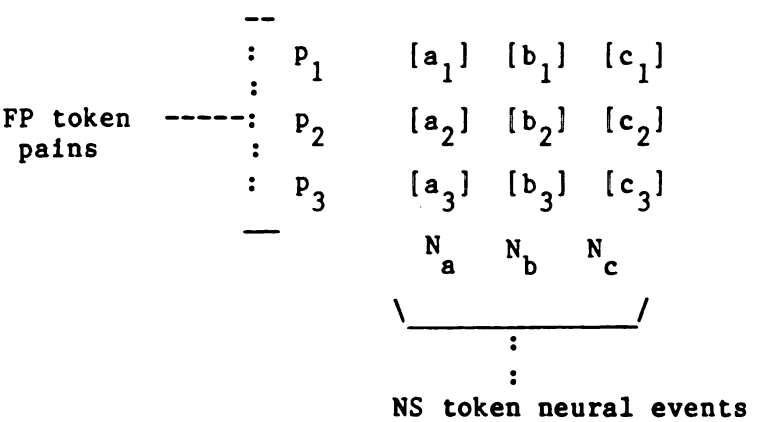

Figure 4.2 : Sub-Token Identities

In the above figure, each p_i (from FP) consists of $\{a_i, b_i, c_i\}$ and purportedly denotes a token pain. Each N_j (from NS) denotes a token neural event, consisting of $\{j_1, j_2, j_3\}$. Each j_i would be

viewed by the materialist as the component parts (sub-tokens) of the tokens specified by N_j. The important point is that everything is physical.

Consider, what FP would call a token pain. Under EM, this token-use of `pain' might refer to something that is actually occurring in the nervous system. But the reference of that term is quite confused from the point of view of NS. The actual reference of p_1 will be the neural happenings denoted by, for example, a_1, b_1, and c_1. But $\{a_1, b_1, c_1\}$ constitutes neither a type nor a token neural event within NS. On this view, FP and NS do not match up even at the level of tokens. However, even without a token/token identity theory, there are physical extensions for FP terms, confused as they may be.

The splitting of tokens into sub-tokens is not really as objectionable as it might seem at first glance. There is, after all, a hierarchy of types and a related hierarchy of tokens. The firing of a c-fiber involves many things, each of which is a token at a certain level of description. There will be a graded potential in the cell body reaching the threshold to initiate an action potential down the axon. There will be various ion transfers and ion gate openings and closings, etc. Each of these neural happenings is a token of a type of event. NS consists of a set of conceptual categories which constitutes an hierarchical ordering of these types of events and each type has its tokens.

One of the claims of this version of EM would be that the purported referents of tokens of the indefinite singular terms of FP

end up being a nonsensical potpourri of neural events and/or of events which make up neural events. It must be remembered that NS is a theory which offers a description of the world only at a certain (broad) level. Sooner or later one drops out of NS proper and begins giving a description of brain events in terms of some other theory, say biochemistry or biophysics. The bottom line is that, although a token of `pain´ may refer to some set of physical events whenever it is used correctly within FP, it may be the case that what it refers to is an unrelated juxtaposition of physical events that do not constitute a type or a token from the point of view of NS. In fact, it may be equally bad from the point of view of any physical theory at any level, be it NS, biochemistry or whatever.

While reading the arguments for EM, there are times when it seems as though this non-token identity is being posited. One example is to be found in Churchland´s discussion of the "epistemic significance" of psychological (i.e., phenomenal) qualia. After presenting, and arguing for, the network theory of meaning, he considers an objection to the view that none of the meaning of psychological terms comes from the qualitative character of the mental states to which they (are meant to) refer. Churchland considers two possible responses. First it may be admitted that some of the meaning of, e.g., `pain´ comes from the hurtfulness of being in pain. In this case however, there still will be a private language problem. Since part of what I mean by `pain´ is something that you can never know, we speak (partially) different languages. The second response is the more interesting in the present

context. Here qualia are given a significant role in the greater scheme of things, but they are given only an epistemological role, not a semantic role.

> "The second compromise concedes to qualia a significant role in the introspective **application** of sensation-terms, but still attempts to deny that their role enjoys any **semantic** significance...Qualia, therefore, have an epistemological significance, but they are without semantic significance for terms in an intersubjective language."[38]

Thus, the qualitative character of a pain may allow one to make a "spontaneous observation judgement" about what kind of a state one is in. The hurtfulness of pain may allow us to noninferentially apply the term `pain' to our current situation.

This concession seems clearly to imply some form of identity (perhaps even a token/token form). Allowing for the epistemological significance of the qualitative character of mental experiences is open only to dualists and materialists who maintain that mental terms refer to something physical. If there were absolutely nothing that `pain' referred to, nothing actually denoted by that word in any given instance, then it is difficult to see how this talk of the epistemological significance of the hurtfulness of pains could make any sense. If there is not at least some mental/physical identity between a particular pain (à la a folk psychological description of the situation) and a particular physical state, then what sense can be made of talk of epistemological significance of the qualia of perceptual experience? If `qualia' is a locution of a defunct theory and does not

refer to any physical feature at all of one's brain, then how can it (the qualia) have epistemological significance? This kind of a move simply is not open to an eliminative materialist whose position is so pure as completely to disallow any identification of the extensions of tokens of mental state terms and those of brain state terms. (It should be noted that the epistemological significance of something like a pain seems to make the most sense under a token/token identity theory.)

3.) Eliminative materialism and the total rejection of identity: One might formulate EM as a position which rejects even a general, ill-defined form of identity just described. This most radical version maintains the doctrine that FP terms, both types and tokens, have no reference or extension AT ALL. On this view, `pain' would be very much like `Zeus', in that there is absolutely nothing to which it refers. It would not refer even to the extent to which `a thunder bolt of Zeus' might be thought to refer, confusedly and under a wrong description, to lightning or a sudden discharge of atmospheric electric potential. What is proposed here is not merely confused reference, but NO reference.

This construal of EM is the most extreme and the most problematic. On this view of EM, the excruciating pain of a dualist who is being boiled alive is no more real than is Zeus. Even if you are now submerged in boiling water and screaming about being in terrible pain, there really is nothing there that you are actually picking out with your term `pain'. What you call pain is not even identical with

anything physical. Now that is an extreme view. If the "absurdity" objection of the proponent of FP holds any force, this is where it is most weighty.

There are passages where such a view is at least hinted at. During a discussion of the ontological implications of intertheoretic reduction, Churchland says that cross-theoretical identity claims are not a proper part of an intertheoretic reduction. Correspondence rules need not be construed as identity claims. They may be mere ordered pairs, in which case we will

> "...need only the minimal assumption that the second element of each pair truly applies where and whenever the first element of each is normally **thought** to apply. Such an assumption, note, is strictly consistent with the idea that the first elements (the expressions of [the old theory]) do not apply to reality at all."[39]

While responding to some objections to EM, Churchland considers the claim that "the bald statement of eliminative materialism is that the familiar mental states do not really exist."[40] He goes on to reject the claim that such a statement must be meaningless since it denies the existence of beliefs, one of which is embedded in the above statement. However, what is instructive here is that he does not dispute the content of the sentence, i.e., that EM does claim that mental states do not exist. The apparent implication of these passages is that the mentalistic ontology of FP is a myth. There just is no such thing as a mental event. There is, therefore, no type/type, token/token nor any other mental/physical identification to be made.

- problems

The denial of any form of mental/physical identity poses certain problems. These problems concern both the explanatory uses of FP terms and the epistemic status of the proposed referents of those terms. We have a very strong intuition that, in our own case anyway, what we call qualia do exist. Even if folk psychology is a radically misleading theory, even if it is so wrong that it cannot be reduced to NS, it is still the case that the claim that sincerely uttered tokens of `this pain is awful' do not refer to anything is too strong a claim. There may not be a specific token brain event to be identified with the referent of `this pain'. But it seems that tokens of `this pain is awful' must refer, however badly, to something. We know when to apply the term. There are clear-cut cases of pains and of non-pains, at least when we are operating within FP.

Further, even the most adamant eliminative materialist must admit that folk psychology has been a tolerably successful theory over the centuries. The human race has survived, and even prospered through its use. When I am in a state which I describe (within FP) as being in pain, I will generally engage in certain kinds of behavior. I will grimace. I might say that I am in pain. I will usually take whatever appropriate measures I can to remove the cause of that pain and in the future I will avoid those situations which have previously caused me pain, thus facilitating my survival. If there is absolutely nothing to which I refer when I say that I am experiencing pain, if there is absolutely nothing there which `pain' is getting at, then, not only is

it is difficult to see how so-called mental states could be epistemologically significant but the apparent explanatory power of FP will be quite a mystery.

Since FP does work to some extent we must admit that I can know something about myself based on what FP labels psychological or mental states. What I know based upon my mental states can affect my behavior. Since, according to the materialist's view, all behavior has a physical cause and only a physical cause, the so-called mental state must be something physical. Thus, to take this most extreme eliminativist's view is unacceptable.

4.) Eliminative materialism as a paradigm shift: Despite the suggestiveness of such passages as the one above, it is probably not fair to saddle the proponents of EM with quite the view just described. The claim that FP terms fail to refer must be understood within the context in which it is presented. There are several key points argued for by the eliminativist. The first is that NS is, or eventually is going to be, a superior theory to FP. The second is that elimination is preferable to reduction as a method of transition from FP to NS. The third is that all of the meaning of theoretical terms comes from their role a theoretical framework. The fourth is that all terms are theoretical terms. Hence, the only distinction to be made between observation terms and theoretical terms is one of degree. Finally there is the claim that the theory one accepts implies specific (weak) ontic commitments.

Hence, given the relative merits of NS over FP, we should adopt NS. Given the merits of elimination over reduction, we should simply abandon FP. Now, the claim to be considered is that, from the point of view of EM, even token uses of `pain' fail to refer for the following reasons. First of all, once we move from FP to NS, `pain' will be essentially meaningless since it is not a term that is defined within NS. Second, since `pain' is as theoretical as any other term, there is no special epistemic status to be given to reports of being in pain. Finally, since ontology is implied by theory and FP is to be dropped, the weak ontic commitments of FP will be dropped as well, at least to the extent that they are not also implied by NS. Given all of this, there may be a way to accept much of the essence of the extreme form of EM presented above, while avoiding most of the charge of absurdity.

Consider the case of caloric, and the difference between the terms `caloric' and `unicorn'. `Caloric' was meant to refer to a substance. From the point of view of contemporary physical theory, there is no substance that the term referred to. But there was something going on, something that was really there. What was really there was (what is now described as) the kinetic energy of molecules. According to our best current theories, that is the actual phenomena which the caloric theory was in fact addressing, insofar as it was addressing anything.

`Unicorn' is another term that does not refer to anything. There are no unicorns. But unlike `caloric', there is nothing at all that `unicorn' was getting at. There were (presumably) no observations that were made which the existence of unicorns could serve to explain.

Except for the fanciful writing of fiction and fairy tales, `unicorn´ simply does not refer to anything, not to a substance, to an event, or to anything else.

> "...since we were unable to **identify** caloric fluid with kinetic energy (according to the old theory, caloric is a material **substance**; according to the new theory, kinetic energy is a form of **motion**), it was finally agreed that there is **no such thing** as caloric. Caloric was simply eliminated from our accepted ontology."[41]

According to this moderately radical version of EM, `pain´ is more like `caloric´ than `unicorn´. The claim is that `pain´ does not refer to a type, nor to a token, nor even to any set of sub-tokens of the physical brain and its activity. Further, there is nothing in the extension of `pain´ to be identified with anything in the extensions of any NS terms. Still, it is not the case that there is nothing which the folk psychologist is trying to describe with the word `pain´. Rather, it is the case that "what" some people are trying to describe as `pain´ is better described under NS in a much different way. Hence, there is a clear sense in which there might not be pains even though there is something very real going on whenever someone sincerely reports that they are in pain.

Another example of this view can be found in a passage from Rorty:

> "It would seem that the verb in such statements as `Zeus's thunderbolts are discharges of static electricity´ and `Demoniacal possession is a form of hallucinatory psychosis´ is the `is´ of identity, yet it can hardly express **strict** identity. The disappearance form of the Identity Theory suggests that we view such statements as elliptical for e.g. `What people used to call "demoniacal possession" is a form of hallucinatory psychosis...[Similarly,] what people now call "sensations" are identical with certain brain processes."[42]

One must be very careful about how this "non-strict" identity is to be interpreted. It is certainly not like the relationship between temperature and mean molecular kinetic energy. Movement from a discussion of the thunderbolts of Zeus to talk of discharges of atmospheric electricity does not constitute a shift to a "deeper" or different level of description. What is being proposed here is not just a version of the type/type identity theory, nor of the token/token theory for that matter. Although there was indeed something going on in the world which our distant ancestors described by `thunderbolt of Zeus', there is no hope of reducing such a notion to, or translating such a term as, `sudden discharge of atmospheric electricity'. The reason lies in the network theory of meaning and the connectedness of the conceptual categories of the theories involved. If lightning really were identical to the thunderbolts of Zeus, then lightning would be the type of thing that is "thrown down from Mt. Olympus by a deity."

Perhaps we could convince an ancient believer that the bright flashes of light we see in the sky during a storm is electrical discharge and not a manifestation of an irritable god. Now, that may not dissuade the person from his belief in the existence of Zeus. He might only grant that lightning is not a thunderbolt of Zeus. Perhaps he will even admit that there are no such thunderbolts. But he would never be convinced that there was ABSOLUTELY NOTHING at which he was getting when he sincerely said "there is one of Zeus' thunderbolts." He was simply misdescribing lightning. Hence the "non-strict" identity. Now, contrast that with the case of the person who gives up

his belief in the existence of Zeus. Here he would say that there was nothing at all at which he was getting when he used `Zeus´ in a purportedly referential way. Here there is a total failure of reference. Zeus is not even "non-strictly" identical with anything countenanced by our contemporary theories.

The claim made by Rorty is that `sensation´ misdescribes brain processes much as `thunderbolt of Zeus´ misdescribed lightning. The wrong description results from an incorrect judgement being made. In fact, as Rorty makes out the case, there is little if any difference between misnaming and misjudging.[43] Such is not surprising since both must occur within the context of a theory (hence the link between the network theory of meaning and the theory-ladeness of all perceptual judgements). Further, we must be careful to distinguish our descriptions from what is being described. If one accepts one's theory as true, then it is easy to conflate these two levels. This is evident in the passage by Rorty above. To be precise, that passage should read: "What people used to call "demoniacal possession" is now called a form of hallucinatory psychosis...What (some) people now call "sensations" are, under a different theory, called "brain processes."

Given the fact that our judgements about the world must be formulated within the context of a theory, any of them may be wrong if the theory within which they are made is wrong. This includes perceptual or observation judgements. Those judgements are supposed to concern what exists. If we drop FP and accept NS, then such judgements will change and it will no longer make any sense to claim that pains

exist. Such claims will literally be meaningless. The shift from FP to a physicalistic theory may well involve "an ontological displacement of rather jarring proportions."[44] That all of this is being claimed is evidenced by the following:

"The eliminative materialist holds that [folk psychology]...is a **false** theory. Accordingly, when we finally manage to construct an adequate theory of our neurophysiological activity, that theory will simply displace its primitive precursor. [Folk psychology] will be eliminated, as false theories are, and the familiar ontology of common-sense mental states will go the way of the Stoic pneumata, the alchemical essences, phlogiston, caloric and the luminiferous aether."[45]

"[Folk psychology]...is no better off for being our current matrix of perceptual judgement...The conviction that the world instantiates our ordinary observation predicates cannot be defended by a simple appeal to the `manifest deliverance of sense'. **Whether or not the world instantiates them is in the first instance a question of whether the theory which embeds them is true**, and this question in turn is primarily a matter of the relative power and adequacy of the theory as a means of rendering the world intelligible."[46]

* * * * *

Conclusion

One thing should be evident at this point. There is more involved in EM, especially the more radical versions, than merely an analysis of the relative merits of FP and NS. The network theory of meaning is not part of NS. Nor is the related claim that there is no distinction between observation terms and theoretical terms. The analysis of the advantages of elimination over reduction has nothing to do with NS. And even though it is often implied that NS involves the strong ontic commitments of a physicalistic monism, there is little to be found in NS proper to support such a claim. You will not, for example, find much space in a neurophysiology text devoted to the topic of the ultimate ontological character of neural tissue.

The philosophical position termed "eliminative materialism" is something more than a theory. It is better thought of as a world view, gestalt or paradigm. In fact, as FP has been construed in this work it too is more than a mere theory. The next thing to be done is to lay out FP and EM as competing paradigms. Such a construal will go a long way towards explaining the apparently irresolvable conflicts that have raged between proponents of these two views.

Chapter 5

Dualism versus Materialism:

Competing Paradigms

Introduction

There is an important distinction to be drawn between theories and paradigms. The arguments presented in favor of a completed neuroscience (NS) over folk psychology (FP), as theories meant to explain and predict human behavior, are not necessarily operative as arguments in favor of materialism over dualism, as encompassing world views which include strong ontic commitments. The most natural (and perhaps the only) way to evaluate the arguments for NS or FP, as theories, is from within some larger conceptual framework or other. One might thus use the hypothetico-deductive method to justify the weak ontic commitments of FP relative to those of NS from within the larger framework of dualism as a paradigm. Similarly one might justify NS over FP from within a materialistic paradigm. The method of justifying these paradigms will, however, be a very different matter.

Until now, FP has been presented as a common-sense THEORY which functions, in part anyway, to explain and predict human behavior. However, as has been stressed, it is a view which is intimately related to dualism. When FP is conjoined with dualism it forms the conceptual framework into which most of us are born and which we learn to use by default. As a theory, FP does not NECESSARILY imply the strong ontic commitments of the dualist, although it does imply certain weak ontic

commitments, e.g., to pains, beliefs and desires. However, it is quite reasonable to view FP as more than a mere theory. It is a world view which is a de facto dualism. Hence, in what follows, `dualistic FP' will be used when it is the paradigm that is to be emphasized and `FP' will be used, as it has been throughout this work, to refer to our common-sense belief-desire conceptual framework as a theory which we use to explain and predict human behavior.

NS proper does not NECESSARILY imply the strong ontic commitments of the physicalist, although it does imply weak ontic commitments to such things as axons and synapses. The mere conjoining of NS and physicalistic monism underdetermines materialism as a paradigm. It could be either reductivistic or eliminativistic. If one adds the network theory of meaning and a generous dose of skepticism concerning the merits of FP and/or the likelihood of intertheoretic reduction, the result will be a world view or paradigm which is eliminative materialism (EM). If one is less convinced that FP is unworthy of continued support, one might adopt the view which Churchland terms "revisionary materialism", a view which allows for partial reduction and partial elimination of the indefinite singular terms, the conceptual categories and the weak ontic commitments of FP.

The things left now to do are: 1.) present sketches of dualistic FP and of EM as paradigms, and of FP and NS as theories of human behavior, and to summarize their strengths and weaknesses. It will be useful to describe the paradigms of dualistic FP and EM as completely as possible in order to make very clear their many points of

disagreement; 2.) show how construing dualistic FP and EM as paradigms, in contrast to FP and NS as theories, brings out the source of the apparent irresolvability of the various disputes between proponents of these two conceptual frameworks; 3.) explicate the difficulty in developing criteria for choosing between paradigms; and 4.) take a last look at dualistic FP and EM.

* * * * *

The Competing Paradigms

- folk psychology

Following is a list of the many constituent elements of FP as a theory and of the dualistic FP as a paradigm. There is no intent here to argue that this list is exhaustive, nor that each element is a necessary part of FP, especially a dualistic FP, as accepted by everyone. It is very doubtful that there is just one well-defined conceptual framework that can characterize common-sense. Rather, this list should be viewed as a fairly complete, consistent and reasonable approximation of what are here being termed "FP" and "dualistic FP."

FP as a theory:

1.) Is composed of interrelated terms and statements, both law-like generalizations and singular (observation) statements, expressing its conceptual categories and the relations between them. Many of

the laws are tacit and have never explicitly been stated by anyone.

2.) Functions to explain and predict human behavior in terms of propositional attitudes and perceptual experiences.

3.) Implies weak ontic commitments to the types-of-things for which it has indefinite singular terms, e.g., to beliefs, desires, pains and sensations of red.

Dualistic FP as a paradigm:

4.) Makes strong ontic commitments to both mental and physical states, events and/or substances.

5.) Postulates inherently private, internal subjective states which are not objectively accessible.

6.) Posits a non-physical mind which causally interacts with the physical brain and consequently is at least partially contrary to physical science.

7.) Recognizes or implies the possibility of an inverted spectrum, of mindless ones, of knowing what it is like to see red or to be a bat as something over and above the knowable physical facts, etc.

8.) Generally does, or at least can, allow for the possibility of life after bodily death.

9.) Maintains that human free will (in the Libertarian sense) exists, and that the notion of personal agency is correct.

10.) Can thus easily and straightforwardly allow for a strong concept of human responsibility and for retributivist theories of punishment.

11.) Can, in addition, base its meta-ethics on the inherent goodness/badness of certain mental or experiential states, e.g., pleasure and pain.

12.) Maintains that some, if not all, mental states are indubitable. We are certain of their existence and of their character. This is most clearly the case with perceptual experiences, but may also hold for the propositional attitudes.

13.) Differentiates observation terms from theoretical terms.

14.) Accepts the sensation theory of meaning, as a version of the extensional theory of meaning,[1] as the (correct) account of meaning of perceptual or observation terms. Allows for the acceptance of the network theory of meaning for theoretical terms.

15.) Accepts that a private language is possible in the case of mental state terms, at least those reporting sensory experiences.

Advantages: As a theory, there are some things to be said in favor of FP. Embeded in it is a wealth of traditional wisdom concerning our interactions with others. It expresses, by definition, our common-sense beliefs about ourselves. It has served us fairly well over the years. It has developed over a very long time and is reasonably complete, in the sense that it has something to say about most aspects of our lives. To a certain extent it is even testable

empirically. We can, for example, test for Smith's belief that it is raining outside by asking him if he believes it is raining outside.

As a paradigm, a dualistic FP allows for the possibility of life after death and unembodied minds or souls. It seems to be based on what has traditionally been taken for granted as self-evidently real, namely the contents of our minds and the apparent privacy of our inner states. Thus, it accounts very well for the qualitative character or the phenomenal qualia of our perceptual experiences, e.g., what it is like to feel pain or to see a bright shade of red. The undeniable reality of these sensory qualia together with their private character, allows for such interesting thought experiments as the inverted spectrum. It accords with our moral intuitions, that people sometimes act of their own free will and are thus responsible for those actions and deserve praise and blame for them. It also recognizes that certain things, e.g. pleasure, are inherently good and some things, e.g. pain, are inherently bad without need for justification. It also lets perceptual experience be the ultimate judge of the rightness or wrongness of certain hypotheses, since observations and the meaning of observation terms are to some extent independent of theory. From this summary it would seem that FP, as a theory within dualistic FP, has quite a lot going for it. Its truth and value are not often questioned by the non-academician.

Disadvantages: Unfortunately, perhaps, there also are many problems associated with accepting FP. As a theory, the language and conceptual categories of FP are exceedingly idiosyncratic, relative to

physical science. It is often vague, with many of its key concepts imprecise, lowering its testability and falsifiability. Similarly, relative to scientific theories, it fairs poorly when evaluated according to most of the other meta-theoretical criteria which have been proposed. FP is a very old theory with pre-scientific origins. It has not progressed significantly in the last few hundred (thousand?) years. In the areas addressed by it and NS, FP generally has a lower degree of explanatory power and predictive success, perhaps most importantly due to its imprecision.

As a paradigm, this view suffers from all of the well known problems of dualism. It is to some degree untestable and unfalsifiable due to its inherently private and subjective mental states, which brings it immediately into question. The exact relationship between minds and brains is as yet unspecified. Given the fact that minds are, in principle, not objectively observable, this situation is apt to continue. If the relationship is causal, then the laws of conservation in the physical sciences will be violated. Finally, the dualistic ontology of FP seems to violate the desideratum of ontic simplicity and thus may be susceptible to an application of Ockham's Razor. All in all, it is not in an enviable position. Given all of the associated problems, it is not surprising that there has been a tendency to move away from FP, from dualism and from the positive strong ontic commitments of mentalism generally.

- eliminative materialism

Following is a list of the many constituent elements of NS as a theory and EM as a paradigm. The latter is construed as broadly as possible and may even include functional descriptions, e.g., of second-order functional states of physical systems, whenever they are appropriate and consistent with the overall picture painted by the eliminative materialist. Again, the intent in not to argue that this is an exhaustive list, nor that every element is necessary. The list represents a brief but fair approximation of EM as presented in the literature, supplemented wherever necessary to make it as complete as possible.

NS as a theory:

1.) Is composed of interrelated terms and statements, both law-like generalizations and singular (observation) statements, expressing its conceptual categories and relations between them. It subsumes various sub-theories, e.g., those of neurophysiology, neuroanatomy, biochemistry, etc.

2.) Functions to explain and predict human behavior in terms of neural events, e.g., the isomerization of rhodopsin molecules in retinal rod cells, acetylcholine generating end-plate potentials at neuromuscular junctions, etc.

3.) Implies weak ontic commitments to the types-of-things for which it has indefinite singular terms, including electrical, chemical and structural events occurring within the brain.

EM as a paradigm:

4.) Makes strong ontic commitments only to physical states, events and/or substances. (Here, electrical events must be included under "physical.") Could allow for second-order functional states of physical systems.

5.) Maintains that all states, including our internal ones, are, at least in principle, publicly and objectively accessible.

6.) Rejects the notion of a causal interface between a non-physical mind and the physical brain, qua interactionism or epiphenomenalism.

7.) Does not, indeed cannot, recognize the meaningfulness of such notions as an inverted spectrum, people who have functioning brains but are mindless, or being able to know what it is like to see red or to be a bat as something over and above the knowable physical facts, etc.

8.) Does not acknowledge the possibility of life after bodily death.

9.) Maintains a deterministic doctrine, either hard or soft determinism. Denies the notion of personal agency or interprets it along physicalistic and deterministic lines

10.) Is therefore, more consistent with consequentialistic theories of punishment, qua behavior modification, than with retributivism.

11.) Must construe the notions of goodness and badness as being either meaningless or as being defined physicalistically, e.g., in terms of survival value for the individual and/or species. An attempt

might be made to define them in terms of what an individual likes or dislikes, but then `likes´ would require a physicalistic construal.

12.) Maintains that all judgements are theory laden, including introspective judgements. Hence, it rules out the possibility of any indubitable philosophical foundation based on theory neutral observations.

13.) Maintains that there is no real difference between observation terms and theoretical terms, nor therefore, between observation statements and theoretical statements. May even deny a distinction between analytic and synthetic statements.

14.) Accepts the network theory of meaning as giving a complete account of the source of meaning for all terms.

15.) Maintains that a private language is impossible, and that the sensation theory implies such a private language and hence must be rejected.

16.) Maintains that an intertheoretic reduction from FP to NS is not workable, not necessary and probably not possible. Argues instead for the outright elimination of FP.

Advantages: Given the progress and current success of contemporary NS relative to FP, it seems a **prima facia** reasonable hope that, in the not too distant future, NS will be superior to FP in terms of explanatory power and predictive success. This is especially so if one´s physicalistic theory is broad enough to allow for second-order functional levels of description and analysis. Given the

hypothetico-deductive method of justification, both NS and its weak ontic commitments will then be relatively well supported. There is already a higher degree of testability and falsifiability of NS given its thoroughly objective nature and its greater precision, and NS does not require translation in order to integrate with other scientific theories.

As a paradigm, the ontological monism of EM avoids all of the difficulties with dualism. Its bottom-up approach is attractive. Since it rejects reductivism it does not have to support the claim that FP is reducible to NS, nor does it face the (top-down) constraints implied by reductivism. An eliminative materialist is free to accept "the best theory available."

The recognition that one's theory intervenes between one's experiences and the objective properties held to exist in the world seems correct and is important, as is the recognition that even one's introspective judgements are theory laden. There is certainly something to the claim that the actual qualitative identity of one's experiences (however defined) is irrelevant to the objective features one takes oneself to perceive in the world. That is why no significant behavioral problem would arise even if there were such things as inverted spectrums.

Disadvantages: The main objection to NS, from the point of view of a dualistic FP, is that it is always going to be incomplete as a description of humans, principally because a.) the phenomenal qualia of

perceptual experience are beyond its scope, and b.) it does not address the mind as the non-physical origin of human voluntary action. Naturally the underlying basis for this claim of inadequacy is the dualistic strong ontic commitment of FP. There is a third objection which comes from the sophisticated functionalist which maintains that c.) NS does not, and perhaps cannot, operate at a sufficiently high level of description to be able adequately to handle such complex human attributes as are described within dualistic FP, e.g., "vague beliefs that some people are more lucky than other people." These objections are met, at least in part, when one places NS within the broader context of the EM paradigm. One other objection that might be brought against NS is that it currently cannot explain much human behavior. The hope that a future completed neuroscientific theory will yield a better explanation is, at this point anyway, just a hope.

There are also various objections to EM as a paradigm. (These topics are important and are only going to be introduced here. They will be picked up again below when EM and dualistic FP are evaluated as paradigms.) EM rules out the possibility of life after death. It denies, or redefines, human free will and personal agency. This move will tend to undermine morality unless "right," "wrong," "good" and "bad" can be defined within its framework. EM portrays FP as such a poor theory and paradigm that it is not even deserving of intertheoretic reduction, but instead is to be dropped outright. However, if FP really is so wrong it is difficult to understand how we, as a race, have managed to get along in the world and with each other

as well as we have. Naturally, the extent to which such points constitute objections depends very much on the paradigm toward which one is sympathetic. One might maintain that we have not gotten along all that well, partly perhaps because of such foolish notions as life after death, moral rightness and wrongness, etc.

More important in the present context is the trouble that EM (and NS) have when trying to account for what dualistic FP calls the phenomenal qualia of perceptual experience. EM (and NS) seem unable adequately to capture first-person reports of what it is like to see red or to feel pain. It may be objected that, in this, they have failed to recognize what was right with the Cartesian approach. They seem to miss the indubitability of the EXISTENCE, and of the NATURE or CHARACTER, of what is described within dualistic FP as the qualia of seeing red or feeling pain. (The same might be said of propositional attitudes, e.g., that we have indubitable knowledge of the existence and the nature or character of such things as beliefs and desires. However, the qualia case seems to be the stronger, and it will thus be the main focus of attention.)

There are three levels at which indubitable knowledge about qualia might be claimed. 1.) One might claim to have indubitable knowledge of the existence of something which, within FP, is best described as a pain. 2.) Besides the mere knowledge that something exists, one might further know (indubitably) that it is a certain type-of-thing, e.g., that it is a pain, or that it is a c-fiber firing. Here the debate turns on theory selection. Is this thing, whose existence is known

indubitably, of a type countenanced by FP, NS, both or neither? One might even have indubitable knowledge that some pain is a "stabbing pain" or a "throbbing pain." 3.) The most difficult, and perhaps the most important, sense in which one might claim to have indubitable knowledge of the nature of something which exists has to do with the ultimate ontological category to which that thing belongs. In short, can we know that, for example, some particular thing which FP labels `pain´ is in fact a mental something? Although the proponent of dualistic FP might accept all three of these kinds of indubitable knowledge claims, the first seems the most intuitively plausible. It may, even by itself, be enough to cast serious doubt on EM, as will be discussed in the last section of this chapter.

- reductive materialism

The reductive materialist will accept NS and the conceptual categories of FP insofar as they have been reduced to NS. As a paradigm, psychoneural reductivism would be very similar to EM. The following are two differences between the paradigm of reductive materialism and that of EM.

Reductive Materialism as a paradigm:

7.) Will allow for the meaningfulness of an inverted spectrum only if it can be given a physicalistic interpretation. Generally must deny the possibility of people who have functioning brains but are mindless, since that will be an out right contradiction, and of

being able to know what it is like to see red or to be a bat as something over and above the knowable physical facts.

16.) Maintains that an intertheoretic reduction from FP to NS is not only workable, but generally the correct way to proceed.

Advantages: As mentioned in chapter 3, there are certain advantages to the reductivistic approach. One has a pre-made set of conceptual categories with which to work. All that is done is to "reinterpret" what we have known all along, accepting the weak ontic commitments of FP while denying the mentalistic strong ones of the dualistic FP paradigm. This reinterpretation has the very significant advantage of allowing for expansion and progress of FP by making it a sub-theory of science generally. There is a sense in which this tack appears to preserve the traditional wisdom acquired through the ages. Hence, there is also a sense in which the reductivist seems to admit the reality of minds and mental events. Qua weak ontic commitments, the reductivist can allow that there are pains and sensations of red and that they are quite different from one another. `Pain´ thus expresses a useful conceptual category and denotes a particular type-of-thing in the world. The one (big) difference is that this pain is now interpreted as a neural event rather than as a non-physical event. In general, the reductivist enjoys all of the ontological advantages of the materialistic monist, particularly simplicity and theoretical integration.

Disadvantages: The reductivist accepts the reality of the qualitative character of perceptual experience, qua weak ontic commitments to phenomenal qualia as types-of-things. However, he rejects such qualia in the sense of a strong ontic commitment to mental events. Given this, and the acceptance of the type/type identity theory, the reductivist must somehow explain how the excruciating pain associated with spilling a large mug of scalding hot tea onto one's lap "just is" a bunch of c-fibers firing. Further, as the dualist will be quick to point out, there is an implied denial of human free will (again in the sense of the libertarian rather than the soft determinist) and thus human values generally will be undercut unless (and until) they can be accounted for under the materialist's scheme of things.

In addition to the general problems with materialism, the reductivist must resolve the problems associated with his proposed transition from FP to a physical theory. The success of the proposed reduction will depend heavily upon the isomorphism between FP and (a sub-set of) NS. If FP is to be reduced to NS then FP must divide the world into types in a way that, from the point of view of NS, is correct. This means that our ancestors must have serendipitously anticipated contemporary neuroscience when they were developing their notions of human nature. There are also all of the translational problems that will be associated with the conversion of FP language into NS language. Finally, there is the pragmatic, methodological consideration of the implied top-down approach, seemingly forced upon

the neuroscientist by the reductive strategy. There are good reasons to believe that it would be better to examine the human nervous system and develop an appropriate set of conceptual categories from there rather than to have to take the approach of trying to find neural correlates to such folk psychological notions as beliefs in gods and fear of falling.

- concluding remarks

Now that the competing paradigms have been presented, several points concerning the debate between them must be made. The first has to do with Churchland's example of the infrared people. Churchland presents this story as an argument for the network theory of meaning, and thus indirectly in favor of EM. Examination of this case will help to explicate the relationship between EM and the network theory. It will also serve to further support the claim that the dualistic FP and EM are competing paradigms. The second topic concerns the analogical arguments that have been presented in support of the possibility of the elimination of FP. The third concerns the examination of some of the disputes between proponents of EM and proponents of a dualistic FP, with an eye towards showing how viewing the dualistic FP and EM as paradigms explains much of the apparent irreconcilability of those disputes.

There is a general problem which is going to be present whenever one is dealing with an eliminativism. A pure form of EM will so thoroughly reject FP that there will be serious communication problems

resulting from the sparsity of common vocabulary. This is one of the features which points toward the construal of EM and the dualistic FP as paradigms. However, it does make it difficult to discuss the key issues. There are two approaches that one might take in an attempt to overcome these difficulties.

1.) The first involves the interpretation of eliminativism which was proposed by Rorty, as presented in the last section of chapter 4. Recall that this view maintains that there is no "strict identity" between lightning and thunderbolts of Zeus, even though their use of `thunderbolt of Zeus´ was getting at something, i.e., discharges of atmospheric electricity. One way of thinking about this move is to adopt FP statements hypothetically; e.g., "insofar as `qualia´ refers to anything, then what it refers to is a brain state." One advantage of this proposal is that the proponent of EM can provisionally adopt the views of dualistic FP in order to try and point out "internal" problems with the competing paradigm. In the infrared peoples´ case, for example, the claim would be that even if we allow that mentalistic terms are meaningful, even if we temporarily work within the FP paradigm, still the claim that the meanings of words comes from sensory experience must be false.

2.) In order to allow for discussion, the paradigm herein called `EM´ might be construed in a broad enough sense that it will allow that at least some intertheoretic reductions and identifications might occur. It might also be construed so as to encompass second-order functional state descriptions insofar as they concern physical systems.

Allowing for EM to be so construed permits one to discuss the strongest possible materialistic paradigm.[2] It will be primarily eliminativistic, thus avoiding the problems of general intertheoretic reduction. But it will allow for an occasional identification, thus enabling us to keep those few parts of FP that are worthwhile. It is, thus, close to what Churchland has called "Revisionary Materialism." Still, it will be in keeping with the primary tenet of EM proper, namely that no wholesale intertheoretic reduction is to be condoned. (There will, of course, be a problem with this piecemeal "reduction" insofar as one accepts the network theory of meaning.)

Throughout the remainder of this chapter the materialistic paradigm that will be under consideration is EM. However, EM will be construed in a broad enough manner so as to include an occasional case of reduction/identification between FP and NS and/or a provisional adoption of those parts of FP required for discussion.

* * * * *

Problem Cases viewed from the Competing Paradigms

- the infrared people

Recall that Churchland maintains that the heterophonic sensation-guided translation of the infrared peoples' vocabulary into ours leads to unacceptable results.[3] Translating their `cold' as our `black', or our `white' as their `hot', results in such absurd translations as "food keeps better in a black place" and "snow is hot." Such

translations are not correct, hence the problem. He states that "the impossibility of the heterophonic translation...is just the impossibility of the thesis that the meaning of the common observation terms at issue is given in sensation." In other words, the sensation theory of meaning must be incorrect. Alternatively, the homophonic translation, where their `hot' (as seen) means the same thing as our `hot' (as felt), has every empirical virtue that a translation can have. This implies that "...the meaning of the relevant observation terms has nothing to do with the intrinsic qualitative identity of whatever sensation just happens to prompt their non-inferential application in singular empirical judgements."[4]

One problem should be cleared up immediately. Consider the way that the example of the infrared people has been set up. They see hot things in a way that is "qualitatively similar" to the way that we see white things. Our objective intentionalities are different due to the different retinal response to electromagnetic radiation. In a strict sense, if one is to be a steadfast proponent of EM then the example makes no sense whatsoever. What could be meant by the instruction to suppose that "...so far as the intrinsic nature of their visual sensations is concerned, the world `looks' to them much as it looks to us in black-and-white prints..."?[5] What is a "heterophonic sensation-guided translation" anyway? Objections of this sort can be avoided, and discussion can continue, by working with the broad interpretation of EM presented above.

The example of the infrared people works as an argument for the network theory of meaning by way of arguing against the sensation theory. Recall that the sensation theory is an integral part of the FP paradigm, and the network theory is an integral part of the EM paradigm. Hence, to argue against the sensation theory is to argue against the FP paradigm. It will be argued below that the example leads to a problem for the sensation theory and for the dualistic FP only if one already accepts EM (and thus the network theory). Hence, the extent to which the infrared example poses a problem depends upon one's paradigm. The proponent of EM will have a problem disproving the correctness of the sensation theory of meaning if the only arguments available come from within the materialistic paradigm. This is typical of interparadigmatic disputes.

The broad EM account: There are various ways in which one might try to undercut the apparent difficulties presented by the infrared case. Consider, for example, that such terms as `intrinsic nature of visual experience' refer to neural states. On this view, "the experience we judge to be a sensation of whiteness, the infrared people judge to be a sensation of heat" will be a claim about differing second-order neural states operating on similar first-order neural states initiated at our respective sensory transducers.

Thus construed, there would be no translational problems. We say "snow is white" and mean that our seeing snow sets up a HUMAN brain state HB_1 in us. When they say "fire is hot" they mean that their seeing fire will set up an INFRARED PERSON brain state IB_1 in them. If

we assume that HB_1 and IB_1 are qualitatively (i.e., first-order neural) identical, then there is a sense in which our `white', qua $HB_1$, will be their `hot', qua IB_1. Snow (indirectly) affects our sensory transducers the way that fire affects theirs.

However, if we set up a translation manual we would have to recognize the ambiguity inherent in observation terms such as `white', `black', `hot' and `cold'. There is a sense in which they might be used to report on our INTERNAL states (on what Churchland called the "objective intentionality of sensation"). Here we would translate their `hot', as a report of their own neural state, $IB_1$, produced by such things as fire, as our `white', as our report of our neural state, HB_1, produced by such things as snow. Hence, in cases of self-reference or first person reports, their "hot" IS our "white". This sort of a translation would work just fine. It would convey just the right sort of information if they wished to know something about the similarity of the states (phenomenal or first-order neural) produced when we see snow and they see fire.

This would not however, imply that we would have to translate our claim ABOUT THE EXTERNAL WORLD that "snow is white" into their claim about the same external world that "snow is hot", based on the qualitative similarity of their IB_1 to our HB_1. The origin of the apparent problem in the example of the infrared people is an ambiguity inherent in the observation terms. The problem is only apparent, since it results from an equivocation on the key terms. The ambiguous use of those terms consists in their referring to something internal to us (or

to the infrared people) on the one hand, and as referring to something external on the other.

However, it might be that, at least to some extent, the extensional theory of meaning is implicitly accepted in the above analysis. Part of what we might mean by `appearing white´ is captured by the intended referent of that term. Hence, the same term can mean very different things when used internally and externally. Insofar as the infrared people and we use observation terms extensionally and externally, there will be no translational difficulties. We can both say that snow is cold. Insofar as we all use observation terms extensionally and internally there is also no problem. Snow appears white to us and cold to them, even though the visual appearance of snow as cold for them is "like" our phenomenal experience of black. The problem arises when we mix the internal and external uses. But that problem simply is one of equivocation.

The relationship of their hot-as-seen to our hot-as-felt is similar to the relationship of a ball-as-round based on our seeing it and a ball-as-round based on our feeling it. This relationship, and the distinction between the internal and the external uses of observation terms is akin to Berkeley´s view on the matter. Within the context of his idealism, Berkeley maintained that the proper objects of sight are not external to the mind and that, since seeing and touching are so different qualitatively, we never see and touch the same thing.

"...a man is easily convinced that bodies and external things are not properly the object of hearing, but only sounds, by the mediation whereof the idea of this or that body, or distance, is suggested to his thoughts. But then one is with more difficulty brought to discern the difference there is betwixt the ideas of sight and touch, though it be certain a man no more sees and feels the same thing than he hears and feels the same thing."[6]

The dualistic FP account: Now, the materialist's account above could be converted from a NS explanation to a FP explanation by the superimposition of the sensation theory of meaning over the extensional theory. One could accept a mentalistic phenomenal experience, rather than a neural state, as the referent of the internal use of `white'. Since Churchland presents the infrared example as an argument against the sensation theory, it will be worthwhile to reiterate the points made above in the context of the FP paradigm, while being a bit more formal.

An observation term can be used descriptively, or referentially, in two very different ways. It may be used to describe, or refer to, some object out there in the external world which is observed. The same term may also be used to report something internal, e.g., a neural event or an experience associated with an observation. For example, one may use the term `red' to describe the reflective characteristics of a physical surface. But `red' may also be used to report or describe an internal state. Within a dualistic FP, this internal state is thought to be something mental, being of a particular phenomenal type. Within EM it can only be thought to be something physical, e.g., as a particular neurophysiological type. On either view of the internal state, there is a distinction to be made between the internal

and the external uses of observation terms. Since this ambiguity in involved in the dispute between the dualistic FP and EM, it will be beneficial to introduce a technical device to eliminate it.

The subscripting device: The method that I will use is to subscript key words in order to keep clear which type of reference is in use. When I wish to refer to something internal to the perceiver the word will be subscripted with an `i', to indicate "internal." When using an observation term to refer to something external to the perceiver an `e' will be used, for "external." Almost any type-word can be subscripted in this way. Adjectives and adverbs can be subscripted, as may nouns and even verbs.

For example, I may be seeing something$_i$ red (i.e., having a visual experience best described within FP by "a red thing"), even though I am fairly sure that there is nothing$_e$ out there in the external world, at least nothing that is red$_e$. This was, presumably, the case with MacBeth. I may even believe that red$_e$ does not exist out there in the world, perhaps because I do not believe in the existence of so-called secondary qualities. In adjectival and adverbial clauses such as `red square', it will be most efficient to subscript only once and read the clause as if it were hyphenated. Hence, rather than `large$_e$ red$_e$ square$_e$', `large red square$_e$' will be written and taken to mean `large-red-square$_e$'. Additional subscripting will be used only if necessary. (No mixed uses of subscripting, e.g., "red$_i$ square$_e$", will be allowed since that tends to reintroduce the ambiguity.)

Although it is not intended that the discussion should turn to the debate over the primary/secondary quality distinction, it is fairly clear that the subscripting device here described is relevant to that debate. One might interpret the claim that "ideas of primary qualities resemble those primary qualities while ideas of secondary ones do not" as a claim of resemblance between such things as square$_i$ and square$_e$ but non-resemblance between such things as red$_i$ and red$_e$. (We will leave the meaning of `resemblance' open here.)

There is a sense in which one can use this subscripting device and yet remain ontologically neutral with regard to strong ontic commitments.[7] If one says that one sees something red$_i$, then one is committed in the weak sense to the conceptual category which countenances red internal types-of-things. However, it might be meant in either the dualistic FP sense or the materialist sense. That is, one could intend to use the term `red$_i$' to refer to a mental state or to a brain state. Hence, both the (reductive) materialist and the folk psychologist should be content with the use of `red$_i$'.

The situation is fairly analogous with the use of terms with the `e' subscript. If one claims to see a red square$_e$ one would thereby be committed to the existence of red squares out there in the world. That would not, however, force much in the way of strong ontic commitments upon one. That red square$_e$ could be a material/corpuscular object, a relatively strong gravitational distortion in space/time, an idea in the mind of God, or just about anything else. The distinction between weak and strong ontic commitments may be a bit more clear in the

internal case because there we have two competing paradigms more obviously in contention.

In summary, use of such terms as `red$_i$´ and `red$_e$´ will remove the basic ambiguity inherent in observation terms. `Red thing$_e$´ is meant to pick out an objective property of something out there in the world. `Red thing$_i$´ is meant to pick out a property or state of the perceiver who is making the observation. It is now time to take another look at the case of the infrared people. Through the application of this subscripting device it is possible to find a way out of the problem posed by Churchland.

The infrared example with sub-scripting: Recall that the problem with the "heterophonic sensation-guided translation" of the infrared people's vocabulary into ours resulted in such absurd translations as "food keeps better in a black place." This translation was the result of the sensation theory of meaning as applied to their use of the term `cold´. When they see that something is cold they have a visual experience which is qualitatively similar to the visual experience that we have when we see something black. Hence, according to the naive sensation theory, i.e., pre-subscripting, their `cold´ means the same thing as our `black´, due to similar sensory experiences associated with the application of the terms.

The above subscripting device allows for the correct analysis of the situation through the recognition of the ambiguity of the relevant observation terms. First, assume that the infrared people know about

colors in the world through a means other than direct observation, much as we know about infrared and ultraviolet. Next, let us take the view of an omnipotent observer so that we can know about the qualitative similarity of their phenomenal states when they see hot things to our phenomenal states when we see white things. (Notice that this omnipotent observer move is only required when we are working within dualistic FP. In EM (and NS) we could do it by monitoring neural states.) Given these two assumptions, we can construct the following abbreviated translation manual:

	Infrared peoples' language		Normal human language
1.	hot_e	:	hot_e
2.	$cold_e$	:	$cold_e$
3.	$white_e$	:	$white_e$
4.	$black_e$	:	$black_e$
5.	hot_i (visual)	:	$white_i$ (visual)
6.	$cold_i$ (visual)	:	$black_i$ (visual)
7.	hot_i (tactile)	:	hot_i (tactile)
8.	$cold_i$ (tactile)	:	$cold_i$ (tactile)
9.	$Snow_e$ is $cold_e$	:	$Snow_e$ is $cold_e$
10.	$Snow_e$ is $white_e$	:	$Snow_e$ is $white_e$

Since the infrared people do not have the necessary sensory transducers to detect electromagnetic radiation between the infrared and the ultraviolet, they have no internal terms for white and black. This can lead to confusion when the discussion turns to the phenomenal experiences generated by external objects. We can avoid such difficulties if we place an `R' immediately before a term referring to an internal state of the infrared people and an `H' for normal humans.

For example, "R $Snow_i$ is R hot_i" and "H $Snow_i$ is H $white_i$". Given this convention and the above translation manual, we can now remove the apparent problem brought out by the example of the infrared people.

The solution:

The (false) claim that

(a) "food keeps better in a black place"

should be replaced by the true claim that

(b) "$food_e$ keeps better in an R $black_i$ place".

Similarly,

(c) "snow is hot"

should be replaced by

(d) "$snow_e$ is H $white_i$" (visual) or

(e) "$snow_e$ is R $cold_i$ (visual)

Notice that

(f) "$white_e$ appears H $white_i$"

brings out the deficit of the infrared people's sensory systems. There is no similar statement in their language since they cannot see_i $colors_e$. There will thus be no statements of the form:

(g) "$white_e$ appears R $..._i$"

where "..." stands for any term.

Hence, Churchland's infrared example presents a problem for the dualistic FP only insofar as one rejects the sensation theory of meaning, or at least ignores one of its crucial elements, namely the

recognition of the ambiguity of observation terms. Without that recognition, the claim that some terms get their meanings from sensory experience cannot be adequately appreciated. Actually, the example fails even from a materialist's point of view since, in the above discussion, all of the terms subscripted with an `i' can be interpreted physically, e.g., within NS as referring to first-order sensory neural states.

- historical analogies

From the point of view of common sense, there is an apparent absurdity in the bald statement of the eliminative materialist, that someday we will realize that sensory experiences such as feeling pain and sensing red simply do not exist. Naturally, the way that this is supposed to be able to occur is through the abandonment of FP. If we drop the theory to which the concepts of pain and sensation of red belong, and within which "pain" and "sensation of red" are defined, then claiming that you are in pain or sensing red will be quite meaningless. Referring back, once again, to the last section of chapter 4, we see that one way (Rorty's) of formulating the claim of EM is:

1) What some people now call "sensations" are really just brain processes

In order to supply some intuitive appeal to this approach, several analogies have been presented which purportedly illustrate the underlying "non-strict" identity.[8] Consider the following examples:

2) What people used to call "Zeus's thunderbolts" are discharges of electricity

3) What people used to call "caloric fluid" is nothing but the motion of molecules

4) What people used to call "witches" are nothing but psychotic women

In all of these cases we simply dropped one type of description in favor of the other. Rorty's suggestion of the "what people used to call `X' is nothing but..." format is a useful one. For one thing, the underlying "non-strict" identity diminishes our tendency to look for the extensional equivalence of such things as "caloric" and "mean molecular kinetic energy", which should make the eliminativist happy. A second point in its favor is that it forces us to keep distinct our descriptions and that which we are intending to describe. If we borrow the sub-scripting device presented earlier, and we are careful to recognize that our current ways of analyzing things are also descriptions, we can restate the above as:

2a) What people used to call "Zeus's thunderbolts$_e$" are now described as "discharges of electricity$_e$"

3a) What people used to call "caloric fluid$_e$" is now described as "the motion of molecules$_e$"

4a) What people used to call "witches$_e$" are now called "psychotic women$_e$"

The thesis to be defended here is that the analogies presented are faulty and that the appeal that they apparently have for the proponents of EM is due to a communication problem with proponents of FP. The disanalogy becomes apparent when we attempt to restate number one.

1a) What people now call "sensations$_e$" will one day be described as "brain processes$_e$"

In numbers 2a), 3a) and 4a) above, all of the terms are subscripted with an "e" because the objects under discussion are external; they are out there in the world. However, in the case of sensations, we are addressing something that is internal to a perceiver. We are talking about ourselves. Sentence 1a) would thus be better formulated as:

1b) What people now call "sensations$_i$" will one day be described as "brain processes$_i$"

Further, the major issue is really the first-person reports. The apparent absurdity of the ontology of EM supposedly comes from the indubitability associated with my experiencing my pain, not from my recognition of someone else's pain. Descartes argued that the existence of HIS thinking was indubitable for HIM. In order to focus on the real objection to EM, let us construe 1b) as a first-person account. Since the materialist maintains that his is the correct view

(and since he need not be intending to predict the future), we can convert 1b) to:

1c) What any given subject now calls "my sensations$_i$" is correctly described by that subject as "my brain processes$_i$"

The use of "my sensation$_i$" indicates the relative nature of the subscripting device in use here. The "i" does, after all, indicate internal to some subject. My neural states are internal for me and external for you. The same is true in the case of my sensations.

There is a final important clarification required concerning the use of the "i" and "e" subscripts. For both the dualist and the materialist, the "e" subscript is to be used when talking about the world out there, which is external to my physical body. The dualist will also use the "e" subscript when talking about the physical brain, even his own. The materialist, on the other hand, will use a term subscripted with an "i" to refer to anything (physical) going on inside the boundary of the skin. The dualist will use "i" terms only to refer to things internal to the mind.

The dispute between the materialists and the dualists can now be characterized as a dispute over the use of the "i" or the "e" subscript where our neural states are concerned. For the materialist, if "sensation$_i$" refers to anything then it must refer to something physical, e.g., a brain state. Hence, in all cases the internal/external boundary must be the skin. For the dualist, the physical body is external to the mind, so "sensation$_i$" will be used to

refer to a mental experience. Hence, the ambiguity of observation terms brought out by subscripting takes on different forms for the materialist and the dualist. Within FP, the subscripts "i" and "e" indicate mental or physical. Within EM those subscripts indicate inside or outside the skin.

The disanalogy between the sensations/neural process case and the Zeus's thunderbolt/lightning, caloric/mean kinetic molecular energy, and witch/psychotic woman cases should now be clear. These latter are all cases of ontological elimination that occurred in the external world. That is not the level of dispute between the dualist and the materialist. The dualist can legitimately claim, at least from his point of view, that there has never been a single elimination, or reduction, of the relevant sort to be used as an analogy.

The claim that the materialist is making, FROM THE POINT OF VIEW OF A DUALISTIC FP, is:

1d) What I now call "my sensations$_i$" would be correctly described by me as "my brain processes$_e$"

The difference between 1c) and 1d) exemplifies the difference of the main claim being made by EM as viewed from within EM and from within the dualistic FP. From within a dualistic FP, the claim of the materialist is as expressed in 1d) and, as such, is quite absurd and without historical precedent. From within EM, the claim being made is as expressed in 1c). Insofar as the "i" subscript merely indicates "inside the skin", the cases expressed in 2a), 3a) and 4a) constitute

perfectly good analogies. What occurs inside the skin is just as physical as what goes on outside of it. Judgements concerning what is going on inside the skin are just as theory-laden as judgements concerning what goes on outside of it. We can eliminate a theory containing the term "sensation" just as easily as, and for the same reasons that, we can eliminate a theory containing "Zeus." The only thing special about first-person accounts is that each human brain can engage in self-assessment without the mediation of sensory transducers.

The point of the above discussion is not to argue for or against EM, or dualistic FP. The point is, rather, that the analogical arguments presented in favor of EM are convincing or absurd to the extent that one accepts the main tenets of materialism or dualism. That these arguments can be either convincing or absurd is explained by the fact that there is a significant, indeed crucial, difference in what is accepted as the main claim being made by the eliminativist. A proponent of EM says something like 1) above, and means by it something like 1c). The dualist interprets it as something like 1d) and then they battle about the coherence of the claim. This conflict may not be resolvable so long as the parties remain in their respective paradigms. The materialist cannot mean by 1) what the dualist means by it, viz. 1d), and the dualist really cannot accept the claim of the materialist that 1c) captures all there is to say on the matter.

- on the physical definition of "qualia"

Nagel: In an attack on reductivism, and on physicalism generally, Thomas Nagel attempts to show that there is something which we may call the "subjective character of experience" and that it is just what has not been accounted for by the physicalist. Nagel states that:

"It is useless to base the defense of materialism on any analysis of mental phenomena that fails to deal explicitly with their subjective character... If physicalism is to be defended, the phenomenological features must themselves be given a physical account. But when we examine their subjective character it seems that such a result is impossible. The reason is that every subjective phenomenon is essentially connected with a single point of view, and it seems inevitable that an objective, physical theory will abandon that point of view."[9]

For Nagel, this subjective character is what is involved in conscious experience and for any organism which is conscious, it must be like something to be that organism. We know, for example, what it is like to be us. We do not, however, know what it is like to be a bat. The echolocation used by bats is not at all like any of our perceptual experiences, hence we can't know what it is like to be a bat. Until physicalism can account for this subjective character of experience it will be inadequate.

Clark: This topic of the subjective character of experience is addressed by Austen Clark.[10] He attempts to show that one can construct a physicalistic definition of qualia based on a functional definition of qualitative similarity. He maintains that this can be done without circularity and without being based on unanalyzed

phenomenal properties, as is the case with J.J.C. Smart's method where "sensation of red" is defined as "the kind of thing that normally goes on when I see something red".

His method is based on a purely physicalistic notion of indiscriminability, as an ascription of a statistical relationship between sensory inputs and behavioral responses in a forced choice setup. A sensory input is a stimulus activating a sensory transducer. The behavioral output involves the subject's distinguishing between two elements presented in a test situation. If the subject's correct responses over time approach randomness, then it can be said that the two elements are indiscriminable. Clark points out the fact that indiscriminability is non-transitive, recognizing the existence of "just noticeable differences" (JND), and thus distinguishes between indiscriminability and identity. For example, the element x may be indistinguishable from y, and y may be indistinguishable from z, even though x and z are distinguishable. In such a case, x cannot be identical to y (nor y to z) even though they are indistinguishable.

According to Clark, the outcome of such psychophysical techniques would be an ordering of potentially detectable differences in stimuli. This would give one the ability to describe physically the notions of "same qualia" and "different qualia", or "looking phenomenally the same" and "looking phenomenally different", namely in terms of behavioral response to sensory transducer activity. This description could go beyond mere indiscriminability, and pick out qualitative

identity for classes of stimuli where each element is indistinguishable from the other elements in the same set of stimuli.

If Clark is correct, then sensory qualities can now be viewed as corresponding to classes of objects whose stimulatory encodings fall into equivalence classes which are qualitatively identical. For example, color will be handled with a relational account such that two things are the same color if they fall in the same equivalence class with respect to phenomenal identity as measured psychophysically. We can go on and name individual equivalence classes indexically, by picking out an exemplar which characterizes the class and then defining that class, e.g., we might use the name `red' for a class of colors. On this account it is the case that qualia are properties of neural states, since it is the stimulus as an event in a sensory transducer which is being measured psychophysically.

Clark takes his analysis to yield a reply to Nagel's call for an objective characterization of phenomenal properties. Clark maintains that psychophysics can tell us "what it is like to be a bat" by picking out the discriminatory equivalence classes for the bat. If the bat can distinguish x from y, then it cannot be the case that, from the bat's point of view x is phenomenally the same as y. "In short, all the qualitative, phenomenal, or subjective likenesses and differences among experiences of the bat could be identified extensionally from the structure of its discriminations. So we get an `objective' characterization of what it is like to be a bat."[11]

Within the present context it is a difference in their respective notions of "what it is like to be a bat" that is of central importance. It is highly doubtful that Clark's account is the kind of thing that Nagel has in mind. Clark recognizes this in the sentence following the above passage where he says: "Of course such a description does not enable us to experience the world the way the bat does, so in that sense it does not answer the question `What is it like to be a bat?'" However, Clark's main point is that an objective characterization of what it is like to be a bat is possible.

What we have here is a basic disagreement over the important sense of what it is like to be something. Nagel's sense can be viewed as constituting an exemplar for dualistic FP as a paradigm. Clark's sense reveals a significantly different theoretical point of view where the importance of Nagel's case is greatly reduced, even to the point of insignificance. It is replaced by an account of what it is like to be something which comes from psychophysics. There is no need to argue that Clark's analysis constitutes an exemplar for a scientific theory. One need only view it as an attempt to respond to an argument in favor of FP. The main point is that what is an exemplar in one theory is a matter of secondary interest and secondary importance (at best) in another. Thus, what is going on here has much in common with a paradigm shift.

- Mary and her brain states

Frank Jackson has argued against what he calls "the thesis of Physicalism", which he takes to be the view "that all (correct) information is physical information",[12] or that the actual world is entirely physical and hence "complete physical knowledge is complete knowledge simpliciter."[13] As an alternative view he maintains that if you have all of the physical information you still will not have information concerning the hurtfulness of pains, the itchiness of itches, the taste of a lemon or the smell of a rose.

He presents two related thought experiments in support of his view. First of all there is Fred, a person whose visual sensory system is such that he can easily distinguish two shades of red that no other person is able to distinguish. He can, for example, perform sorting tests of red objects at which others fail. Jackson maintains that this example shows that not all information is physical information. Even if we knew all of the physical facts about Fred, we still would not know something that Fred knew. Even if we knew all about Fred's physiology and behavior, there would still be something about Fred that we would not know. We would not know what it was like for Fred to see these two colors of red. We have all of the physical information about Fred, yet we do not know everything there is to know about him. Therefore, physicalism leaves something out.

This is one formulation of Jackson's "Knowledge Argument", which purports to show that "one can have all of the physical information without having all the information there is to have."[14] He

distinguishes his argument from the above view presented by Nagel,[15] since he concentrates on knowing rather than on "what it is like to be" something. What is important for Jackson is the fact that there is something about Fred's experience of which we are left ignorant even given all of the physical information. He takes this to be different from knowing what it is like to be Fred, in the sense of sharing his experiences "from the inside." Rather, there is a quality of his experience about which we cannot know given only physical information. Jackson's view also stands in clear contrast to the way that Clark would describe Fred's situation, based on psychophysical measurements.

Jackson's second example involves Mary, a neuroscientist who has remained in an achromatic environment for her entire life. She has, however, managed to acquire all of the physical information that there is to be had concerning physiology, anatomy, biochemistry, etc. about the outside world. In particular, she knows all about what goes on when other people see ripe red tomatoes. She has all of the physical information about them, yet, when she finally leaves her achromatic environment and sees her first ripe tomato, she will learn something new about other people. She will acquire information about them that she did not have before. She will learn what it was like for them when they were seeing red tomatoes. What is important is that she will learn something about the previous experiences of others, something that she did not know while in her lab even though she had all of the physical information. Hence, her previous information was incomplete.

Hence, the physicalist's thesis is false. The argument is summed up as follows[16]:

ARG. 1

(1) Mary (before her release) knows everything physical there is to know about other people.

(2) Mary (before her release) does not know everything there is to know about other people (because she learns something about them on her release).

Therefore,

(3) There are truths about other people (and herself) which escape the physicalist story.

It is important to distinguish knowing-that from knowing-how, and knowledge-by-acquaintance from knowledge-by-description. Jackson claims that it "can hardly be denied" that Mary acquires something very significant of a knowledge kind when she first ventures out of here achromatic lab.[17] What she gains is "knowledge-by-acquaintance", and perhaps some "knowing-how". She is now acquainted with red things, she knows how to see colors. She will now be able to imagine what seeing red is like, etc. This new knowledge-by-acquaintance or knowing-how can be adequately handled by the physicalist and non-physicalist alike. However, Jackson maintains the further claim that Mary has certain additional knowledge-that after leaving her lab. She now knows that it was like such and so for all of her colleagues when they were seeing red all that time she was in her impoverished environment. It is this

knowledge-that concerning qualia which lies at the heart of Jackson's Knowledge Argument.

The physicalist's account: Whether or not one shares Jackson's propensity to be a "qualia freak", it behooves one to look at his argument from a materialist's point of view. What does this talk about Mary knowing something amount to? Insofar as one allows for partial reductions within EM, one will say that by "knowledge" we must understand a certain type of brain state. (Included under "a certain type of brain state" will be something about the relationship of the brain state to the world such that knowledge will be a different state from belief.) If one were more purely an eliminativist, then "knowledge" would simply be dropped in favor of the conceptual categories and language of NS. Whichever version of materialism is held, it can be said that:

(a) If `knowledge' refers to anything then what it refers to is something neural.

The referent may be a token or a type of neural state or event. Let us use `brain state' as a most general term to indicate this neural something. If we (temporarily) accept the antecedent of (a) above, and agree to speak of types rather than tokens, then the following statements will be true:

(b) `Knowledge-how', `knowledge-that' and `knowledge-by-aquaintance' all refer to different types of brain states.

If Mary knows all of the physical facts then she is in a certain complex brain state, or has had a particular range of (types of) brain states. However, Mary has never been exposed to chromatic light. Hence, there are certain sorts of brain states in which she has never been, i.e., those constituting knowledge-by-aquaintance of colors.

When Mary steps out of her lab her brain undergoes certain changes in response to new and unique types of stimuli, and she now has knowledge-by-aquaintance of red. This much Jackson is willing to grant. But Jackson makes the further claim that she now knows something new about other people, something over and above the physical facts. Specifically, she knows that this is what it was like all along for them to see red. This is the central and important claim.

For the materialist, if this knowledge-that is anything it is a (type of) brain state. Such a materialist might view knowledge-that as a secondary, proximal or central neural state whose cause (and object?) is a primary, distal or peripheral neural state, i.e., the state of a sensory transducer. For example, knowing what it is like to see red might be secondary to seeing red, which is the brain state associated with retinal stimulation by chromatic light. Since there are several levels of knowledge, i.e., of brain states, it will be helpful to set up a materialist's translation manual.

(c) "sees red" = "brain state x" (BS_x)

(d) "knows what it is like to see red" = "brain state y" (BS_y)

(e) "knows what it is like for others to see red" = "brain state z" (BS_z)

Given the above manual, Jackson's argument begins with the claim that the achromatic Mary has never been in BS_x, BS_y or BS_z. She has never seen red (doesn't have that knowledge-by-aquaintance) and thus does not know what it is like for herself or others to see red.

It might be granted that some "knowledge-that" brain states can be caused equally well by chromatic stimulation or by achromatic stimulation, or by auditory stimulation, etc. Much of our so called "propositional knowledge" is like that. It does not follow however, that ALL knowledge-that can be so caused. It is quite possible that there is some such secondary knowledge-that which can only be initiated by certain primary knowledge-by-aquaintance. For example, BS_x might be required for the initiation of BS_y, which in turn might be required for BS_z. Thus, the initiation of these new brain states might be possible only when Mary is first exposed to a chromatic environment. On this view, to say that "Mary learns something new about others" after she sees her first tomato is just to say that BS_z is initiated in her only after, and only on account of, being in BS_x.

If we accept this as possible, then the above argument (Arg. 1) does not refute physicalism. In fact, premises (1) and (2) are inconsistent. If we assume that the first premise is true, then Mary must have been in BS_z since that brain state is part of the set of brain states which constitute "knowing everything physical there is to know about other people". However, this makes premise two false, since learning something new here means newly acquiring BS_z. On the other hand, if premise two is true, then the first premise is false, since

the knowledge under consideration is a BS_2 and premise two says that Mary has never been in the type of brain state which is referred to by "knowing what it is like for others to see red", i.e., she has not been in BS_z. She could not have been in such a state since she has not been in BS_x or BS_y, which are the requisite causes.

It is worth summarizing this argument:

ARG. 2

(1) If `knowledge', `knowledge-that', `knowledge-how', and `knowledge-by-aquaintance' refer to anything, then they refer to (types of) brain states.

(2) To say that "Mary knows everything physical there is to know about other people" is to make a claim about Mary's brain states (and their relationship to certain facts in the world).

(3) If `learn' refers to anything, then it refers to the acquisition of a (special sort of) new brain state. (Or, alternatively, to a new functional/behavioral disposition which ultimately depends on a new brain state, at least for creatures like Mary.)

(4) To say that "Mary learns something new about other people when she is released from her lab" is to make a claim about her acquisition of new brain states.

(5) Hence, after Mary is released: "She now knows what it is like to see red", "she now knows what it was like for others to see red", "she has new knowledge-that concerning others" and "she has new knowledge-by-aquaintance related to seeing red" are all true statements. However, each of these must be interpreted along the lines of having new brain states, brain states of a type that she did not have while in her lab. Specifically, Mary now has BS_x, BS_y and BS_z.

(6) Hence, before her release, either Mary did not know everything there was to know about others, i.e. she had never been in BS_z, or she did not learn anything new upon being released from her lab, i.e. did not newly acquire BS_z.

(7) Hence, Jackson's argument fails since premise one is inconsistent with premise two, as can be seen by translating into brain state talk, whereby premise one says Mary had a certain type of brain state and premise two says that it was newly acquired by her when she stepped out of the lab.

Analysis: The dispute between Jackson and the materialist can be characterized as a disagreement over the status of BS_z, as either knowledge-by-aquaintance or knowledge-that. This is a problem concerning the extent to which (d) and (e) above are to be included in the type of brain state identified with knowledge-that or the type identified with knowledge-by-aquaintance. Both parties agree that Mary gains new knowledge-by-aquaintance since, ex hypothesis, she has never seen anything red before leaving her lab. BS_x is clearly this type of

knowledge. If, as the above argument states, BS_y and BS_z can only be caused by BS_x, then there is an important sense in which they too are best characterized as knowledge-by-aquaintance. Perhaps it is best to say that BS_z is a derivative or second-order "knowledge-by-aquaintance" brain state. Knowledge-that, on the other hand, would be just propositional knowledge which can be acquired in a variety of ways.

The question of the causal origins of BS_z is partly empirical and partly a matter of where the burden-of-proof lies. The physicalist might maintain that there are many brain states which can only be caused by very specific other brain states. It is a matter of "hardware" in the central nervous system. It is fairly clear that BS_y is such a state, being caused only by BS_x. Otherwise Mary could well know what it is like to see red without ever having had her cone cells stimulated. (She could, no doubt, have seen red without chromatic stimuli if she could initiate BS_x along other than the normal causal channels.) There is nothing especially problematic with the claim that BS_z is such a state as well. Psychoneural reductivism relies heavily and unabashedly on an optimistic view of a completed neuroscience. When we have the empirical evidence it may well be clear that BS_z is the type of brain state which can only be caused by such things as BS_x. At any rate, the important point here is that, for Jackson's argument to work he has the burden-of-proof to show that BS_z is not, or cannot be, such a brain state. Without that premise, his argument that the physicalist cannot account for our intuitions concerning Mary simply fails.

A proponent of Jackson's view will, of course, have a different opinion about where the burden-of-proof lies. The intuitions of those in dualistic FP indicate that Mary could know everything physical there is to know about other peoples' brains and yet learn something new about their experiences of red when she first sees red for herself. It is not so much a matter of what knowledge is as a matter of what the knowledge is about. She knows all about their brains but not about their experienced qualia. Therefore, there is more to know about than what is physical.

One way for the materialist's counterargument to succeed is to make the assumption that knowing what it is like for others to see red is something that can only be caused by seeing red for oneself. But, knowing "that it is like so-and-so to see red" seems to be propositional knowledge or knowledge-that. If propositional knowledge can be acquired in any number of ways, then if the physicalist wishes to maintain that THIS particular kind of knowledge-that can only be gained through a causal chain beginning with retinal stimulation, then the burden-of-proof is on him.

Jackson focuses on the objects of our knowledge, what it is that we know, what knowledge is ABOUT. In the above discussion, the physicalist has been presented as concentrating on what knowledge IS, i.e., a brain state, rather than on what it is about. But another way that the materialist might argue is simply to maintain that, insofar as there is knowledge about anything, it is about what is physical. Hence, in Jackson's knowledge argument, if there is a way for Mary to

acquire all of the physical facts without ever experiencing chromatic stimulation, then premise (2) is simply false. If premise (2) is true, then premise (1) is false. This begins to sound more purely eliminativistic and evidences communication problems.

The debate between Jackson and the materialist exemplifies an interparadigmatic dispute between EM and a dualistic FP. The claim that Mary (before her release) does not know what it is like to see red is very much like an exemplar for dualistic FP. Jackson seems implicitly to recognize this. He says that "the knowledge argument is a valid argument from highly plausible, though admittedly not demonstrable, premises to the conclusion that physicalism is false"[18], and that his task "is to present an argument whose premises are obvious to all."[19] There is a reliance on convincing rather than proving.

This fits quite well if one views the dispute as resulting from a paradigm shift. It will be impossible to prove something from within a dualistic FP to someone in a materialistic paradigm, since there is insufficient common grounds for such proof. Further, the exemplar of a dualistic FP, e.g., knowing what it is like for someone to see red, is without force in an alternate paradigm such as EM. Insofar as one allows for partial intertheoretic reductions or "non-strict" identities, the "knowing what it is like to..." construction will be viewed as an archaic way of talking about brain states. There simply is no mysterious element (qualia) left unaccounted for. People's verbal behavior, e.g., saying "I now know what it was like all along

for others to see red", is only a (significantly misguided) report of the acquisition of a new type of brain state.

The above discussion indicates that EM and dualistic FP constitute two different paradigms, it also serves as an example of the advantages of construing EM in the broad sense operative here, rather than in a narrow or purely eliminative sense. Discourse between those in dualistic FP and those in the broad EM is easier, since they can share at least some of the same vocabulary and conceptual categories, e.g., "knowledge". However, insofar as `knowledge´ is construed as referring to a type of brain state, something must be done to work out the details and remove the vagueness of the central claims put forth by the materialist. This is, of course, the point of relying on a "completed neuroscience."

In the case of a pure eliminativism, the antecedent of "if it refers to anything then `knowledge´ refers to a brain state" is going to be denied. Once one drops FP, `knowledge´ no longer represents a working concept. In order to bring it back in, e.g., as referring to a type of brain state, it must evolve as a useful part of NS. The hard-core eliminativist will have difficulty communicating with Jackson, whose whole argument turns upon words and notions which are not in NS or EM and are thus meaningless from the pure eliminative materialist´s point of view. The difference between dualistic FP and materialism is thus more clearly a difference between paradigms the farther one moves toward a purely eliminative view.

Jackson's knowledge argument is presented as a problem for the materialist. However, problems are only well defined within the context of a paradigm, and what is a problem in one may not be in another. Strictly speaking, the knowledge argument is not even meaningful within a pure eliminativist's view of NS.

* * * * *

Concluding Remarks

At this point, it would be nice to be able to formulate a strategy for adjudicating between the paradigms of dualistic FP and EM. It turns out that this is a most difficult thing to do. Part of the problem lies in the fact that one must operate within some paradigm or other in order to make any evaluation whatsoever. Whether one arrives at a positive or a negative evaluation of some doctrine depends heavily upon the paradigm within which that evaluation is carried out. As far as is known, there is no way to work from a neutral point of view. For there is no (known) "super paradigm", no paradigm which contains all paradigms, and nothing analogous to "the set of all sets." Nor is it possible to formulate judgements while outside of any and all paradigms. (This is just an extension of the claim presented in chapter 4 that all judgements are theory-laden.)

However, there a few small things which can be done toward a relative appraisal of dualistic FP and EM. Most of us are able to work within either paradigm and, to some extent anyway, to shift back and forth between them. We can provisionally adopt one then the other and

try to work out their implications. We can identify the common ground between the two paradigms, then analyze and compare them relative to this common ground. For example, both dualistic FP and EM place a high value on internal consistency. So we can check each view for this trait. It is also possible to set aside misguided approaches to paradigm evaluation, e.g., those which involve the use of the criteria presented above for the evaluation of theories.

- evaluating paradigms: criteria for theory evaluation

When discussion turns to the topic of evaluating paradigms, one's initial inclination might be to try and use the same criteria that are used for theory evaluation. Unfortunately, these criteria are not all that helpful when we have moved to the level of paradigms. As discussed in chapter 2, these criteria are intimately related. Virtually any argument will involve more than one criterion. For example, the argument from ontic simplicity also involves explanatory power. (Given the relationship between explanation and prediction, it will involve predictive success as well.)

However, Ockham's razor can be legitimately applied only if one thing will do in place of two. That, of course, is just what dualistic FP claims is wrong with materialism. The materialist just cannot adequately account for what, within dualistic FP, is called the qualitative character of perceptual experience. Excruciating pain, or seeing a brilliant shade of red just is not addressed by EM or NS. Simplicity of ontology may well be a good criterion for judging between

theories. However, it is useable for judging paradigms only insofar as the two paradigms under consideration agree on what there is to explain. From the point of view of dualistic FP, NS fails even to address, much less explain, what is the most obvious fact about human existence, namely the (special nature of the) qualia of perception. Hence, from the point of view of dualistic FP, the ontology of NS is not just simple, it is inadequate. It fails to recognize, much less explain, something about which we must be concerned.

Consider the following as an example of this form of interparadigmatic dispute concerning what requires explanation. Churchland claims that, since no particular quale is required for the inference of a given objective feature of the world, since any objective feature can be conceptually tied to any experience, not only will one do as well as another but they might be eliminable altogether. The view expressed here is that qualia, if they do exist, are just unnecessary "middle-men" in the process of perception. As long as the relevant causal connections remain intact, these qualia might well be eliminated altogether.

This is an argument for the possibility of the elimination of experience itself. This argument is plausible only insofar as one is already sympathetic to EM. If everything that there is to explain is physical, e.g., human behavior, then talk about qualia may well be dropped. However, within dualistic FP the qualia themselves are one of the main things to be explained. It really does not matter much if they are mere middle-men. Even if they are mere epiphenomena, they

still cannot be ignored if they are known to exist, which they are according to the dualist.

(One further comment should be made here, concerning this argument of Churchland's. Regardless of which of the two paradigms one accepts, the argument that since no particular qualia is required, they might all be dispensed with, is fallacious. Consider the following analogy.

Let x = "it is raining"

y = "the sprinklers are on"

z = "the temperature has fallen to the dew point"

w = "the sidewalks are wet"

If we assume that x, y, and z are each individually sufficient for w, then none of them is individually necessary for w. Any one will do as well as another. However, assuming nothing else sufficient for w, we would not claim that x, y and z are all eliminable and it still be possible that w. From the fact that no particular qualia is required for our perceptual judgements, it does not follow that no qualia whatever is required. Since this is not a particularly elusive point, perhaps the best way to interpret Churchland's move is to view the argument as one which is presented in the context of a paradigm where qualia are not some special feature of the world which need explaining.)

The fact that there is a fundamental disagreement concerning what there is for us to explain (and predict) is very closely related to the notions of observation, testability and falsifiability. As they relate

to paradigms, all of these topics are intimately tied to the notion of strong ontic commitments. Falsification requires testing. Testing relies on observation and judgements concerning those observations. How one interprets one's observation will depend on the paradigm one accepts. Of particular importance here is the claim that might be put forward by the proponent of dualistic FP, that each and every observation constitutes a refutation of EM, since observations involve the phenomenal qualia of perceptual experience. The strong ontic commitments of a paradigm can thus play a crucial role in testing and falsifying. Hence, the testing of paradigms will involve a certain circularity since the interpretation of the test will depend upon the paradigm.

Consider the difference between testing a theory and testing a paradigm. There is a sense in which both FP and NS as theories are testable. Certainly many, if not all, of the hypotheses within them are testable. There is relatively little dispute over the testability of hypotheses within NS. Within FP, I might formulate the hypothesis that Jones took his umbrella with him today because he believed that it was going to rain. Given that hypothesis, and a few other assumptions, I would predict that Jones will say "Yes" if I ask him if he thinks it will rain. I can thus test the hypothesis by asking him about his beliefs and observing his response.

Thus theories can be tested according to the hypothetico-deductive method. Given that excellence of a theory is the measure of ontology, both the theory and its ontic commitments can be justified. However,

it is clear that what is getting justified are weak ontic commitments. Theories imply weak ontic commitments, due to quantification over their indefinite singular terms denoting types-of-things in the world. The situation is different when we consider EM and FP as paradigms with their associated strong ontic commitments.

- paradigms and strong ontic commitments

The problems with testing a paradigm run very deep indeed. Testing relies heavily on observation. Within EM, there is no distinction between observation terms and theoretical terms, since all observational judgements (as reported through the use of observation terms) are theory-laden, or in this case paradigm-laden. If we shift from a dualistic FP to EM (and NS), we will stop judging ourselves to be in pain and start judging that our c-fibers are firing. Hence, the observation of our introspective inner states may not serve to support dualistic FP over EM since those internal states will be judged to be qualia or neural events depending upon the theory, and paradigm, to which one subscribes.

From the point of view of dualistic FP, the very act of observing serves to deny the truth of EM. Each time we observe anything we have a mental experience, since an observation is a mental event. Hence, every observation serves to falsify EM. Of course, for the eliminative materialist none of this is at all persuasive. An act of observation just is a physical event involving physical stimuli from the outside

world impinging upon one's sensory surfaces. If a judgement is anything at all it is a neural event.

The truly nasty part about the strong ontic commitments of paradigms is that there seems to be no support for them. What support there is involves an undesirable circularity. Justification of the weak ontic commitments of such theories as FP and NS through empirical testing will not answer any questions concerning the ultimate nature of such things as beliefs or neurons. If there is no way to verify that something is mental, or that it is physical, then there is a very serious question concerning the usefulness (and meaningfulness?) of such notions. Shades of logical positivism!!

Dualistic FP: Even so, as mentioned earlier, there does seem to be something indubitable about our knowledge of the existence of that which we would call, within dualistic FP, a sensory experience. When Macbeth saw his dagger_i there was no doubt about its existence_i despite any firm convictions about the nonexistence of the dagger_e. On the other hand, even in the case of first-person accounts of seeing something_i we are faced with the problem of interpretation. Even at the level of ultimate types-of-things judgements are involved, not judgements concerning existence_i, but judgements concerning the type-of-thing that the something_i happens to be.

From the point of view of dualistic FP, this indubitability of existence_i may well be the "mark of the mental." Our knowledge of the (physical) world_e, as traditionally characterized, is fallible. This

includes such basic knowledge as that of the mere existence of something$_e$. There are many avenues of support for this view. Dreams, hallucinations and illusions, thought experiments such as Descartes' "evil spirit" and more contemporary mad-scientists implanting electrodes in the brain, and real-life scientists, e.g., W. Penfield, actually stimulating the brain with microelectrodes and causing such things as the smell of a rose$_i$.

It makes little difference whether such knowledge is based upon direct sensory experience or upon extensive theorizing. Knowledge of the external world, even of the mere existence of things$_e$ out there, is fallible. It is clear, at least to the dualistic folk psychologist, that there is a striking contrast between such fallible knowledge of things$_e$ out there and of things$_i$ "in the mind." This point will retain its force even though "in the mind" is a judgement that may well be fallible. It will lose its force if "indubitable knowledge" is thought to be just another type of brain state.

Eliminative Materialism: EM is in no better shape relative to strong ontic commitments than is dualistic FP. The hypothesis that everything is physical is no more testable or falsifiable than is the claim that non-physical qualia exist. Physics and chemistry test for many things. But they do not, indeed cannot, test for the physicality of anything. We can test the claim that water is H_2O, and that H_2O can be separated into hydrogen and oxygen by an electrical current. Such tests involve weak ontic commitments and the (law-like relationships

existing between the) conceptual categories of theories. But how would one check to see that H_2O really is physical?

There are two ways that one might interpret materialism. It might be a view which uses "matter", or "physical substance" or "material bodies" as mere short-hand terms to denote all of the types-of-things countenanced by the (physical) sciences. If this is what is maintained, then EM makes no strong ontic commitments at all. All that is being claimed is that we should accept the "best" theory available, that that theory is a scientific one, and that we should thus accept the (weak) ontic commitments as implied by those theories.

Essentially, all that one would be maintaining on this interpretation of materialism is that whenever we say anything about anything, we can always add "and it is physical". Whenever we attach a predicate to the variable "x", we can add "& Px". There is something distinctly vacuous about such a view. Besides, there is something compelling about the notion that the only thing common to everything that exists is existence. That would seem to be one of the reasons why we no longer treat existence as a predicate. It already applies to everything. "Is physical" would be the same as "exists."

The second possible interpretation, the one that seems more commonly made, is that materialism is a position which does make positive strong ontic commitments to the physical and only to the physical. This goes beyond the acceptance of the weak ontic commitments implied by scientific theories. It makes a claim concerning the ultimate or basic nature of all of the particular things

that are actually referred to by the type-words of any theory whatsoever. (Insofar as FP terms actually refer, what they refer to is something physical.) Such a claim is not a scientific one. It is not something that can be tested, verified or falsified. Nor is it the case that a theory such as NS implies the existence of physical matter in this strong sense.

Construing "being physical" as existing in space will not help much. (Consider the views of Berkeley and Kant.) From the epistemological side, verification that something is in space will require sensory experience and judgements concerning those experiences (even granting some physicalistic view of "experience"). Further, attempts to define "located in space" will almost surely rely on the notion of being physical, since the two are so closely related. It would seem that "x is located in space" (or "x is physical") is on equal footing with "x is a mental experience".

- eliminative neuroscience

What then can be said about EM as a paradigm? Recall the following points which have been made earlier in this work.

1.) Virtually all of the terms of a theory specify types. It is the willingness to apply such terms to quantified variables which results in one's (weak) ontic commitments concurrent with the acceptance of some theory.

2.) The main arguments for the rejection of FP in favor of NS involve the superior explanatory power and predictive success of the latter and the better integration and synthesis of NS with the bulk of scientific theories.

3.) The main arguments for elimination of FP rather than reduction appeal to such things as undue top-down constraints.

4.) The best measure of (weak) ontology is excellence of theory. If we accept NS, we must accept an ontic commitment to such types-of-things as c-fibers, action potentials and axons. In accepting other scientific theories, we will also be committed to the biochemical substances which constitute neurotransmitters, etc. We will also be committed to molecules, atoms and sub-atomic "particles". But we need not be committed to "matter" per se.

5.) None of the arguments in favor of NS over FP can be used in support of materialism. NS, together with its conceptual categories, may well be the stronger theory. However, the conceptual categories, and the type-words which are their linguistic representatives, specify only WEAK ontic commitments. There is nothing to support the strong ontic commitments of the materialist. Even if excellence of theory is the best or only measure of ontology, one is not justified in accepting a materialistic monism based on the success of NS.

Given the above, it is reasonable to maintain that "eliminative materialism" is something of a misnomer. To be more precise, if a proponent of the eliminativistic approach wished to argue for the

displacement of FP by NS, then what should be adhered to is a position best described as eliminative neuroscience. Materialism, per se, is a view which makes an unsupported, and perhaps unsupportable, strong ontic claim. Nothing much is gained by such a claim and, in its interesting sense, the term `materialism´ may not even be meaningful.

* * * * *

* * * * * * * * *

ENDNOTES

Chapter 1

1. By even numbers it is meant the abstract entities, if there are such, not the individual numerals such as "4" which may be instantiated in a variety of mediums including, I suppose, pigskin.
2. A classic formulation of this kind of substance dualism comes from Descartes, particularly his Meditations. See The Philosophical Works of Descartes, Vol.1, Haldane, E.S. and Ross, G.R.T., 1981, Cambridge University Press.
3. For an interesting article here see D. Dennett, "Where Am I" found in his Brainstorms, 1981, MIT Press, Cambridge, Mass.
4. To get just a hint of the naivete of this classic formulation, and to get a feel for some of the contemporary scientific research which is relevant, see Patricia S. Churchland's "Consciousness: The transmutation of a concept", Pacific Phil. Quarterly, Vol.64, 1983, pp. 80-95.
5. ibid.
6. For background reading on some of the traditional arguments against interactionism, and against dualism generally, see for example, R.J. Hirst, The Problems of Perception, 1959, Humanities Press, and J. Shaffer, "Mind-Body Problem", in the Encyclopedia of Philosophy, 1967, Macmillan Pub. Co., Vol.5, pp. 336-346.

7. This problem takes on an intriguing new twist when one considers the quantum theoretic models of contemporary physics. In quantum theory the role of the observer is crucial. The observer seems always to be treated in the old Newtonian-mechanistic manner, generally with a significant overlay of mentalism. Quantum wave functions are disrupted when an mind acquires an appropriate type of experiential knowledge. Someone looks up into the night sky and sees a star, thus disrupting a photon quantum wave function. See Taking the Quantum Leap by F.A. Wolf, 1981, Harper & Row.
8. One could not, however, maintain that minds simply are part of the physical system and still be consistent with substance dualism.
9. For a reductivistic version of emergentism see M. Bunge, Scientific Materialism, 1981, D. Reidel Pub. Co., Dordrecht, Holland.
10. F. Jackson, "Epiphenomenalism".
11. For an interesting review of many such cases see P.S. Churchland, 1983, op.cit.
12. Note how this line of reasoning could be extended to support the thought that not all humans have minds. Minds, like blue eyes, might be instantiated in only part of the population. If mindedness were an epiphenomenon then, by definition, there would be no behavioral test by which those with minds could be distinguished from those without.
13. This is not to be confused with the religious view which stands in contrast to evolutionary theory.

14. Shoemaker, S. "The Inverted Spectrum", Journal of Philosophy, Vol. LXXIX, No. 7, July 1982, pp. 357-381.

15. Nagel, T. "What is it like to be a bat?", Philosophical Review, No. 83, 1974, pp. 435-450.

Chapter 2

1. Haldane, E.S. and Ross, G.R.T. trans., The Philosophical Works of Descartes, 1982, Cambridge Univ. Press, Vol. 1, Prin. VIII, P. 221 and Prin. XI, P. 223.
2. Matter and Consciousness, Paul M. Churchland, 1984, (2nd. ed. 1988), The MIT Press., P. 56. (Hereafter `M&C')
3. See for example: Churchland, P.M. Scientific Realism and the Plasticity of Mind, Cambridge Univ. Press, Cambridge, 1976 (Hereafter `Plasticity'); M&C; Churchland, P.S. Neurophilosophy, The MIT Press, 1986, Cambridge, Mass.; W.V. Quine, "Epistemology Naturalized", found in Ontological Relativity and Other Essays, 1969, Columbia Univ. Press, N.Y. Skinner, Beyond Freedom and Dignity, 1971, Random House, N.Y.
4. Quine, W.V., 1969, op. cit.
5. Philosophical Foundations of Physics, 1966, Basic Books, Inc. Or see Scientific Knowledge, J.A. Kourany ed., 1987, Wadsworth, Inc., pp. 122-138.
6. See for examples: K. Popper, Conjectures and Refutations: The Growth of Scientific Knowledge, 1963. P.M. Churchland, Plasticity and M&C. W.V. Quine and J.S. Ullian The Web of Belief, Ch. VI, 1978, 2nd. ed. Random House, N.Y.
7. This seems so very simple and straightforward. However, there are some very interesting and complex problems lurking in the background. The main difficulty lies in the fact that the theory

that one is attempting to justify may itself be involved in establishing the correctness of the predicted observation statement. Not only is the perceptual experience associated with the observation influenced by the theory, but the meanings of the words in the observation statement are also largely determined by the theory within which they are embedded. More will be said about this later when the network theory of meaning is discussed.

8. Churchland, P.S. 1986. op. cit. P. 279.

9. 1978, op. cit.

10. In certain cases, an instrumentalistic view seems quite reasonable. Even if one were a realist relative to current and resistance, one would likely view amps and ohms as "useful fictions". One need not maintain that an amp, per se, actually exists, except perhaps as a real unit of measurement. In fact, it may be the case that one cannot reasonably maintain that amps exist. It is a unit of measurement which we invent and use because it is useful. We could have just as easily invented a "schmamp" (equaling 1.263 amps) which would be very useful if we had the desire to work with complex mathematical formulae. But now either schmamps are real only as units of measurement, or they become real "things" as soon as we invent the terms (hence we create them), or they were always really out there and we now recognize them (in which case everything imaginable already exists). It is pretty clear that the first alternative is the most reasonable to accept.

11. For an explanation see Quine, W.V., 1960, Word and Object, M.I.T. Press, Cambridge, Mass., P. 112.
12. See especially "On what there is", Review of Metaphysics, 1948, Vol. 2, P. 32.
13. Word and Object, P. 242.
14. Psychosemantics, 1987, The MIT Press, Cambridge, Mass. See especially the Preface for an interesting account of explanation within this theory.
15. Plasticity
16. M&C
17. Skinner, B.F. 1971, op. cit.
18. Found in The Complete Plays of Aristophanes, M. Hadas, ed. 1962, Bantam Books, New York. P. 161.
19. In the example, `irritable' and `pain' are theoretical terms, `grimace' and `hitting one's thumb with a hammer' and `bodily injury' are observation terms; 2. is a theoretical law, 1. and 3. are composites functioning as correspondence rules for Carnap, 4. is perhaps best viewed as an empirical law, 5. is a singular statement of fact, as is 6. which is the prediction.
20. Plasicity, M&C, and Stich, S. From Folk Psychology to Cognitive Science, 1983, MIT Press, Cambridge, Mass.
21. See Plasticity P.2 for a presentation of this view.
22. Rorty, R. "Mind-body identity, privacy and categories", 1965, in C.V. Borst, op. cit.
23. Plasticity, and M&C.

Chapter 3

1. Neurophilosophy, op. cit., pp. 278-279.
2. M&C and Plasticity.
3. Propositional attitudes are those (mental) states which express a certain attitude toward a proposition. For example, "fear", "believe", and "desire" are all terms expressing propositional attitudes, since one might fear (or believe or desire) that the stock market will soon collapse. "The stock market will soon collapse." is the proposition that one has a certain attitude toward. For a further explanation of this topic see M&C.
4. U.T. Place, "Is consciousness a brain process?", The British Journal of Psychology, XLVII, 1956. Also found in C.V. Borst, ed. The Mind/Brain Identity Theory, St. Martin's Press, N.Y., 1975, pp 42-51.
5. J.J.C. Smart, "Sensations and brain processes", The Philosophical Review, LXVIII, 1959. Also found in C.V. Borst, ed., ibid., pp 52-66.
6. D.M. Armstrong, "The nature of mind", Inaugural lecture of the Challis Professor of Philosophy at the University of Sydney, 1965. Also found in C.V. Borst, ed. ibid., pp67-82.
7. J.J.C. Smart, op. cit.
8. J.J.C. Smart, ibid.
9. See "Empiricism, semantics, and ontology", found in Meaning and Necessity, 2nd. edition, 1956, The Univ. of Chicago Press, pp. 205-221.

10. See "On Carnap's views on ontology", found in The Ways of Paradox and Other Essays, 2nd. edition, 1976, Harvard Univ. Press, pp. 203-211.

11. U.T. Place, op. cit.

12. Although I do not wish to argue here for the claim here that all predication is the specification of relationships, it seems clear that many predicates do indeed pick out relationships. Even one-place predicates often are relational. If one asks what is meant by "x is red", the way to answer is to specify the relationship between x and other colored objects (some red and some not), with things like spectrophotometers and human sensory systems. The tie in with the weak sense of ontic commitment should be clear. The categorization of things in the world involves the specification of types, which in turn involves comparing things and thus establishing relationships.

13. This is the first introduction to a central and very thorny problem. Partly it is a matter of ontology, partly of semantics, and largely of the theory that one maintains. Also involved is the notion of a judgement, e.g., judging that one is indeed in pain. A full discussion of this topic will not come until later. However, you the reader are urged to keep this issue "in the back of your mind," as it will crop up again and again. It really is the main point of attack by the dualist/mentalist against the materialist.

14. The hedge here will be explained in chapter 5.

15. Leaving the notion of "wrong" vague at this point.

16. For a discussion of these topics see The Tao of Physics, F. Capra, 1975, Bantam ed. 1980.

17. For a general discussion of metaphysical functionalism see N. Block, "Introduction: What is functionalism?", found in N. Block, N., 1980, op.cit.

18. Actually, just which creatures it rules out will depend somewhat on the way that the reduction is set up, that is on the bridge laws and on certain empirical matters of fact. For example, if psychoneural reductivism were to maintain that pain just is the firing of C-fibers, then any creatures not endowed with such fibers could not possibly be in pain. The objection posed by the functionalist is that we should not rule out the possibility of creatures very different from ourselves instantiating such states as pain, belief, fear, etc. At least we should not rule out this possibility a priori and by definition. It should remain an empirical question whether or not a given creature can experience pain. In each individual case, a judgement will be based upon the behavior of the creature and whether or not there is a functional state of its physical system which fulfills the role that the firing of C-fibers plays in our system.

19. For an analysis of mental states in terms of causal roles, see two articles by David Lewis, "Psychophysical and theoretical

identifications" and "Mad pain and martian pain". Both can be found in N. Block, 1980, op.cit.

20. This view is very close to the view of the "functional state identity theorist" as described by N. Block in "What is Functionalism", op.cit.

21. To be accurate, "second-order" should be replaced by "higher-order" and "first-order" by "lower-order", since it is not at all clear that a c-fiber firing is a first-order predicate. The main point however is that functional state identity theory places mental events at a meta-level relative to psychoneural reductivism.

22. N. Block´s article; op.cit.

23. As described by N. Block, op.cit.

24. This is closely related to the "functional specifier" view in Block´s article, op.cit.

Chapter 4

1. It should be noted, however, that there is nothing particularly inconsistent with accepting much of what the functionalist wishes to maintain. As generally characterized, functionalism is a special version of type/type identity theory and a psychofunctional reductivism generally goes right along with it. However, as mentioned in chapter 3, an eliminativistic formulation is possible, since what one would be interested in would be the construction of a theory whose conceptual categories stressed causal/functional roles and which was consistent with scientific theories generally.
2. See M&C and Plasticity.
3. M&C, Plasticity, etc.
4. M&C
5. M&C P.43
6. Plasticity, P.7
7. Plasticity, P.7
8. Plasticity, pp.14-15
9. Plasticity, P.16
10. Plasticity, P.25
11. Plasticity, P.16
12. Plasticity, P.14
13. This seems to parallel the traditional distinction between seeing and seeing-as.

14. M&C P.80

15. Plasticity, P.25

16. This is one formulation of what may be termed the "active perception thesis".

17. Plasticity, pp 16-25

18. M&C P.56

19. M&C P.56

20. Plasticity, P.61

21. For a general discussion of this topic see Churchland's Plasticity, sections 2 and 3. Also, refer back to chapter 2 of this work.

22. The important implication here is that our perceptual judgements, and the observation terms used to express those judgements, will no longer have any special epistemic status. This is just the point of interface between the thesis that perception is an active affair and the network theory of meaning.

23. M&C P. 63

24. M&C P. 60

25. This is one of the points made by Jerry Fodor in his article "Observation Reconsidered."

26. In essence, this is just a particular version of the view that languages are public affairs and hence the meanings of terms must have an objective source.

27. Plasticity, p.13

28. Plasticity, P.15

29. Plasticity, P. 37

30. Plasticity, P. 40

31. Plasticity, pp 38-39

32. Plasticity, P. 15

33. Plasticity, P. 15

34. M&C, pp. 73-81

35. M&C, pp. 73-74

36. Ch. 2, sec.5

37. Plasticity, P. 44

38. M&C, P. 60

39. Plasticity, P. 83

40. M&C, P. 48

41. M&C, pp. 44-45

42. "Mind-body identity, privacy, and categories", found in Borst, C.V. 1970, op. cit., P. 190.

43. Rorty, ibid. P. 204

44. Plasticity, P. 34

45. Plasticity, P. 114

46. Plasticity, P. 24

Chapter 5

1. See Ch. 4, sec. 3.
2. For an indication of how this move might help the materialist, see the discussion below of a physical definition of qualia and of Mary and her brain states.
3. See Ch. 4, sec. 3.
4. Plasticity, P. 11.
5. Plasticity, P. 9.
6. George Berkeley, Essay Towards a New Theory of Vision, paragraph no. 47. See for example, Berkeley: Works on Vision, C.M. Turbayne, ed. 1963. Bobbs Merrill, N.Y.
7. For a discussion of weak and strong ontic commitments, see Ch. 2.
8. See for example: Richard Rorty, "Mind-body identity, privacy, and categories", op. cit., M&C and Plasticity.
9. "What is it like to be a bat?", in Ned Block's Readings in the Philosophy of Psychology, Vol. 1, 1980, pp. 159-168.
10. Austin Clark, "A physicalist theory of qualia", The Monist, Vol. 68, No. 4, Oct. 1985, pp. 491-506. See especially pp. 504-505.
11. "Epiphenomenal qualia" The Phil. Quarterly, Vol. 32, 1982., pp. 127-136.
12. "What Mary didn't know", The Journal of Philosophy, May 1986, Vol. 83, No. 5, pp. 291-295.
13. Ibid.
14. Op. cit. 1982, P. 130

15. In "What is it like to be a bat?", op. cit.

16. Jackson, 1986, op. cit. P. 293

17. ibid. 1986, P. 294

18. ibid. P. 295

19. Jackson, op. cit., 1982, P. 128