



1973

Mind-Body Identity, Incorrigeability and Conceptual Revision

Gordon D. Cohen
Loyola University Chicago

Recommended Citation

Cohen, Gordon D., "Mind-Body Identity, Incorrigeability and Conceptual Revision" (1973). *Master's Theses*. Paper 2689.
http://ecommons.luc.edu/luc_theses/2689

This Thesis is brought to you for free and open access by the Theses and Dissertations at Loyola eCommons. It has been accepted for inclusion in Master's Theses by an authorized administrator of Loyola eCommons. For more information, please contact ecommons@luc.edu.



This work is licensed under a [Creative Commons Attribution-NonCommercial-No Derivative Works 3.0 License](https://creativecommons.org/licenses/by-nc-nd/3.0/).
Copyright © 1973 Gordon D. Cohen

MIND-BODY IDENTITY, INCORRIGIBILITY

AND CONCEPTUAL REVISION

Gordon D. Cohen

1973

Submitted in partial fulfillment
of the requirements for the Degree
of Doctor of Philosophy in Philosophy,
Loyola University of Chicago

TABLE OF CONTENTS

CHAPTER		PAGE
I	INTRODUCTION.	1
II	LAWS AND SENSATIONS	23
III	SENSATION-EXPLANATIONS AND CONCEPTUAL REVISION	59
IV	INCORRIGIBILITY AND MIND-BODY IDENTITY.	106
V	INCORRIGIBILITY AND THE EEG ARGUMENT.	155
	SELECTED BIBLIOGRAPHY	181

CHAPTER I

INTRODUCTION

Classical materialism maintained that a man was nothing over and above a vast arrangement of particles in motion. Modern materialism differs from classical materialism in refusing to specify the kind of entities which ultimately make up the furniture of the world, including man. These entities may turn out not to be particles or anything like particles. It is the science of the future to which we must look for the answer to this question. But modern materialism agrees with classical materialism in the prediction that everything in the world, including the behavior of man, will be completely explicable in terms of entities which fall wholly within the province of physics. Because of this emphasis on the idea that we must await the outcome of scientific inquiry, perhaps in some far distant future, to know what man is, modern materialism is sometimes called scientific materialism. Professor J.J.C. Smart, a leading exponent of scientific materialism, confesses his faith, in the following well-known passage, in the power of science finally to come up with the definitive answer to this question which has perplexed philosophers from at least

the time of Socrates:

It seems to me that science is increasingly giving us a viewpoint whereby organisms are able to be seen as physicochemical mechanisms: it seems that even the behavior of man himself will one day be explicable in mechanistic terms. There does seem to be, so far as science is concerned, nothing in the world but increasingly complex arrangements of physical constituents. All except for one place: in consciousness [S]ensations, states of consciousness, do seem to be the one sort of thing left outside the physicalist picture, and for various reasons I just cannot believe that this can be so. That everything should be explicable in terms of physics . . . except the occurrence of sensations seems to me to be frankly unbelievable.¹

It should be clear from this passage that Smart would consider the scientific materialist to have made out his case once he has shown that the occurrence of sensations is explicable in terms of physics. This might seem at first glance a strange position to take, if not a disingenuous one, in view of the serious objections which have been raised to any form of materialism by Thomistically oriented philosophers and contemporary phenomenologists with respect to thoughts and their intentionality. It is nevertheless an understandable position for those philosophers who, following Descartes, reject anything as a criterion of the mental, such as intentionality, which might serve to exclude

¹"Sensations and Brain Processes," Philosophical Review, vol. 68 (1959), p. 142.

sensations from the realm of the mental. Most defenders of scientific materialism, in fact, follow Descartes on this point. In any case, the issue which is currently at the center of controversy between scientific materialists and their opponents concerns the status of sensations rather than thoughts.

The attempt to show that the occurrence of sensations is explicable in physicalistic terms has consisted in recent years in the defense of what is now simply called the Identity Theory. The central claim of this theory is that it makes sense to assert that empirical inquiry will discover that sensations are identical with certain brain processes. The theory does not claim that empirical inquiry will discover that sensations are identical with certain brain processes but only that it makes sense to assert that it will. This is in line with the previously mentioned hesitation to speculate now upon the outcome of future empirical inquiry concerning the nature of man. But, in view of such diffidence, why should it be thought important to establish the meaningfulness of asserting that empirical inquiry will discover that sensations are identical with certain brain processes? The answer is that materialists cannot be

satisfied with a mere correlation between sensations and physical states or processes. There is a sense in which some forms of relation which have been thought to hold between sensations and physical states or processes make the occurrence of the former explicable in physicalistic terms. Such is the case with epiphenomenalism, which maintains that the occurrence of sensations can be wholly accounted for by occurrences within the body. Such forms of relation do not help the materialist, however, since he wants to hold that the occurrence of sensations is explicable in physicalistic terms in the sense that the former is nothing over and above the latter. His position is, therefore, that sensations are identical with certain physical states or processes in some sense of "identical." Moreover, for analogous reasons, materialists cannot be satisfied with a mere correlation between properties of sensations and properties of those physical states or processes with which sensations have been identified, as would be maintained by the double-aspect theory. Does this mean that the corresponding properties must also be held identical in some sense of "identical?" There seems to be no general agreement among materialists

on this point. Indeed, the problem of what to do with properties has proved a most difficult one and has occasioned some of the subtlest reasoning on this topic.

Now there is a straightforward sense of "identical" in which sensations have been claimed to be identical with certain physical states or processes. This is the sense which expresses the relation of "strict identity," that is, the relation such that

$$(x)(y)((x = y) \supset (F)(Fx \equiv Fy)).$$

Thus, if a given sensation is strictly identical with (say) a certain brain process, then all (nonintentional and non-modal) properties truly predicable of the sensation are truly predicable of the brain process, and conversely. On the face of it, however, strict identity cannot hold between sensations and brain processes, or, for that matter, between sensations and any other physical states or processes. And this "cannot" is logical. It is not simply that some properties truly predicable of a sensation are not in fact truly predicable of any physical state or process, or conversely. Rather, some properties truly predicable of the one are apparently not predicable at all of the other. For the predicates

which express these properties seem to be of a different logical type than the expressions which refer to the other. They seem to belong to different "logical categories" or to different "logical spaces" or to different "language-games." Thus, while we certainly can say that a pain, for example, is throbbing or becomes more or less intense, it would seem to be not merely false but, in a certain sense, senseless to apply these predicates to physical processes such as brain processes. Again, while a brain process can be correctly said to be occurring three feet from the kitchen table, to apply this predicate to the experience of having a pain would seem to result in a certain kind of nonsense. Further, the relation of identity between sensations and brain processes would have to be contingent, since (to mention only one reason) a person who reports sensations need know nothing about brain processes. But contingent identity can hold only between terms which are independently identifiable, and this seems to imply that sensations must have some properties that are not reducible to physicalistic properties. In view of considerations such as these, philosophers, especially so-called "linguistic" philosophers, have charged materialists with committing "category mistakes" or with being guilty of

"conceptual confusion." Materialists, accordingly, have thought it important in recent years to defend the view that it is sensible to assert that empirical inquiry will discover that sensations are identical with certain brain processes, however diffident they may be about what empirical inquiry will in fact discover.

It may be thought that the shortest way with objections of this kind is to argue that sensations are identical with certain brain processes in a sense other than that of strict identity. But this is not the case. For it is difficult to see what other sense of "identical" there can be which does not reduce simply to mere "correlation." For this reason, some philosophers who defend the Identity Theory claim that it is the sense of strict identity in which sensations are identical to certain brain processes, but then go on to argue that this claim does not involve one in category mistakes or conceptual confusion. Smart is one of the most notable of such philosophers in recent years. Smart argues that sensation reports are neutral between psychic and physicalistic logical categories because such reports are simply classifications of sensations in terms of bare

similarities.¹ Thus, "I have a pain," according to Smart, is roughly equivalent to "What is going on in me is like what goes on in me when a pin is stuck into me." The analysis in this case simply reports similarities without saying wherein these similarities consist. One focus of controversy over this translation version of the Identity Theory concerns the adequacy of such proposed translations. It has been objected, for example, that these rough equivalences cannot be transformed into strict equivalences without a corresponding loss of their neutrality.² Richard Rorty, another well-known defender of the Identity Theory, regards objections of this kind as ineluctable for any version of the Identity Theory which claims strict identity between sensations and brain processes and so has chosen the more difficult task of trying to make clear a sense of identity which is neither strict identity nor a relation that reduces to mere correlation and which could sensibly be asserted to hold

¹See ibid., pp. 149-50. The version of the Identity Theory which Smart defends was originally put forward by U. T. Place ("Is Consciousness a Brain Process?," British Journal of Psychology, vol. 47 (1956), pp. 44-50), but it is Smart's article that has received most attention.

²See Kurt Baier, "Smart on Sensations," Australasian Journal of Philosophy, vol. 40 (1962), p. 66, n. 23, and James W. Cornman, "The Identity of Mind and Body," Journal of Philosophy, vol. 59 (1962), pp. 489-90.

between sensations and certain brain processes. In the course of the article in which Rorty introduces his own theory,¹ Smart's theory is summarily discussed and dismissed by appeal to the difficulty of providing adequate translations. But even in this early article, it is clear that Rorty's disagreement with Smart runs much deeper. The basic issue dividing them is not so much the possibility of providing adequate translations as it is the possibility of reporting bare similarities. The latter seems to imply that there is an activity which can reasonably be called "awareness" prior to the learning of language. Rorty, however, wishes to ally himself with Wittgenstein and many of his followers in the battle against such a notion of awareness--a battle, indeed, which he takes as having already been won. More specifically, what Wittgenstein and his followers have argued and Rorty accepts is that the possibility of picking out something as an object of awareness requires that it be identifiable as an item of some definite sort and that the classification of items just consists in the classification

¹"Mind-Body Identity, Privacy, and Categories," Review of Metaphysics, vol. 19 (1965), pp. 24-54.

of expressions for them according to the conceptual scheme embedded in a language. Expressions for sensations, accordingly, must be tied up logically, that is, in conformity with the rules of the conceptual scheme, with other sorts of expressions if the former are to be referring expressions at all.

Now this view would seem to be inconsistent with the claim that it makes sense to assert the identity of sensations and brain processes. For if there is no possibility of reporting bare similarities, the need for providing independent identifications of both terms between which the contingent identity relation is supposed to hold can apparently be satisfied only by the recognition of irreducibly psychic properties. And then it is not easy to see how the Identity Theorist can escape the charge that he is committing "category mistakes" or is guilty of "conceptual confusion." Rorty proposes a way out for the Identity Theorist who sympathizes with Wittgenstein's attack on the notion of pre-linguistic awareness. We can simply change the conceptual scheme which is embedded in the language we speak.

[T]he classifications of linguistic expressions that are the ground of [the Identity Theorist's] opponents' criticism are classifications of a language which is

as it is because it is the language spoken at a given stage of empirical inquiry. But the sort of empirical results that would show brain processes and sensations to be identical would also bring about changes in our ways of speaking. These changes would make these classifications out of date There is simply no such thing as a method of classifying linguistic expressions that has results guaranteed to remain intact despite the results of future empirical inquiry.¹

Once these changes are made, we can go on (sensibly) to assert that sensations are identical with certain brain processes, meaning by "identical" here "the sort of relation which obtains between, to put it crudely, existent entities and non-existent entities when reference to the latter once served (some of) the purposes presently served by reference to the former--the sort of relation that holds, e.g., between 'quantity of caloric fluid' and 'mean kinetic energy of molecules.'"² To put it less crudely, Rorty is proposing that the statement

Sensations are identical with certain brain processes
should be analyzed as

What people now call "sensations" are identical with
certain brain processes

where the word "identical" in the analysans, but not in

¹Rorty, ibid., pp. 24-25.

²Rorty, ibid., p. 26.

the analysandum, signifies the relation of strict identity. The reason why the word "identical" in the analysandum cannot signify the relation of strict identity is that the expression being used for what is asserted to be identical with certain brain processes does not belong to the same logical category to which some predicates true of brain processes belong. The analysans neatly avoids this difficulty by embedding the offending expression in a context in which it occurs non-referentially. Although the word "sensations" does occur in the expression used to refer to what is asserted to be identical with certain brain processes, "sensations" is not itself being used to refer; it is not being used at all but only mentioned. And so far as the expression which is being used to refer is concerned, viz., "What people now call 'sensations,'" there seems to be no reason why it should not be in the same logical category as any predicate true of brain processes.

As subtle and ingenious as this strategy is, it is a strategy, I think, for losing the war in which Wittgenstein and some of his followers were engaged when, according to Rorty, they won the battle against the notion of pre-linguistic awareness. Rorty admits, and indeed it is

difficult to deny, that the discourse in favor of which sensation-discourse is to be eliminated would serve only some of the purposes presently served by the latter. That brain-discourse would not serve, at least initially, the expressive purposes served by sensation-discourse goes without question. Such purposes, however, are not what is at issue here. In any case, they could be provided for in other ways if necessary. More to the point, sensation-discourse tends to give rise to certain kinds of questions about ourselves and the world and leaves room for certain kinds of answers to these questions while excluding others. And it does so in virtue of the type of discourse which it is--in virtue, that is, of the set of conceptual relations peculiar to it. As Wittgenstein says:

Concepts lead us to make investigations; are the expression of our interests, and direct our interests.¹

Neurophysiological discourse gives rise to other kinds of questions and excludes certain kinds of answers which sensation-discourse permits. Would not, then, the elimination of sensation-discourse impoverish our understanding

¹Philosophical Investigations, tr. G.E.M. Anscombe (New York: Macmillan, 1953), sec. 570.

of ourselves and the world? Rorty does not think so. Talk about "quantity of caloric fluid" is also associated with certain kinds of questions and permissible answers which are out of place in talk about "mean kinetic energy of molecules." But the rejection of these kinds of questions and answers has not diminished but, on the contrary, enhanced our understanding of the world. Again, talk about demons is associated with certain kinds of questions and permissible answers which are out of place in talk about germs and hallucinations. Yet, we are none the poorer for having allowed demon-discourse to fall into desuetude. Why should sensation-discourse be thought to occupy a privileged position, immune to the kinds of criticism which led to the elimination of demon-discourse and talk about "quantity of caloric fluid?" One question sometimes deserves another. What is our justification for replacing talk about "quantity of caloric fluid" with talk about "mean kinetic energy of molecules," demon-discourse with talk about germs and hallucinations? Rorty's answer is simplicity and the fact that all the predictive and explanatory advantages of modern science are retained. Now the standards of justification appropriate to modern science may indeed be relevant

to the evaluation of these two kinds of discourse, but why should these standards be thought to have the sort of magisterial neutrality which gives one the right to employ them in evaluating any discourse whatever? If the notion that there is no such thing as pre-linguistic awareness means anything at all, it means at least that there is no such neutral point from which one can evaluate the different sets of conceptual relations, or logical spaces, which make up the conceptual scheme embedded in the way we talk. Wittgenstein remarks:

How is the word 'justification' used? Describe language-games. From these you will also be able to see the importance of being justified.¹

Again, he says:

Our mistake is to look for an explanation where we ought to look at what happens as a 'proto-phenomenon.' That is, where we ought to have said: this language-game is played.²

The problem of justification, I think, is the most serious one which Rorty's version of the Identity Theory has to face. It is a problem about which I will have a great deal

¹Ibid., sec. 486.

²Ibid., sec. 654.

more to say in later chapters. Rorty is not unaware of this problem and offers by way of solution a general theory of the conditions under which a term may cease to have a referring use. This theory, however, as I will argue in Chapter II, has no legitimate application in the case of sensation-terms. The reason why it lacks such an application is connected with a second problem. Identity Theorists are right, I believe, to reject the view that it cannot make sense to assert that empirical inquiry will discover that sensations are identical with certain brain processes unless it also makes sense to assert that empirical inquiry will discover that other mental states, such as thoughts, beliefs, desires, intentions, and attitudes, are identical with certain brain processes. For many reasons, it is implausible to maintain that mental states such as these are identical with brain processes. This does not mean, of course, that materialism must inevitably come to shipwreck over such mental states, for there are good reasons for thinking that they will eventually submit to behavioristic explanations (although there are also good reasons for thinking that they will not). Mental states comprise a highly diverse lot and should not all be expected to fall

under one type of explanation. But it does not follow from their diverse character that they are not in any sense logically connected with one another. I accept P. F. Strawson's view that "the topic of the mind does not divide into unconnected subjects."¹ Each type of mental state is inextricably bound up with the others, interwoven with them. I also hold, as do Strawson and even Rorty, that the connection between them is in some sense logical. Given this view of the mind, which is shared both by defenders of Eliminative Materialism and their "linguistic" opponents, it ought to be asked what effect the elimination of sensation-discourse would have on other types of mental discourse. If the elimination of sensation-discourse would make significant parts of other commonly employed types of mental discourse unintelligible, then this at least would be an important disanalogy between such elimination and the elimination of demon-discourse or discourse employing the concept of caloric fluid. And it is primarily analogies with the elimination of other types of discourse to which Eliminativists appeal to make out their case. This is a

¹Individuals (Garden City, New York: Anchor Books, 1963), p. 109.

question, however, which has barely been discussed. The reason, I believe, is twofold: first, a concentration on too narrow a range of cases and, second, an uncritical acceptance of an oversimplified view regarding the nature of the logical connections between mental states. By concentrating only on cases in which sensation-discourse either replaces the primitive, natural expressions of sensation, such as crying or grimacing, or explains them, the connection of sensation-discourse with other types of discourse--in particular, discourse in terms of desires, intentions, and attitudes--has been overlooked. As Wittgenstein points out:

We surely do not always say someone is complaining, because he is in pain. So the words 'I am in pain' may be a cry of complaint, and may be something else.¹

Sensation-discourse is employed in the explanation of a whole range of human actions, respecting many of which explanation in wholly physiological terms or even physiological-cum-behavioristic terms would be singularly inappropriate. This is a point I try to bring out in Chapter II. If, however, the logical connections between mental states were

¹Op. cit., p. 189.

of the nature of linguistic conventions easily modifiable in light of future scientific advances, as Rorty holds, the elimination of sensation-discourse need not make unintelligible significant parts of these other types of discourse. Now I would agree that the logical connections between mental states are based on contingent facts. But it does not follow from this that these connections are contingent on the state of empirical inquiry. In Chapter III, I argue that the contingent facts underlying the connection between pain and wanting are such that the elimination of that part of sensation-discourse employing the concept of pain would make unintelligible a significant part of discourse employing the concept of wanting. With respect to the suggestion that it might be possible also to eliminate the latter type of discourse, I try to show that its elimination would make unintelligible a significant part of discourse employing the concept of intention and, further, that there is a logical absurdity in the idea that the existence of intention-discourse is contingent on the state of empirical inquiry. That such consequences would follow from the elimination of sensation-discourse would not be enough to show that the latter could not be eliminated, however, if it were the

case either that that part of wanting-discourse or that part of intention-discourse which would thereby become unintelligible could also be eliminated without affecting the rest of these types of discourse. But the parts of wanting-discourse and intention-discourse which would be affected are exceedingly common and exceedingly important. For it is in their terms and only in their terms that a great variety of human actions--actions of the kind I discuss in Chapter II--become intelligible. In view of the extensive range of human phenomena which these parts of wanting-discourse and intention-discourse explain, therefore, it is doubtful whether they could be eliminated without thereby radically affecting the rest of these types of discourse and hence even our present concept of a person. That such radical changes would follow upon the elimination of sensation-discourse constitutes, I believe, an important disanalogy between such elimination and the elimination of demon-discourse or discourse employing the concept of caloric fluid and casts serious doubt on the claim that the logical connections between mental states--at the very least, the logical connections between sensations, desires, and intentions implied in those parts of wanting-discourse and intention-discourse which the

elimination of sensation-discourse would make unintelligible--are contingent on the state of empirical inquiry.

Now if the logical connections between these mental states are not contingent on the state of empirical inquiry, this has important consequences for a third problem facing Eliminative Materialism and, in general, the Identity Theory. This is the problem of incorrigibility, which has received considerable attention in recent literature. It is maintained by opponents of the Identity Theory that first-person present-tense sensation statements are incorrigible in a sense in which physiological statements cannot be incorrigible and that this shows that physiological statements could not in principle replace sensation statements but could at most provide only evidence for their truth. Identity Theorists have sought to meet this objection either by challenging the significance of incorrigibility in the sense in which it cannot be a feature of physiological statements or by trying to show that it would be reasonable in the light of future scientific advances to eliminate this feature. In Chapter IV, I consider the major arguments which have been offered against the significance of incorrigibility for the Identity Theory and argue that they

fall short of their goal. Arguments on which Identity Theorists tend most to rely in support of the claim that it would be reasonable in the light of future scientific advances to eliminate incorrigibility are versions of what has come to be called the electroencephalogram (EEG) argument. I take up this claim in Chapter V. I argue there (1) that the EEG argument does not show this claim to be true and (2) that no merely empirical argument can show this claim to be true. In my argument for (2) I contend that the logical connections between sensations, desires, and intentions implied in those parts of wanting-discourse and intention-discourse which explain actions of the kind I discuss in Chapter II are dependent on the feature of incorrigibility and, hence, that this feature is not contingent on the state of empirical inquiry since these connections are not.

CHAPTER II

LAWS AND SENSATIONS

Eliminative materialism claims that there are no such things as sensations. This claim is intuitively implausible, because it seems to imply that people who report sensations hold false beliefs. Proponents of eliminative materialism either accept this consequence and try to explain away its implausibility by assimilating the ordinary language of sensations to an all-pervasive scientific theory or else argue that the claim that there are no such things as sensations does not have this consequence and try to account for its intuitive implausibility in some other way, e.g., the practical inconvenience of eliminating sensation talk.

Feyerabend, for example, adopts the first approach and argues that the reason it sounds so implausible to say that people who report sensations hold false beliefs is that the beliefs to which talk about sensations commits one are part of a theory so pervasive that many facts with which one might seek to compare the theory in order to test it are formulated in terms of the theory and therefore already

prejudiced in its favor.¹ This approach, I think, does not remove the implausibility but merely serves to shift it from the alleged consequences of the claim that there are no such things as sensations to what seem to be consequences of the claim that sensations are theoretical constructs. One consequence which seems to follow from the latter claim, for example, is that we do not directly experience sensations. For something is a theoretical construct only if we do not directly experience it. But sensation-terms are clear cases of terms used to report what we directly experience.² Another consequence of making sensations into theoretical constructs is that statements expressed by sentences such as

(1) I have a toothache

would seem to commit those who assert them to the statement expressed by

(2) There are pains

in the same way that statements expressed by sentences such as

¹See "Materialism and the Mind-Body Problem," Review of Metaphysics, vol. 17 (1963), pp. 50-53.

²For the development of this objection, see James W. Cornman, "Mental Terms Theoretical Terms, and Materialism," Philosophy of Science, XXXV (1968), 45-63.

(3) Those tracks (on the glass surface of a Wilson Cloud Chamber) were made by electrons
commit those who assert them to the statement expressed by

(4) There are electrons.

But the move from (1) to (2) is just not on a par with the move from (3) to (4). A statement requires circumstances.

(1) expresses the statement that I have a toothache only in certain circumstances. Suppose a stranger came up to me, uttered (1), and then quickly walked away. Did he make a statement? What statement? Was he a foreigner who did not know the language and thought that (1) expressed a warning to watch out for the open manhole I was approaching? Did he mistake me for an accomplice to whom he wanted to relay an important message in code? Was he a motorist whose new car had just broken down again and who in exasperation wanted to tell the first person he met that he had a "lemon," but thought that (1) would be a better expression for conveying what he wanted to say? It might be objected that the stranger, at least, knew what he meant by uttering (1), and this is sufficient for (1)'s having been used to make a statement. But can a person make a statement if he is addressing someone who does not understand what he means?

He can try to make a statement. A person often searches for "the right words," using first one form of expression and then another until he hits upon one which enables him to say what he has been trying to say. He succeeds in saying what he has been trying to say only when the person or persons (or, at least, some of them) whom he is addressing come to understand what he means.¹ That a person cannot make a statement unless those whom he is addressing are capable of understanding what he means is even clearer. Can a person talk to a drapery hook? Can he even try to talk to a drapery hook? He can, of course, go through the motions--he can utter words while in its presence. I am not saying that there must always be someone else whom a person is addressing if he is to succeed in making a statement. A person can talk to himself, tell himself things. But this consideration is irrelevant to the question of whether the stranger who, in my example, came up to me and uttered (1) had succeeded

¹It might be thought that a person can succeed in making a statement if someone who overhears his words understands what he means even though the person or persons addressed do not. But it is more natural to describe this case as one in which the overhearer knows what the speaker is trying to say rather than what the speaker is saying.

in making a statement. For he was not telling himself that he had a toothache.

Now the circumstances in which (1) can be used to state that I have a toothache are not circumstances in which (2) can be used to make a statement. If I had been acting irritably, or turned down an invitation, or made an appointment with the dentist, I could use (1) to inform someone who wondered what was up that I had a toothache. But in these circumstances, to what question could (2) be used to give an answer? What possible information could (2) be used to convey? In the circumstances of ordinary life where (1) has its home, (2) has no use. Indeed, if someone were to come up to me and utter (2) outside a philosophical context, I would be at a loss as to what he meant. If this is true, however, then I do not see how it can be maintained that statements expressed by sentences such as (1) commit those who assert them to the statement expressed by (2). If (2) fails to express a statement in the circumstances in which (1) succeeds in expressing the statement that I have a toothache, it would seem that the use of (1) carries with it no such existential commitment. Or if it does, this would at the very least be a queer sort of commitment. Such queerness,

however, does not affect the transition from (3) to (4).

I think it is plain that the circumstances in which (3) can be used to state that those tracks were made by electrons are circumstances in which (4) also has a use. Indeed, one could easily imagine a discussion between adherents of the corpuscular interpretation of quantum physics and adherents of the wave interpretation in which both (3) and (4) were used to make statements.

It may be objected to the foregoing that statements expressed by sentences such as (1) must carry with them existential commitment regarding pains. If the words "a toothache" in (1) were not used in such a way that they purport to designate an object distinct from the person uttering (1), there would be nothing for (1) to be about and hence it would be meaningless. But this is not true.

Consider the sentence

(5) The grocer had a glimpse of the man who stole Smith's wallet.

(5) does not express a statement about the man who stole Smith's wallet, but about the grocer's glimpse of this man. Now clearly it is implausible to say that (5) commits the person uttering it to a belief in the existence of glimpses.

The words "a glimpse" need not ostensibly refer to an object distinct from the grocer in order that there be something for (5) to be about. The genitive inflection in the phrase "the grocer's glimpse" does not signify the relation of belonging to, as it does in the phrase, "the grocer's apron," but rather the manner in which the grocer saw the man who stole Smith's wallet.¹ The proper analysis of (5) is not

($\exists x$)($\exists y$)($\exists z$)(x is the grocer & y is the man who stole Smith's wallet & z is a glimpse of y & x had z)

¹It may be said that although the grocer does not have a glimpse in the same sense in which he has an apron, it does not follow that the genitive inflection in the phrase "the grocer's glimpse" does not signify the relation of belonging to. It could be that the grocer has a glimpse only in virtue of having something else, viz., an act of seeing, in the sense in which he has an apron. But this suggestion is implausible for two reasons. First, there is no ordinary sense of the expressions "having," "possessing," and "belonging to" in which it makes sense to speak of "having an act of seeing" or "possessing an act of seeing" or "an act of seeing belonging to someone." It is true that these former expressions can be given a sense which make these latter expressions meaningful. Since, however, it does make sense to speak of "having an apron" or "possessing an apron" or "an apron belonging to someone" in the ordinary sense of "having," "possessing," and "belonging to," it follows that the given sense of these latter expressions would not be the sense in which the grocer has an apron. Second, the grocer has an apron in a sense in which it is possible for the apron not to belong to anyone at all. But it is implausible to maintain that the grocer can "have" an act of seeing in a sense in which it is possible for the act of seeing not to belong to anyone at all. These remarks apply mutatis mutandis to the expressions "having a toothache" and "feeling nauseous" which I assimilate in the text to "having a glimpse."

but rather

$(\exists x)(\exists y)(x \text{ is the grocer} \ \& \ y \text{ is the man who stole}$
 Smith's wallet & x saw y fleetingly).

Analogously, it is quite possible that the proper analysis of sentences such as (1) will assimilate them to sentences such as

(6) I am feeling nauseous.

The expression "have a toothache" on such an analysis will not be further analyzable and will serve merely to characterize a way of being conscious.

Rorty avoids these difficulties connected with Feyerabend's version of eliminative materialism by maintaining that words for sensations are observation-terms, although, indeed, he insists that "the distinction between observation-terms and non-observation-terms is relative to linguistic practices (practices which may change as inquiry progresses)" ¹ On this position, however, there seems to be no way of reconciling the intuitive implausibility of saying that people who report sensations hold false beliefs with the claim that there are no such things

¹"Mind-Body Identity, Privacy, and Categories," p. 40.

as sensations. To claim that there are no such things as that to which an observation-term purports to refer seems to imply that people who report them hold false beliefs. Thus, for example, to claim that there are no unicorns seems to imply that people who say such things as "I chased a unicorn out of my clover-patch yesterday" hold false beliefs. Accordingly, Rorty argues that this implication does not hold with respect to sensations. His argument consists in sketching a general theory of the conditions under which such an implication does not hold. The theory-sketch is this.

(1) X's are the subjects of both inferential and non-inferential reports; (2) empirical discoveries are made which enable us to subsume X-laws under Y-laws and to produce new X-laws by studying Y's; (3) inferential reports of X's cease to be made; (4) non-inferential reports of X's are reinterpreted either (4a) as reports of Y's, or (4b) as reports of mental entities (thoughts that one is seeing an X, hallucinatory images, etc.); (5) non-inferential reports of X's cease to be made (because their place is taken by non-inferential reports either of Y's or of thoughts, hallucinatory images, etc.); (6) we conclude that there simply are no such things as X's.¹

I want to begin my discussion of Rorty's version of eliminative materialism with an examination of this theory. In particular, I want to examine stage (2) of the theory. For

¹Ibid., pp. 35-36.

stage (2) to have an application to the case of sensations, there must be such things as sensation-laws which occur essentially in sensation-explanations. But what are sensation-laws? It is worthy of note that Rorty does not provide us with any examples of such laws. In fact, there is a marked scarcity of such examples in recent materialist literature. Scientific materialists should find this scarcity embarrassing, since it only serves to strengthen the conviction of their opponents that sensation-statements play a different role in our lives, have a different use, than do neurophysiological statements and hence that sensation-laws, if there are such, are not statements of a type which can be subsumed under neurophysiological laws.

To clear up this question, it will be useful to call to mind the sorts of things which go under the heading of "sensations." Besides pain, there are the sorts of things which make up the well-known list assembled by Gilbert Ryle in The Concept of Mind:¹ thrills, twinges, pangs, throbs, wrenches, itches, prickings, chills, glows, loads, qualms, hankerings, curdlings, sinkings, tensions, gnawings and

¹(New York: Barnes and Noble, 1949), pp. 83-84.

shocks. Now what does a throb-law look like? I cannot think of any. Is this because throb-laws are so terribly complicated that no one has yet been able to formulate one of them with sufficient precision? But then how is it possible for them to occur essentially in explanations which even the most untutored persons can understand, as when someone asks me why I shook my head like that and I say that I felt a sudden throbbing in my right ear? Or is it that throb-laws are so terribly obvious that no one has yet bothered to formulate one of them? Then it ought to be easy to think up one. Here is a candidate.

(L.1) Anyone who feels an unpleasant throbbing in his ear will shake his head in such-and-such a manner.

The trouble with this candidate is that it is false. A person might rub his ear instead. Suppose we replace (L.1) by a slightly more complicated candidate.

(L.2) Anyone who feels an unpleasant throbbing in his ear and thinks that shaking his head in such-and-such a manner will be the best way to remove it, will shake his head in such-and-such a manner.

The trouble with (L.2) is that it seems to be restricted to cases in which the sufferer deliberates over a number of possibilities until he decides upon one and then carries out his decision. But in most cases nothing of the sort happens. In fact, in most cases of a person's shaking his head because of a throbbing in his ear, the person is not thinking of anything at all, or, at least, not of anything related to the throbbing. He feels the throbbing and just shakes his head. There are some cases, indeed, in which the person is so engrossed in what he is doing, as when he is playing chess, that he does not realize that he has shaken his head until someone asks him why he has. Suppose now that instead of replacing (L.1) by a more complicated candidate, we replace it by one that is more general.

(L.3) Anyone who feels an unpleasant throbbing in his ear will tend to do whatever is conducive to removing it.

The objection to (L.3) is that it could not possibly be true. For there are many different and incompatible things conducive to removing an unpleasant throbbing in the ear. Perhaps the following will do:

(L.4) Anyone who feels an unpleasant throbbing in

his ear will tend to do something conducive to removing it.

But (L.4) does not explain why the person shook his head rather than, say, rub his ear. The reason why throb-laws are so hard to find is not that they are too obvious for anyone to have thought it worth the effort to formulate them.

The same considerations apply mutatis mutandis to pang-laws, pricking-laws, curdling-laws, etc. Of the sensations which make up Ryle's list, itches seem to be least objectionable as subjects of possible laws. When someone asks me why I am rubbing my nose and I reply that I have an itch, it seems not too implausible to say that my reply serves as an explanation only if it makes a tacit appeal to some such law as

(L.5) People who have itches tend to scratch the place that itches.

I think, however, that the case for the essential occurrence of itch-laws in itch-explanations stands or falls with the case for the essential occurrence of other sorts of sensation-laws in sensation-explanations. The sensation-explanations for which it is most plausible to claim a

tacit appeal to sensation-laws are pain-explanations. This is no doubt one of the reasons why pains have received so much attention in recent materialist literature. Itches seem to stand somewhere between throbs and pangs on the one hand and pains on the other in an ascending order of plausible candidates for subjects of possible laws. The considerations which weigh in favor of pains would serve, I think, to balance the considerations deriving from throbs and pangs which weigh against itches. For this reason I want to turn now to pains.

To the best of my knowledge, materialists have given only one explicit example of what is supposed to count as a pain-law. The example is this.

(L.6) People tend to avoid things with which they have had painful experiences.¹

A few other examples have been suggested. These are all of the following sort.

(L.7) People in pain tend to cry out.²

¹See Hilary Putnam, "Minds and Machines," in Sidney Hook (ed.), Dimensions of Mind (New York: New York University Press, 1960), p. 171.

²The expression "cry out" is not meant to be restricted only to the cries, say, of a child, but includes

(L.8) People in pain tend to grimace.

(L.6), (L.7), and (L.8) have a suspicious look about them which makes it doubtful whether they are statements of a kind that can be subsumed under neurophysiological laws.¹ Before I explore this point, however, I want to call attention to the richness of pain-explanations, consequent upon the very important role they play in our lives, and to the corresponding failure of this handful of examples adequately to represent the full range of pain-laws which, on the present view, would have to be postulated. Consider the following examples of pain-explanations:

(E.1) He said that he felt bad because he was hungry and had a headache.²

(E.2) I declined the invitation because I had a headache.

within its range of application groans and exclamations such as "Ouch!" and "Oh, how my head hurts!"

¹Rorty never explains anywhere what he means by the subsumption of one set of laws under another set of laws. I think it is clear, however, that he means at least that the subsumed laws are less basic or fundamental than the laws under which they are subsumed, in the sense that the former are dependent on the latter.

²This example is Putnam's. See ibid., p. 170.

- (E.3) I won't be at the chess tournament tonight because I have a headache.
- (E.4) I expected to pass out any minute because of the pain.
- (E.5) I finally promised her that I would hire someone to help me, because the pain involved in moving about continued to get worse.
- (E.6) His constant suffering wouldn't let me forget my guilt.
- (E.7) I prayed to God to end my life then, because I could no longer endure the suffering.
- (E.8) I gave him permission to amputate the foot because I could bear the pain no longer.
- (E.9) I ordered him to amputate the foot because of the pain I was in.
- (E.10) I sold my business because I was in too much pain to work any longer.
- (E.11) I finally married her, because the pain prevented me from taking care of myself.

Now it might be thought that no further laws need be postulated for (E.1-11) than (L.6) or those immediately derivable from (L.6). But, first, this is clearly not the case

for (E.4-6). Expecting to pass out any minute is not a way of avoiding painful experiences, although passing out is. Similarly, promising her that I would hire someone is not a way of avoiding painful experiences, although hiring someone is. And remembering my guilt is not a way of avoiding his pain or the pain I feel at his suffering. Even (E.1) seems to be an explanation to which (L.6) is irrelevant. One's saying that he felt bad is not a way of avoiding his hunger and headache. Nor is (L.7) relevant to (E.1) in most cases. One's saying that he felt bad can sometimes be assimilated to the cry "Oh, how my head hurts!," but not always.

Second, since pain-laws, on the present view, are statements of a kind which can be subsumed under neurophysiological laws, the expression "avoid things" in (L.6) would have to be capable of being spelled out in wholly neurophysiological terms. But no statement or set of statements wholly in neurophysiological terms would be sufficient to explain the occurrence of the acts of avoidance in (E.8-11). Giving permission involves more than just saying "you may" A child or a madman can say "You may amputate my foot," but the person to whom this is said does not thereby have the right to amputate. Or even a normal adult can say this to

someone whom he mistakenly believes to be a licensed physician, but it is clearly false that the person to whom this is said has been granted permission to amputate. A person can grant someone permission to amputate only if the former is legally of age and has not been judged legally insane, and the latter has been licensed to perform operations by a governmental agency authorized to issue such licenses. To explain satisfactorily the occurrence of the act of avoidance in (E.8), therefore, the relevant pain-law would have to be capable of expansion into a statement which mentions such things as laws, customs, practices, authority, institutions, and, in general, an organized community. But it is difficult to see how these sorts of things could be spelled out in wholly neurophysiological terms. How, for example, could the law that people become responsible for their actions at the age of 21 consist in "the firing of C-fibers?" Analogous considerations apply to (E.9-11). I cannot order someone to do something unless I am in a position of authority. A ship's-captain who has just been wounded can order a ship's-doctor to amputate his foot, but a petty officer cannot. Now how could the authority to order the ship's-doctor to amputate one's foot be a discharge of neurons or consist in

the firing of C-fibers? Further, to be in such a position of authority presupposes the existence of military laws and institutions, which would, accordingly, also have to be capable of expression in wholly neurophysiological terms. Again, selling one's business involves more than just accepting some pieces of metal and paper and then walking away. The act of avoidance mentioned in (E.10) essentially involves the notions of a contract and money. But a piece of paper with writing on it is not a contract outside an organized community in which there are laws, government, etc. And pieces of metal and paper are not money apart from the existence of financial institutions. That (E.11) does not differ in this respect from (E.8-10) is obvious; marrying someone involves more than merely saying "I do."

So far I have been concentrating on explanations which presuppose the existence of a social community with its rules, customs, and institutions. It would not be difficult, however, to extend these remarks to explanations which presuppose the existence of a moral or religious community. Consider, for example, (E.5-7). (E.5) and (E.6) are explanations, as I have pointed out above, to which (L.6) is irrelevant. It does not follow, of course, that they are

explanations to which no pain-laws would be relevant. But any statement which might plausibly be claimed to be a pain-law occurring essentially in (E.5) or (E.6) would have to be capable of expansion into a statement which mentioned moral responsibility and moral rules, duties, and rights. To explain the memory of my guilt, for example, is not just to explain the memory of his suffering or even the memory of having injured him. Similarly, praying to God involves more than just saying "Please, dear God," Even a savage can be trained to say these words, but is he then praying? I am not saying that savages cannot pray, or even that they do not pray. Nor am I saying that some of the expressions they employ in the course of praying cannot be translated into the words "Please, dear God," But should we happen upon someone belonging to a tribe unacquainted with any form of worship and train him to say these words and to perform all the appropriate physical movements and even have all the appropriate mental accompaniments, we would still not have the right to say that he was praying. Praying is a custom (a use, an institution).¹ One might

¹Cf. Wittgenstein, Philosophical Investigations, secs. 199-200.

grant that a statement or set of statements wholly in neuro-physiological terms would be sufficient to explain the occurrence of the act of saying, "Please, dear God" but how could such a statement or statements explain the existence of the custom of saying these words? And need one grant even this much? Saying, "Please, dear God," involves more than just uttering certain noises. Uttering these noises is saying something only if a place is already prepared for them in a language, and language is essentially interwoven with shared activities and presupposes agreement in its use.¹ In short, saying something presupposes the existence of an organized community. This point about the nature of language has a significance which extends beyond (E.7). It shows that the claim that there is a statement or set of statements in wholly neurophysiological terms sufficient to account for the occurrence of the event explained in (E.1) is vulnerable to the same sorts of objections which have just been raised in connection with (E.5-11), and even casts doubt on analogous claims vis-à-vis (E.3) and (E.4). (E.3) is not a prediction but a statement of intention.

¹Cf. ibid., secs. 241-42.

I am not predicting that I won't be at the chess tournament tonight because I observe that I have a headache and remember that whenever I had a headache in the past I stayed home. Now what is it like for me to intend not to be there tonight? Perhaps I say to myself or someone else, "I'll not go." Or I imagine myself there tonight and shake my head. Or I do not think about it at all but just stay home or go elsewhere. But intending not to be at the chess tournament tonight involves more than just uttering these words or shaking my head while imagining myself there or staying home. Not being there is the fulfillment of intending not to be there, but it is not the fulfillment of these events. It is not clear whether it even makes sense to speak of these events as having a fulfillment. To try to avoid this objection by pushing the intention back from these events to the brain processes which are their causes would be futile. For speaking of a brain process as having been fulfilled would seem to make sense only if it is another way of saying that the brain process has been completed. It is in language that uttering these words or shaking my head while imagining myself there or staying home makes contact with not being there. In any case, intending not to be at the chess

tournament tonight presupposes the existence of the technique of the game of chess; and this in turn presupposes the existence of the custom of playing games. Again, what is it like for me to expect to pass out? I say perhaps "I think I'm going to pass out." Or I simply grab hold of the nearest chair to steady myself. These events or even their neurophysiological causes are not my expectation of passing out, however, for passing out fulfills the latter but (logically) cannot fulfill the former. Only in language do these events and passing out make contact. It seems, therefore, that none of the events explained in (E.1-11) are such that statements in wholly neurophysiological terms would be sufficient to explain their occurrence. That (E.2) does not differ in this respect is plain. I cannot decline an invitation unless there is the custom of extending and accepting invitations.

I have been arguing so far that, on the view that there are pain-laws occurring essentially in pain-explanations, the pain-laws which would have to be postulated to do justice to the richness of pain-explanations in ordinary life would have to be capable of expansion into statements which mentioned such things as laws, customs, practices, authority,

institutions, and, in general, an organized community. My purpose has been to cast doubt on the idea that such laws would be statements of a kind which can be subsumed under neurophysiological laws. Now it might be maintained that although these pain-laws would be difficult to analyze in wholly neurophysiological terms, to do so would in principle be possible and this is all that the present view really requires. But, first, in view of this difficulty, the burden of proof would seem to be on the materialist to show that it is in principle possible to analyze such things as laws, customs, etc., in neurophysiological terms. A few examples, of even the most elementary sort, would be helpful; yet none seem to be forthcoming. Second, there is some reason to think that such an analysis would not be possible even in principle. I have in mind the essential vagueness of the events commonly explained in pain-explanations. If promising, ordering, giving permission, etc., were in principle analyzable into neurophysiological processes, then, since the latter either occur or do not occur, the former would either occur or not occur. But the law of excluded middle does not seem to apply to the occurrence of the former events. Consider promises, for example. The promise mentioned in (E.5) could

have been given in a form of words prefixed by "I promise" but need not have been. All that might have been said was "Okay, I'll hire someone." Now what made these words the expression of a promise? The situation and its antecedents--the whole history of the incident. But the situation could have left it open which way these words were to be taken. (Cf. breach of promise suits which call for a decision by the court.) Even when the "I promise" formula is used, it is sometimes not clear whether a promise has been given. Suppose I say, "I promise . . ." when I am out of my head with pain. Did I promise? No one would say so. But there are transitional cases. At what point can one say, "It is exactly here, at this degree of suffering, that these words no longer count as the expression of a promise?" It is easy to see how these remarks apply to each one of the events explained in (E.1-11). Have I given permission to amputate my foot if I say, "Go ahead" when I am delirious with pain? At what point do I not do so? Marriage in some places is effected by cohabiting, but after exactly how much time--in days, minutes, and seconds? A captain can order a lieutenant to do something--but when they are both prisoners of war? Here, regulations had to be devised in order to decide

the question. Is it mutiny when the crew refuses the crazed captain's demand to turn their guns on ships of their own fleet? Do I say that I feel bad if I mumble the words, "I feel bad" while in deep sleep? None of these questions would call for decisions if the events mentioned were analyzable in principle into neurophysiological processes but rather for further research--for further observation and experiment.

I want to turn now from the sorts of laws which would have to be postulated to explain the occurrence of the events explained in (E.1-11) to the relatively simple laws (L.6-8). These laws, as I mentioned earlier, have a suspicious look about them which makes it doubtful whether they are statements of a kind that can be subsumed under neurophysiological laws. The feature to which I refer is that they all appear to be a priori propositions. If (L.6-8) are a priori propositions, then, since neurophysiological laws are contingent propositions, (L.6-8) cannot be dependent on such laws.¹ A priori propositions express connections which cannot be confirmed or refuted by experience (although some a priori propositions, e.g., "Demons are intangible," can be shown

¹Cf. supra, p. 37, n. 1.

to have no application to experience) and, hence, which cannot be dependent on correlations expressed by contingent laws. Consider (L.6), for example. (L.6) is true a priori, not because of its form but because of its meaning--that is, because of the connection between the concept of pain and avoidance-behavior.¹ Thus, (L.6) is true a priori, not

¹The propositions (1) "No unmarried man is married" and (2) "No bachelor is married" are both true a priori propositions, that is, true propositions not open to confirmation or refutation by experience. (1), however, is true by virtue of its form alone, while (2) is true by virtue of the meaning of its non-logical terms. Thus (1) is a substitution instance of the tautologous propositional form "No non-A is A," while the propositional form of which (2) is an instance, viz., "No A is B," is not tautologous. The tautologousness of "No non-A is A" resides in its property of having only true substitution instances, a property which "No A is B" clearly lacks. To put the same point differently, (1) is true and remains true under any and all reinterpretations of its non-logical terms; (2), however, although true, does not remain true under any and all reinterpretations of its non-logical terms. Now just as (2) is true by virtue of the meaning of its non-logical terms, (L.6) is also true by virtue of the meaning of its non-logical terms. But the truth of (L.6) does not depend on the meaning of its non-logical terms in the same way that (2) does. The truth of (2) depends on the relation of cognitive synonymy, which may for present purposes be explicated as follows: "A" and "B" are cognitively synonymous if and only if necessarily all and only those things of which "A" is truly predicable are things of which "B" is truly predicable. Now it is not true that necessarily all and only those things of which "pain" is truly predicable are things of which "avoidance-behavior" is truly predicable, especially if it be thought sensible (which I doubt) to predicate "pain" of three-dimensional "time slices" of four-dimensional space-time entities. The meaning-relation on which the truth of (L.6) depends is much more complicated. Let "meaning-relation_{L.6}" stand for this meaning-relation. Then, crudely formulated, meaning-relation_{L.6} is as follows: "A" and "B" are meaning-related_{L.6} if and only if necessarily there are some things of which "A" is (truly or falsely) predicable only if most

because it remains true under any and all reinterpretations of its non-logical terms, but because if someone in normal circumstances (that is, not while rehearsing a play, under hypnosis, drugged, etc.) were sincerely to self-apply the word "pain" or one of its cognates in the absence of appropriate simultaneous or subsequent avoidance-behavior, it would not be clear what he was using this word to mean. If someone in normal circumstances were sincerely to utter the words "My foot is very sore," for example, but showed no fear of or made no objection to an inconsiderate handling of his foot, it would be difficult to guess what he could mean. We could, of course, test his knowledge of the language--and even find out that he was using these words correctly. This would be a case, then, in which pain did occur in the absence of simultaneous or subsequent avoidance-behavior. But how would we find out that he was using these words correctly? That he is able to define "sore" in terms of "pain" or its other cognates would not be helpful,

things of which "A" is truly predicable are things of which "B" is truly predicable. The notion of necessity employed here is not that of logical entailment. I try to explicate this notion as well as refine the foregoing crude formulation of meaning-relation_{L.6} in Chapter III.

for what we would want to know is whether he understood the meaning of any of these terms. The only way of finding this out, it seems, would be to observe his application of these words in circumstances in which either he or someone else is exhibiting the appropriate avoidance-behavior.¹ Thus although any proposition of the form "A is in pain but does not and will not exhibit avoidance-behavior" is open to confirmation or refutation by experience and, hence, is a contingent proposition, "pain" and its cognates presuppose the existence of circumstances in which pain occurs together with simultaneous or subsequent avoidance-behavior for their meaning. The concept of pain, that is, presupposes the existence of such circumstances.

To the foregoing it might be objected, as it has been,² that one need not observe how someone applies a word in paradigm cases to find out whether he understands its meaning; it would be possible, at least in principle, to find

¹The avoidance-behavior would not be appropriate if it consisted in, say, showing fear of or making objection to an inconsiderate handling of one's hand.

²See Hilary Putnam, "Brains and Behavior," in R. J. Butler (ed.), Analytical Philosophy, II (Oxford: Basil Blackwell, 1965), pp. 1-19.

this out by examining (e.g.) his brain processes or the electrical waves emanating from his brain. The general reply to this objection has already been anticipated in my previous remarks about the nature of language. The meaning of a word is determined by the rules governing its use. But it is difficult to see how these rules could be brain processes or electrical waves. For they presuppose the existence of an organized community, with its laws, customs, practices, and institutions. Further, these rules are essentially vague. The use of most of our words is never completely determined; there are always borderline cases which call for decision rather than for application of the appropriate rule. But brain processes or waves are not in this way vague, if, indeed it makes sense to speak of them as being vague at all. I think this general reply to the objection is sufficient, but a limited reply can also be made which grants the possibility of finding out whether someone understands the meaning of a word by examining his brain processes or brain waves. That certain brain processes or brain waves were evidence for a given person's understanding the meaning of the word "pain" could be established only by observation of correlations between brain processes or brain waves of the same

kind belonging to other persons and the presence in them of such understanding.¹ But then the way of telling whether a given person understands the meaning of "pain" which consists in examining his brain processes or brain waves would presuppose the existence of some other way of telling whether a person understands the meaning of "pain."² Now this other way of telling, it seems, would have either to consist in observing how a person applies "pain" in circumstances in which he or someone else is exhibiting the appropriate avoidance-behavior or, at least, ultimately presuppose this way of telling. It is important to notice that this limited reply to the objection under discussion cannot stand alone but must finally fall back upon considerations relevant to the general reply. For a possible rejoinder to the former

¹Even Putnam admits this much. See ibid., pp. 14-17.

²It does not follow, of course, that this other way of telling would have to be capable of successful employment with respect to every person. Thus, the way of telling whether a given person understands the meaning of "pain" which consists in examining his brain processes or brain waves would not presuppose the existence of some other way of telling whether that person understands the meaning of "pain." Whatever plausibility Putnam's argument has derives from this consideration. Putnam assumes a community of "super-super-spartans" with respect to which any other way of telling could not be successfully employed.

is that one could conceivably establish that certain brain processes or brain waves are evidence for a given person's understanding the meaning of "pain" by observing correlations between brain processes or brain waves of the same kind and the presence of such understanding in one's own case. Thus one might attach some suitable electrical detecting instrument to one's own skull and observe the patterns it records when one is using the word "pain" to say something. But surely, the rejoinder continues, I need not observe how I apply "pain" in circumstances in which I or someone else exhibits the appropriate avoidance-behavior to find out whether I understand what "pain" means. The most effective way to counter this move, I believe, is to point out that one knows that one understands the meaning of the word "pain" only if it is true that one understands it, that is, only if the concept of understanding the meaning of the word "pain" truly applies in one's case. And the range of application of a concept is determined by rules governing the use of the expression for that concept. Without such rules it makes no sense either to affirm or deny that a concept truly applies to a given case. Now rules of language are essentially interwoven with shared activities made possible by

agreement in behavior.¹ And particular kinds of rules are essentially interwoven with shared activities made possible by agreement in particular kinds of behavior. Among the particular kinds of behavior which in fact underlie the activities interwoven with rules governing the use of the expression "understanding the meaning of the word 'pain'" is avoidance-behavior.

Grimacing and crying out while in pain are also part of the circumstances whose existence is presupposed by the concept of pain. Thus (L.7) and (L.8) are true a priori in the same way that (L.6) is. The former are true a priori, not because it is impossible for pain to occur in the absence of grimacing or crying out, but because if someone in normal circumstances (that is, not while rehearsing a play, under hypnosis, drugged, etc.) were sincerely to self-apply the word "pain" or one of its cognates in the absence of grimacing and crying out, it would not be clear what he was using this word to mean. If someone in normal circumstances were sincerely to utter the words, "I have a headache" or "I

¹Cf. Wittgenstein's remark, "If a lion could talk, we could not understand him," op. cit., p. 223.

have a toothache," for example, but did not exhibit signs of unusual irritability, it would be difficult to guess what he could mean.¹ We could, of course, test his knowledge of the language--and even find out that he was using these words correctly. But the way of finding this out, it seems, would have to consist in observing his application of these words in circumstances in which he or someone else is grimacing and crying out (as well as exhibiting the appropriate avoidance-behavior), or, if one accepts the examination of his brain processes or brain waves as a possible test, would have to presuppose the existence of circumstances in which (at least) someone applied these words while grimacing and crying out (as well as exhibiting the appropriate avoidance-behavior).

It will be useful to summarize the general argument of this chapter. Most kinds of sensations make implausible candidates for subjects of possible laws. The few kinds

¹One can, indeed, sincerely utter these words in the course of teaching someone else their meaning or by way of giving an example, say, in a philosophical discussion. But then one is not self-applying pain-words at all. One is not using these words to say that one is in pain, because one is not using them to say anything. In such cases, they are merely being mentioned.

of sensations which do not are such that the laws in which they might most plausibly be claimed to occur as subjects either would have to be capable of expansion into statements which mention such things as laws, customs, practices, authority, institutions, and, in general, an organized community, or else would be a priori propositions.¹ In either case it is difficult to see how these laws could be subsumed under neurophysiological laws. But if they cannot, then Rorty has failed to show that the central claim of eliminative materialism, viz. that there are no such things as sensations, does not have the intuitively implausible consequence that people who report sensations hold false beliefs. For stage (2) of the general theory he sketches of the conditions under which such an implication does not hold has no application to the case of sensations. Further, it does not seem possible to explain away the implausibility of this consequence, as Feyerabend tries to do, by assimilating the ordinary language of sensations to an all-pervasive scientific theory. This approach seems merely to replace the

¹It should be plain that the argument for the a priori character of the pain-laws, (L.6-8), applies mutatis mutandis to the itch-law, (L.5).

implausibility of saying that people who report sensations hold false beliefs by the implausibility of the consequences which follow upon such an assimilation.

CHAPTER III

SENSATION-EXPLANATIONS AND CONCEPTUAL REVISION

In my last chapter, I presented a number of examples to show that sensation-statements play a different role in our lives, have a different use, than do neurophysiological statements and drew the consequence that sensation-laws, if there are such, are not statements of a type which can be subsumed under neurophysiological laws. Scientific materialists could easily grant the soundness of this argument, however, without sacrificing anything essential to their position. As things stand now, they could say, sensation-statements play a different role than do neurophysiological statements, and sensation-laws cannot be subsumed under neurophysiological laws. But the role that a given type of statement plays is determined by the place prepared for it in a conceptual scheme, and conceptual schemes can be revised. That we do now operate with a conceptual scheme having the above features is to be attributed to the present stage of empirical inquiry. Should there be significant scientific advances of the appropriate sort in the future, it would be reasonable to revise our conceptual scheme so

as to eliminate these features. That Rorty would be prepared to fall back upon such an argument if he came to doubt the feasibility of subsuming sensation-laws under neurophysiological laws is clear from the pragmatic strain running through his many writings on the subject. He would then still be left with the problem of showing that the central claim of eliminative materialism does not have the implausible consequence that people who report sensations hold false beliefs. In what follows, I propose to examine the claim that it would be reasonable in the light of future scientific advances of the appropriate sort to revise our conceptual scheme in such a way that the role played by sensation-statements in our lives could be taken over by neurophysiological statements.

I want to consider first the nature of the relation between pain and wanting. Aristotle held that pain and wanting are necessarily connected. In the course of discussing whether the soul is distinguishable into parts and, if so, in what sense, he says:

[W]here there is sensation, there is also pleasure and pain, and, where these, necessarily also desire.¹

¹De An. 413b 22-23.

The notion of necessary connection is notoriously vague and shot through with difficulties, and it is not at all clear in this passage what Aristotle has in mind. From some of his later remarks it would be reasonable to assume that he is setting forth a mechanistic theory of what sets a human being in physical motion. Such an interpretation would be warranted by the following passages:

[W]hen the object [of perception] is pleasant or painful, the soul . . . pursues or avoids the object.¹

[T]here is a justification for regarding these two as the sources of movement, i.e., appetite and practical thought; for the object of appetite starts a movement and as a result of that thought gives rise to a movement, the object of appetite being to it a source of stimulation. So too when the imagination originates movement, it necessarily involves appetite [M]ind is never found producing movement without appetite . . . , but appetite can originate movement contrary to calculation, for desire is a form of appetite.²

This way of explaining the connection between pain and wanting is not uncommon in the history of philosophy. We find it, for example, in both Descartes and Hume. Descartes concludes his Meditations with a crude mechanistic explanation of why "those who when they are sick desire to drink or eat things hurtful to them."³ Even in sick people, he

¹De An. 431a 8-9. ²De An. 433a 18-26.

³Meditations, IV, in The Philosophical Works of

notes, the sensation of thirst causes the desire to drink which in turn causes the body to be impelled to drink, just as pain in the foot excites the mind to do its utmost to remove the cause of the evil as dangerous and hurtful to the foot. Again, in The Passions of the Soul, Descartes sketches the mechanism whereby anticipation of pain causes fear which in turn causes the desire to flee which in turn causes the body to be impelled to flee.¹ Hume offers a similar account of the relation between pain and wanting. "'Tis obvious," he says,

. . . that when we have the prospect of pain or pleasure from any object, we feel a consequent emotion of aversion or propensity, and are carry'd to avoid or embrace what will give us this uneasiness or satisfaction.^{2,3}

And in another passage:

DESIRE arises from good consider'd simply, and AVERSION

Descartes, trans. Elizabeth S. Haldane and G. R. T. Ross (2 vols., Cambridge: Cambridge University Press, 1967), I, 194 (cited hereafter as the "HR" translation).

¹The Passions of the Soul, I, 35-40 (HR, I, 347-50).

²A Treatise of Human Nature, ed. L.A. Selby-Bigge (Oxford: The Clarendon Press, 1888), p. 414. (cited hereafter as Treatise).

³Like Aristotle and Descartes (see The Passions of the Soul, II, 86-87 (HR, I, 369-70), I am classifying aversion as a kind of desire; it is a desire to avoid what is painful.

is deriv'd from evil. The WILL exerts itself, when either the good or the absence of the evil may be attain'd by any action of the mind or body.¹

Let us call the interpretation of "necessary connection" based on the mechanistic model presented in these passages the "causal interpretation." To say that pain is necessarily connected with wanting, on this interpretation, is to say that pain is causally connected with wanting. I wish to emphasize that I am not offering an analysis of what is sometimes meant by the phrase "necessary connection" but merely giving a name to one type of connection it is commonly used to denote. The notion of causal connection is not much less obscure than that of necessary connection. Indeed, there are current analyses of both notions according to which it would be incorrect to classify the former under the latter. As obscure as the notion of causal connection is, however, it will suffice for the purpose of distinguishing the sorts of necessary connection I intend to discuss below. Now the causal interpretation of the necessary connection held to obtain between pain and wanting has, I think, little to recommend it. First, the mechanistic model associated

¹Treatise, p. 439. By "good" and "evil," Hume tells us immediately afterwards he means, respectively, the sensations of pleasure and pain.

with this interpretation does not seem to fit many ordinary cases of avoiding what is painful or what gives some prospect of being painful. Suppose I jump back suddenly upon brushing against a hot stove. Certainly it was the pain that made me jump. But need there have been a desire to jump interposed between the pain and my jumping? It would clearly be unnatural to describe this case as one in which I felt a searing pain, I wanted to jump backwards, and then I jumped. Of course, this could have happened. If someone were forcing me against the stove so that I could not move away, then it would be natural to describe me as wanting to move away. But notice the difference between the two cases. Again, if I jump backwards at the leap and bark of the crocodile, I need not have first wanted to jump, or even wanted to jump in jumping, but simply jumped. Secondly, even cases in which there is room for the notion of wanting to avoid what is painful or what gives the prospect of being painful are not cases in which there is a causal connection between pain and wanting. Suppose I want to leave unopened the package I receive because my friend has just warned me that it contains a bomb. In this case, it is not the prospect of pain which has made me want not to open the box but my friend's

warning. To maintain on the contrary that it is the prospect of pain which has caused my wanting not to open the box would be to confuse the cause of wanting with its object. Wittgenstein's remark about fear in the following passage applies mutatis mutandis to wanting:

We should distinguish between the object of fear and the cause of fear. Thus a face which inspires fear or delight (the object of fear or delight), is not on that account its cause, but--one might say--its target.¹

Thirdly, even cases in which what is painful or what gives the prospect of being painful is the cause of my wanting to avoid it are not cases in which there is a causal connection between pain and wanting. Consider again the case in which someone is forcing me against a hot stove. Suppose, further, that he does not know that the stove is hot and asks me why I want so much to move away from it. I would not reply "Because of the searing pain in my back," but rather, "Because the stove is hot." One might be tempted to make the following objection: "But clearly it is not just because the stove is hot that you want to move away from it. The stove can be hot even though you do not want to move away from it. That

¹Philosophical Investigations, sec. 476.

part of your body touching the stove might be anaesthetized so that you do not feel any pain." Now the most this objection could show is not that the searing pain in my back is the cause of my wanting to move away from the stove, but that it is part of the cause of my wanting to move away from the stove. For it would be possible to reply: "But clearly it is not just because of the searing pain in my back that I want to move away from the stove. I can have a searing pain in my back even though I do not want to move away from the stove. I might not be standing near the stove, or I might not think I am standing near the stove but near the radiator, or although I think I am standing near the stove I do not think that the stove is hot but that the pain in my back is due to the wrestling match I am engaged in, or" But to infer that A is the cause of B from the premises that C is the cause of B and A is part of C is to commit a well-known fallacy. If the inference were valid, then one could correctly argue that since Antonius' eulogy of Caesar caused much unrest among the multitude and the statement, "I come to bury Caesar, not to praise him," was part of the eulogy, this statement caused much unrest among the multitude. I do not think, however, that the objection succeeds in

establishing even this much, viz. that the searing pain in my back was part of the cause of my wanting to move away from the stove. For it suggests that when I gave the heat of the stove as the cause of my wanting to move away from it, I expected my auditor to go through some sort of process of inference like the following: "Since the stove is hot and he is leaning against the stove, it follows that he has a pain in his back. Therefore, he wants to move away from the stove, because part of the total situation in which he finds himself is his having a pain in his back." But, of course, I do not expect him to go through any such process of inference. I would have had third-degree burns by the time he finished. If I had truly expected this, I would have said right out that I had a pain in my back without trying to test his intelligence. Further, if this case could be correctly described as one in which such a process of inference occurred, the following conversation would have been reasonable: "Yes, I know that the stove is hot and that you are leaning against it. But what has that got to do with your wanting to move away from it?" "What has that got to do with it?--It follows that I am in pain, you clod!" Such a conversation would be ridiculous. It would not be akin to a

conversation like: "Yes, I know that you have just looked at the barometer and saw that it was falling. But what has that got to do with your wanting to take your umbrella?"

"A falling barometer means rain, you clod!" This conversation at least makes sense. The conversation about the stove would be more akin to that special sort of nonsense which Lewis Carroll immortalized in "The Mad Tea-Party."

The claim that there is a necessary connection between pain and wanting seems to be false on the causal interpretation of necessary connection. We shall have to search for a different interpretation, then, if we are going to make out this claim. Although Aristotle, as we have seen, often talks as though it is the causal interpretation he has in mind when discussing the relation between pain and wanting, it is possible to discern another interpretation in his writings which, I believe, brings his claim closer to the truth. This second interpretation is suggested by the following passage:

If any order of living things has the sensory, it must also have the appetitive; for appetite is the genus of which desire, passion, and wish are the species; now all animals have one sense at least, viz. touch, and whatever has a sense has the capacity for pleasure and pain and therefore has pleasant and painful objects present

to it, and wherever these are present, there is desire, for desire is just appetite of what is pleasant.¹

It is clear, I think, that Aristotle is doing something else in this passage than setting forth a causal mechanism. If it is not immediately clear, the context in which the passage occurs leaves no doubt. The problem to which Aristotle is addressing himself is the definition of soul. "It is evident that the way to give the most adequate definition of soul," he says, "is to seek in the case of each of its forms for the most appropriate definition."² The appetitive soul is one of the specific forms of soul. And while the appetitive soul contains the attribute of sensation in its definition,³ that attribute contains the appetitive soul in its own definition.⁴ Since pain is a species of sensation and desire an essential attribute of the appetitive soul, therefore, it would also follow that pain contains desire in its own definition. This does not mean, of course, that pain must be desired but that whatever has the capacity for pain must also have the capacity for desire. This way of explaining the

¹De An. 414b 1-5. ²De An. 415a 12-13.

³See De An. 414b 28-33.

⁴Cf. Aristotle's discussion of essential attributes which contain the subjects to which they belong in their own definitions, Post. An. 73a 34-b 1.

connection between pain and wanting has also had its adherents throughout the history of philosophy. One finds it clearly expressed in the following passages of Descartes:

[The faculties of imagination and feeling] cannot be [clearly and distinctly] conceived apart from me, that is without an intelligent substance in which they reside, for . . . in their formal concept, some kind of intellection is comprised, from which I infer that they are distinct from me as its modes are from a thing.¹

[A]ll these sensations of hunger, thirst, pain, etc. are in truth none other than certain confused modes of thought which are produced by the union and apparent intermingling of mind and body.²

Let us call the interpretation of "necessary connection" occurring in these passages from Aristotle and Descartes the "definitional interpretation." To say that pain is necessarily connected with wanting, on this interpretation, is to say that the concept of wanting is contained in the definition of "pain." Again, I wish it to be understood that I am not offering an analysis of what is sometimes meant by the phrase "necessary connection" but merely

¹Meditations, VI (HR, I, 190). Descartes classifies pain as internal feeling. Cf. Meditations, II and VI (HR, I, 153 and 189).

²Meditations, VI (HR, I, 192). To say that x is a mode of y, as Descartes indicates in the first passage, is to say that x cannot be clearly and distinctly conceived apart from y, although the converse does not hold. See also Principles of Philosophy, I, 61 (HR, I, 244-45).

giving a name to one type of connection it is commonly used to denote. Like the notion of causal connection, the notion of definition is not much less obscure than that of necessary connection. And the notion of containment to which I have just appealed goes no way, certainly, toward explicating that of definition. Even so, at the intuitive level there is a distinction between definitional connections and causal connections, and this is all the present argument requires to mark off the two interpretations of "necessary connection."¹ I wish, however, further to refine the definitional interpretation of "necessary connection" by marking off two sub-classes--the analytic interpretation and what I shall later call the "conceptual interpretation." To say that pain is necessarily connected with wanting, on the analytic interpretation, is to say either that it cannot rationally be conceived that there is some individual in or

¹It is true that Aristotle accounts for the unity of a definition in terms of a causal connection between its elements. But this connection is not one of efficient causality, as it would have to be to warrant assimilating the interpretation of "necessary connection" based on the mechanistic model to the definitional interpretation. The genus of the definition, according to Aristotle, is the material cause which is determined to the species by its differentiae or, more properly, its last differentia, the formal causes or cause of the definition. See Met. Bk vii, ch. 12 and Bk. viii, ch. 6.

anticipating pain who wants nothing, that is,

- (A) ' $\sim((\underline{x})((\underline{x}$ is in pain \vee \underline{x} is anticipating pain) \supset $(\exists \underline{y})(\underline{x}$ wants $\underline{y}))$)' cannot rationally be conceived as expressing a true statement (expresses a false statement in all possible worlds) (expresses a self-contradictory statement)¹

or that it cannot rationally be conceived that some individual with the capacity for pain lacks the capacity to want, that is,

- (B) ' $\sim((\underline{x})(\underline{x}$ has the capacity for pain $\supset (\exists \underline{y}(\underline{x}$ has \underline{y} & \underline{y} is the capacity to want)))' cannot rationally be conceived as expressing a true statement (expresses a false statement in all possible worlds) (expresses a self-contradictory statement).

Whether the analytic interpretation, as I have set forth, adequately expresses the views of Aristotle and Descartes in the passages just quoted is not clear. I am inclined to think that it is a correct expression, especially version (B), of Descartes' views. The case of Aristotle is much

¹These alternatives all stand in need of clarification, and it is not obvious that once given they will be found to be equivalent. Nevertheless, they will suffice to mark off the two sub-classes of definitional connections I have in mind.

more difficult. There is room for the development of his remarks along the lines either of the analytic interpretation or the conceptual interpretation I discuss below.

The claim that there is a necessary connection between pain and wanting is, I believe, more plausible on the analytic than on the causal interpretation. Even so, I do not think it can be upheld on either the (A) or (B) versions of the analytic interpretation. It succumbs in the end to Hume's dictum that whatever can be conceived distinctly can be conceived to exist separate from each other. Consider first version (A). This version can be distinguished into a stronger and weaker claim according as 'y's range of application is or is not meant to be restricted to objects and actions suitable (or, at least, thought suitable) to avoiding the particular pain x has. Cases which fail to satisfy the stronger claim can easily be imagined. Some people pinch themselves to find out whether they are dreaming. In such cases, although people anticipate pain, they do not want to do anything to avoid it but on the contrary want to do something to bring it on. Again, masochism is a familiar psychological aberration. It is so common, indeed, that to call it an aberration seems no longer appropriate.

In such cases, although people have pain, they do not want to do anything to get rid of it but on the contrary want to prolong it to the last delicious moment. It might be objected that the masochist could be correctly described as wanting to avoid the pain of having no pain. This seems to me to be nonsense. For it implies that if his desire is not fulfilled, then it is fulfilled. If he does not avoid the pain of having no pain, then he has pain and, hence, has avoided the pain of having no pain. These cases satisfy the weaker claim of version (A). There are others, however, which do not even satisfy this claim. A person sometimes has sympathetic pain when someone he loves is injured. If my young son has broken his leg, for example, I may feel pain in my knee even though I have no injury there. Now in such cases the person cannot always be correctly said to want to get rid of the pain. In my example, my sympathetic pain might have been occasioned by guilt feelings at having permitted my son to engage in the activity in which he broke his leg and may help to assuage those feelings. Nor can it always be correctly said in such cases that what one wants is to feel sympathy for the injured person. Sometimes one wants to feel sympathy for an injured person-- especially

in cases where one thinks that one ought to feel sympathy for him but cannot. But normally one does not feel sympathy because one wants to, nor is one's sympathetic feeling the object of any want. It might be objected that one who has sympathetic pain must at least want to do whatever he can to alleviate the pain of the injured person. I think this is a powerful objection vis-à-vis cases like that of my young son. But this case can be slightly altered so as to reduce the force of the objection. Sympathetic pain is not always occasioned by observing the sufferings of someone who is loved. It sometimes arises merely in the course of recounting the details of these sufferings. Suppose now that my son had received mortal injuries and that I am recounting the details of his sufferings after his death. It certainly could happen that I would have sympathetic pain, but I would not want to do whatever I could to alleviate the pain of my dead son. But wouldn't I want at least to have alleviated it? This cannot correctly be said. As Aristotle points out, although wishing may relate to things that could in no way be brought about by one's own efforts, wanting cannot, and what is past could in no way be brought

about by one's own efforts.¹ But must not an individual who has pain at least want something, no matter how far-fetched the connection is between what he wants and the pain he has? I do not think that even this need be granted. But, first, notice how far the defender of the claim that there is a necessary connection between pain and wanting has been pushed. He is no longer in the position of defending version (A) of the analytic interpretation but version (B). So let us examine now version (B). Cases which fail to satisfy this version are also not very hard to find. Consider first the case of a dog. A dog can be correctly said to be yelping with pain when he is being beaten. But can he correctly be said to have the capacity to want? I believe so. When he is scratching violently round the edges and snuffling along the bottom of a door beyond which there lies a piece of meat, he can correctly be said to want the piece of meat. Consider, however, a fly, an ant, or a bee. Why do we think a child cruel who tears the wings off a fly or impales it on a pin to watch its futile struggles but are indifferent when he tears the petals off a flower? Is it not because we think that he is taking delight in the suffering of his

¹See Nic. Eth. Bk. ii, chs. 2 and 3 and Bk. vi, ch. 2.

victim? Suppose we learn afterwards that it is merely a mechanical toy made to look very much like a fly.¹ It seems clear that a fly can be correctly said to have the capacity for pain.² But now consider a fly buzzing around an uncovered dish of honey. We quite naturally say that the fly is drawn or attracted to the honey, but would feel uncomfortable in saying that it wants the honey, unless we understood this as a metaphorical way of saying just that it is drawn or attracted to the honey.³ Contrast this case with that of the dog which is scratching violently around the edges and snuffling along the bottom of the door. To say that he is drawn or attracted to the meat seems to be only a metaphorical way of saying that he wants it. It makes sense to speak of wanting something only where it also makes sense to distinguish wanting it from merely being drawn or attracted to it, which does not imply, of course, that someone

¹We do not normally chastise a child who breaks the arms off his toy soldiers. In fact, we expect it.

²Notice how we hasten to assure each other that the shrill sound we hear when boiling a live lobster is just the air escaping from beneath its shell.

³Consider: A man can be drawn or attracted to someone with whom he does not want to have anything to do. But can a fly be drawn or attracted to something it does not want? Is this merely because a fly always wants what it is drawn or attracted to? Oh, happy creature!

cannot want what he is drawn or attracted to. But in the case of a fly such a distinction makes no sense. It makes no sense in the sense that there are no criteria whose satisfaction would establish that a fly is drawn or attracted to what he does not want. But then a fly cannot be correctly said to have the capacity to want, although it can be correctly said to have a capacity for pain.

Do these difficulties with the causal and analytic interpretations show that Aristotle and Descartes were wrong to claim a necessary connection between pain and wanting? I think that most of us feel there is something right in what they were saying. I would like now to try to disclose the source of our refusal to reject their claim even when faced with such obvious counterexamples to its two most natural interpretations. The concept of pain is such that we would not have that concept unless we also had the concept of wanting. I am guided here by Wittgenstein's remarks:

. . . if things were quite different from what they actually are--if there were for instance no characteristic expression of pain, of fear, or joy; if rule became exception and exception rule; or if both became phenomena of roughly equal frequency--this would make our normal language-games lose their point What we have to mention in order to explain the significance, I mean the importance, of a concept, are often extremely general facts of nature: such facts as are

hardly ever mentioned because of their great generality.¹

One fact of nature, which is clearly of the kind Wittgenstein had in mind, is that most human beings who have pain generally want to get rid of it and that most human beings who anticipate pain generally want to avoid it. Wanting to get rid of pain and wanting to avoid it are characteristic experiences of pain which partly explain our having the concept of pain. This fact, of course, does not fully explain our having the concept of pain. A more complete explanation would have to mention such facts which show that it is natural for most human beings generally to react to the pain-behavior of other human beings by pitying and treating the part that hurts, facts to which Wittgenstein calls attention when he remarks:

Imagine not merely the words 'I am in pain' but also the answer 'It's not so bad' replaced by instinctive noises and gestures.²

If it were not natural for most human beings generally to react this way, if, for example, our natural attitude toward someone exhibiting pain-behavior were the same as our

¹Op. cit., sec. 142 and note.

²Ibid., sec. 310. Cf. also ibid., pp. 179-80 and Zettel, trans. G. E. M. Anscombe (Oxford: Basil Blackwell, 1967), secs. 540-41 and 545.

attitude toward a machine one of whose parts had just broken off, there would be no point to our language-game with the word "pain." But this fact and the language-game which it partly explains do not rule out the possibility that some human beings generally do not pity or try to comfort and heal other human beings who exhibit pain-behavior or even that most human beings sometimes do not react this way. Similarly, that it is natural for most human beings generally to want to get rid of and avoid pain and the language-game which this fact partly explains do not rule out the possibility that some human beings generally do not want to get rid of or avoid pain or that most human beings sometimes do not want to get rid of or avoid pain or even that some individuals¹ which have the capacity for pain and which are in pain lack the capacity to want anything at all.

¹The application of the concept of pain to flies, ants, and bees is based on resemblances between their behavior and the behavior of human beings in similar circumstances. Cf. Wittgenstein's remarks ". . . only of a living human being and what resembles (behaves like) a living human being can one say: it has sensations . . ." (Philosophical Investigations, sec. 281). "Look at a stone and imagine it having sensations.--one says to oneself: How could one so much as get the idea of ascribing a sensation to a thing? One might as well ascribe it to a number!--And now look at a wriggling fly and at once these difficulties vanish and pain seems able to get a foothold here, where before everything was, so to speak, too smooth for it" (ibid., sec. 284).

But if this were not a fact of nature, then, again, there would be no point to our language-game with the word "pain." I shall call this interpretation of the necessary connection claimed to hold between pain and wanting the "conceptual interpretation."

I wish to make it clear, if it is not already clear, that the foregoing is not an argument. It is an attempt to explain the source of our refusal to reject the claim that there is a necessary connection between pain and wanting even when faced with obvious counterexamples to two of its most natural interpretations and to propose a third interpretation of this claim which avoids these counterexamples. Now, it may be objected that the explanation given is really no explanation at all, since it presupposes a distinction between what is essential to a language-game and what is not, and this distinction is just what needs to be explained. The necessary connection between pain and wanting, on the present view, is to be explained in terms of the role wanting plays in the language-game of pain. But how do we know that this role is an essential one? Wittgenstein puts the objection as follows:

But how can I decide what is an essential, and what an inessential, accidental, feature of the notation?

Is there some reality lying behind the notation, which shapes its grammar?¹

Wittgenstein makes no explicit reply to this objection, but it is clear from many of the things that he says² what his reply would be. It is one that I accept. To decide what is an essential, and what an inessential, accidental, feature of a language-game, describe the language-game. Assemble cases which throw light on the importance of a given feature to the language-game in question. In particular, to decide whether wanting plays an essential or inessential role in the language-game of pain, describe the language-game of pain. In Chapter II, I assembled cases which, I believe, show the importance of wanting to the language-game of pain. Declining an invitation, making an appointment to see the dentist, hiring someone to do the heavy work involved in running one's business, a person's selling his business and marrying someone who he knows can be depended on to take care of him all constitute behavior expressive of both wanting and pain. But, clearly, it does not follow

¹Philosophical Investigations, sec. 562.

²See ibid., especially secs. 89-133.

from the fact that such behavior is expressive of both wanting and pain that wanting is an essential feature of the language-game with the word "pain." Such behavior is expressive of many other things besides, some of which it would be most implausible to claim were essential to the language-game. Even granting that such cases exist and are indeed common, can there not be disagreement in their interpretation? And, further, cannot other cases be cited of behavior expressive of pain but not of wanting? Of course. That there is disagreement regarding what is essential to a language-game, however, only shows that further description, the assembling of additional cases, is required. But why should it be thought that there must ever be agreement in the interpretation of cases, that there must ever be agreement as to what is essential and what is not? If there cannot be a private language, if language is essentially interpersonal, as Wittgenstein claims and tries to show, then there must be such agreement.

If language is to be a means of communication there must be agreement not only in definitions but also (queer as this may sound) in judgments.¹

¹Ibid., sec. 242.

This is because "the speaking of language is part of an activity, or of a form of life."¹ Agreement in judgments "is not agreement in opinions but in form of life."² I accept this view.

Now the concept of a brain process is not tied up with the concept of wanting in the same way that the concept of pain is. It is not essential to the language-game in which the word "brain process" has its original home that human beings want to put an end to or avoid the occurrence of certain kinds of brain processes or even that human beings want anything at all. The concept of a brain process in fact belongs to an entirely different region of language than does the concept of wanting--to the order, one might say, of causes rather than to the order of reasons. One of the virtues of the conceptual interpretation of the necessary connection claimed to hold between pain and wanting is that it focuses attention on just this point. What makes it easy to confuse the two orders of explanation, I believe, is, first, the fact that they are both orders of explanation and, second, the failure carefully to distinguish

¹Ibid., sec. 23.

²Ibid., sec. 241.

between the quite different notions of explaining why an event occurred and substantiating its occurrence, that is, describing evidence on the basis of which it would be rational to assert that the event did occur. That the word "explanation" is used in both orders does not bring them any closer together than does the fact that the word "prediction" is also used in both orders. Compare in this regard Wittgenstein's remark:

Examine these two language-games:

(a) Someone gives someone else the order to make particular movements with his arm, or to assume particular bodily positions (gymnastics instructor and pupil). And here is a variation of this language-game: the pupil gives himself orders and then carries them out.

(b) Someone observes certain regular processes-- for example, the reactions of different metals to acids-- and thereupon makes predictions about the reactions that will occur in certain particular cases.

There is an evident kinship between these two language-games, and also a fundamental difference. In both one might call the spoken words 'predictions.' But compare the training which leads to the first technique with the training for the second one.¹

And although the occurrence of a certain kind of brain process could become evidence for the occurrence of a certain kind of wanting (on the ground of observed correlations between the occurrence of brain processes of that kind and behavioral criteria for the occurrence of wantings of that

¹Ibid., sec. 630.

kind), statements describing the occurrence of the former could not, together with general laws, explain why a wanting of any kind occurred. The reason one might be tempted to believe that they can, even apart from the failure to distinguish explanation from substantiation, is that statements describing the occurrence of a certain kind of brain process could, together with general laws, explain the occurrence of events which count as evidence for the occurrence of a certain kind of wanting. Thus, such statements could explain the occurrence of other kinds of brain processes correlated with behavioral criteria for the occurrence of that kind of wanting or even explain the occurrence of the bodily movements which such behavioral criteria involve.

There are many features of the concept of wanting in virtue of which it falls outside the order of explanation to which the concept of a brain process belongs. In the passage just quoted, Wittgenstein asks us to compare the "training" which leads to the employment of a concept in the order of causes with the training which leads to the employment of a concept in the order of reasons. It is easy to see that the training which leads to the employment of the concept of wanting is very different from the training

which leads to the employment of the concept of a brain process. Again, in the passage immediately succeeding this one, Wittgenstein contrasts the "antecedents" of--that is, the "thoughts, actions and so on" which lead up to--propositions employing concepts such as that of wanting with the antecedents of propositions employing concepts such as that of a brain process. And in earlier passages, he speaks of the "surroundings" essential to the employment of a concept.¹ It is easy to see that the surroundings essential to the employment of the concept of wanting are very different from those essential to the employment of the concept of a brain process. The concept of wanting requires institutions, customs, practices. Thus, a person cannot want to be king or want a lot of money unless there exist governmental institutions or financial institutions. But such institutions are not required for the employment of the concept of a brain process. A further feature of the concept of wanting, which Wittgenstein does not explicitly discuss but which has lately received considerable attention, is that a particular occurrence of wanting can be identified only by

¹See ibid., secs. 583-84.

reference to an object in an intentional description, that is to say, in that description of the object wanted under which it is wanted by the individual in question. Not every true description of an object is one under which an individual wants it; only under certain of its descriptions will it be wanted. This feature is also one in virtue of which the concept of wanting falls outside the order of explanation to which the concept of a brain process belongs. Only events which do not essentially fall under an intentional description or whose identification does not essentially involve reference to an object in an intentional description are appropriate explananda for this order of explanation.

The concept of pain does not entirely fall outside the order of explanation to which the concept of a brain process belongs. It, as it were, straddles both the order of causes and the order of reasons. While it is possible to explain the occurrence of a pain by reference to the occurrence of a certain kind of brain process, it is also possible to explain the occurrence of a certain kind of wanting by reference to the occurrence of a pain. That the concept of pain has also a role to play in the order of explanation to which the concept of wanting belongs is a consequence of

the conceptual interpretation of the necessary connection claimed to hold between pain and wanting. Indeed it is only on the conceptual interpretation, I believe, that this fact about pain becomes fully intelligible. The middle position of pain with respect to the two orders of explanation seems to have been recognized by Descartes, for he makes pain as well as all the other sensations "certain confused modes of thought which are produced by the union and apparent intermingling of mind and body."¹ Now it may be said that if it is possible to explain the occurrence of a pain by reference to the occurrence of a certain kind of brain process and possible to explain the occurrence of a certain kind of wanting by reference to the occurrence of a pain, then it should be possible to explain the occurrence of a certain kind of wanting by reference to the occurrence of a pain. And this would mean that the concept of wanting does not after all fall outside the order of explanation to which the concept of a brain process belongs. But this objection assumes that explanation must be transitive, and such is

¹Meditations, VI (HR, I, 192). Descartes classifies desire as an "action of the soul" and says that all our desires "proceed directly from our soul, and appear to depend on it alone" (The Passions of the Soul, I, 17 (HR, I, 340)).

clearly not the case. It is easy to think of cases which show that explanation is not transitive. Thus, for example, "A killed B's father" could be an explanation of why B killed A but not why C (who is A's son) then killed B.

The view that pain is necessarily connected with wanting is on the conceptual interpretation of this necessary connection unaffected by obvious counterexamples to two of its most natural interpretations. This view, I have shown, was held both by Aristotle and Descartes and is in agreement, I believe, with the sorts of things we would be naturally inclined to say about the connection between pain and wanting. But a consequence of the conceptual interpretation is that the concept of wanting falls outside the order of explanation to which the concept of a brain process belongs and that the concept of pain has a role to play in the order of explanation to which the concept of wanting belongs. These considerations cast serious doubt, I believe, on the claim that it would be reasonable in the light of future scientific advances of the appropriate sort to revise our conceptual scheme in such a way that the role played by sensation-statements in our lives could be taken over by neurophysiological statements. We are not yet in a position fully to

assess this claim, however. For a natural reply to this line of argument would be to grant that the concept of wanting falls outside the order of explanation to which the concept of a brain process belongs but then to go on to claim that future scientific advances could make it reasonable to re-describe the relevant phenomena in such a way as to eliminate the concept of wanting in favor of a concept which does not fall outside the order of explanation to which the concept of a brain process belongs--in favor, say, of a complex concept of bodily-movements-cum-brain-processes. I do not think, however, that this second claim is anymore plausible than the first, and for precisely analogous reasons. For intending is necessarily connected with wanting in the same sense of "necessary connection" in which pain is necessarily connected with wanting. This should not be surprising, since, as P. F. Strawson observes, the concepts which clearly imply intention "are inextricably bound up with," "interwoven with" the other concepts which imply the possession of consciousness on the part of that to which they are applied. "The topic of the mind does not divide into unconnected

subjects."¹ I believe it is this notion of the unity of mind which also lies behind the following passage from

Aristotle:

. . . particulars subsumed under [living beings] constitute a series, each successive term of which potentially contains its predecessor, e.g. . . . the sensory power the nutritive [C]ertain living beings--a small minority--possess calculation and thought, for (among mortal beings) those which possess calculation have all the other powers above mentioned [viz. the nutritive, appetitive, sensory, and locomotive powers].²

Now if intending is necessarily connected with wanting according to the conceptual interpretation, the elimination of the concept of wanting will involve also the elimination of the concept of intending. But the existence of this latter concept is surely essential to the existence of the concept of a person. To use the terminology of Descartes to express a view with which, perhaps, he might have disagreed: Intending is the essential attribute of persons.³ If the elimination of the concept of a person were a consequence of the elimination of the concept of wanting, we would be entitled, I think, to conclude that future scientific

¹Individuals (Garden City, New York: Anchor Books, 1963), p. 109.

²De An. 414b 29-415a 10. Aristotle held that the power to feel pain is necessarily connected with, if not included under, the sensory power. See De An. 413b 22-23.

³I do not in fact think that Descartes would have disagreed with this view. I rely here on Jaako Hintikka's profound interpretation of Descartes' cogito argument in his article "Cogito, Ergo Sum: Inference or Performance?," Philosophical Review, Vol. LXXI, No. 1, Jan. 1962, pp. 3-32.

advances could not make it reasonable to redescribe the relevant phenomena in such a way as to eliminate the concept of wanting in favor of a concept which does not fall outside the order of explanation to which the concept of a brain process belongs.¹ That intending is necessarily connected with wanting is a view which, as the last passage quoted indicates, Aristotle held and is in agreement, I believe, with the sorts of things we would be naturally inclined to say about the connection between them. Just as in the case of pain and wanting, however, there are obvious counterexamples to two of the most natural interpretations of this necessary connection. Such counterexamples are otiose vis-à-vis the conceptual interpretation of this necessary connection. In what follows I offer arguments in support of these claims. My arguments will be briefer than the ones offered in support of the parallel claims regarding pain and wanting, since the arguments for both sets of claims are precisely analogous.

In the case of intending and wanting, as with pain and

¹The concept of a person, I hold, is an essential feature of our conceptual scheme. The possibility of eliminating this concept is one which I consider later on in this chapter.

wanting, Aristotle wavers between a causal and a definitional interpretation of "necessary connection." The clearest statements of these interpretations occur respectively in the following two passages:

The origin of action--its efficient, not its final cause--is choice, and that of choice is desire and reasoning with a view to an end.¹

The object of choice being one of the things in our own power which is desired after deliberation, choice will be deliberate desire of things in our own power; for when we have decided as a result of deliberation, we desire in accordance with our deliberation.²

These two passages together provide an account of how someone comes to perform an action. To put it schematically, A wants something X, deliberates on the means of obtaining X, decides on Y, and then rationally wants (chooses, intends) to do Y, which results--if nothing external prevents it--in his doing Y. On this account, wanting is both causally and definitionally connected with intending. A's wanting X is a causally necessary condition (in the sense of efficient causality) of his intending to do something Y which he regards as a means of obtaining X, and A's intending to

¹Nic. Eth. 1139a 31-32. Aristotle uses the concept of "choice" where we would use that of "intention."

²Nic. Eth. 1113a 10-12.

do Y just consists in his rationally wanting to do Y. The objections raised earlier to the view that there is a necessary connection between pain and wanting, where "necessary connection" is interpreted either causally or definitionally (in the analytic sense), have their counterparts here.

First, the mechanism Aristotle sets up between A's wanting something X and A's doing Y with the intention of obtaining X does not seem to fit many ordinary cases of acting with intention. Suppose that while in the supermarket I pass the dairy counter and, suddenly remembering that I am out of eggs, reach out, take the top carton off the third stack of cartons, and put it in my shopping-cart. Now it is certainly true that I wanted eggs and performed this particular action with the intention of obtaining them. But I did not deliberate on the various possible ways of obtaining them, decide on reaching out, taking the top carton off the third stack of cartons, and putting it in my shopping-cart, and then intend to do just that, whereupon I did it. In this case, although the intermediate steps seem to be lacking, wanting and acting with intention are both present. It might be said, therefore, that the case does not provide a counterexample to the view that wanting is a causally

necessary condition of intending. Now there are some cases in which it is plausible to claim that wanting is a cause of intentional action. Suppose someone were to ask me why I am going to the supermarket, and I answer that I am out of eggs. It could be argued that I am giving a causal explanation of my action in terms of my wants. If I do not want any eggs, the fact that I am out of them will not explain why I am going to the supermarket. But whatever analysis of causation one accepts, a causal explanation must at least make what is explained more intelligible than it was before. This condition, however, would not be satisfied by a putative causal explanation in terms of my wants of the action in my original case. If someone were to ask me why I took the top carton of eggs off the third stack of cartons and put it in my shopping-cart, the answer "I wanted some eggs" would not make my action any more intelligible than it was before. Wanting-statements can serve as causal explanations only in certain circumstances.

Secondly, cases can easily be found or invented which fail to satisfy the formula "A's intending to do Y just consists in his rationally wanting to do Y.. Interestingly, there are hints of such cases in the very same discussion.

in which Aristotle presents his definition of "choice."

Consider the following remarks:

[M]en make themselves responsible for being . . . self-indulgent . . . by spending their time in drinking bouts and the like; for it is activities exercised on particular objects that make the corresponding character [I]t is irrational to suppose that a man who acts . . . self-indulgently [does not wish] to be self-indulgent. . . . [T]o the self-indulgent man, it was open at the beginning not to become [a man] of this¹ kind, and so [he is] self-indulgent voluntarily

These remarks suggest the following counterexample to the above formula. A intends to stay away from Jimmy's tonight even though he does not want to stay away, because he knows that unless he can bring himself to stop now he will end up like his father. Now if A does not want to stay away, then a fortiori he does not rationally want to stay away. So A intends to do something he does not rationally want to do. It might be replied that wants run counter to one another, as is well known, so that it is possible to describe this case as one in which A both wants and does not want to stay away. The issue, however, is not whether it is possible to describe this case in such a way but whether it is possible to describe it as one in which A has no desire whatever to stay away. To argue that although A does not want to

¹Nic. Eth. 1114a 5-21.

stay away, A must rationally want to stay away because clearly A wants a temperate character and A regards staying away a necessary means of obtaining one would be futile. For by parity of reasoning one could argue that A must rationally want not to stay away because A wants to get drunk and regards not staying away a means of obtaining this condition. So that A rationally wants both to stay away and not to stay away. And since rationally wanting something X, according to the formula, is intending to do X, it would follow that A intends both to stay away and not to stay away, which is absurd. Now this case would satisfy a weaker version of the formula which, nevertheless, still implied an analytic connection between intending and wanting: If A intends to do something X, then either A must want to do X or else A must want Y and regard X as a means of obtaining Y. The case just presented does satisfy the second disjunct of the apodosis. There are cases which can be constructed, however, which fail to satisfy even this version. For persons can-- and, indeed, often do--intend to do things which they do not regard as a means of obtaining what they want but as a means of obtaining what someone else wants. It might be replied that such cases are ones in which the persons do

want something after all which what they intend to do will be a means of obtaining, namely, they want that the other person's wants be satisfied. But this is certainly not always true. I may intend to do something I do not want to do for someone whom I hate--someone whose wants I do not want to be satisfied. I intend to do it out of fear. But then do I not intend to do it because I want, at least, to preserve myself? Not necessarily. The person whom I fear may have died long ago, as I am well aware, but his influence on me while living might have been such that I cannot now bring myself to act contrary to what I think he would have wanted. But must I not at least want something? This brings us to the weakest version of the claim that there is a necessary connection between intending and wanting, on the analytic interpretation of "necessary condition": If A is able to act with intention, then A must have the capacity to want. The status of this claim is a profoundly difficult issue. It cannot be decided by bringing forward cases but turns on the question of whether the concept of a necessarily perfect being--hence, one that necessarily wants nothing--which nevertheless acts freely is internally consistent. The intelligibility of such a concept is not obvious, but

neither is its unintelligibility obvious.

I do not think, however, that the issue of whether intending is necessarily connected with wanting is as profoundly difficult as this one nor that it must wait for its solution upon the solution of the latter. If the necessary connection claimed to hold between intending and wanting is given a conceptual interpretation, then, I think, the issue can be resolved in favor of the claim. The concept of intending is such that we would not have that concept unless we also had the concept of wanting. That most human beings generally want to do what they intend to do and that most human beings generally want to perform the intentional actions they in fact perform are both contingent facts. But unless there were such general facts we would have no use for the concept of intending. It would have no role to play in our lives. If we were to come upon a tribe of super-ascetics who were in all other respects like human beings except that they never wanted to do what they in fact did, what right would we have to say that they had any wants at all? The distinction between wanting and wishing would seem to lose its purpose in such circumstances, and so we would probably hesitate to use the concept of wanting. But then

would we still be willing to use the concept of intending? The phenomena, as Wittgenstein remarks in another connection,¹ would gravitate towards another paradigm. Paradoxically, we would probably come to view the actions of these super-ascetics in the way we view the movements of leaves blown about by the wind. I am not saying that the concept of a super-ascetic is unintelligible. Once we have the concept of intending we can extend it to individuals who never want to do what they do and thus also come to have the concept of a super-ascetic. We may even be able to extend it to a being which has no wants whatever. But a necessary condition of doing this is that it have some role to play in our lives, and it would have no role to play unless most of us generally wanted to do the things we did.

We are now in a better position to assess the claim that it would be reasonable in the light of future scientific advances of the appropriate sort to revise our conceptual scheme in such a way that the role played by sensation-statements in our lives could be taken over by neurophysiological statements. Earlier I argued that the most satisfactory account of the connection between pain and wanting

¹Op. cit., sec. 385.

can be given on what I have called the "conceptual interpretation." A consequence of the conceptual interpretation, however, is that while sensation-statements do in fact explain the occurrence of certain kinds of wanting, neurophysiological statements could not explain why a wanting of any kind occurred. To avoid the conclusion that sensation-statements play an important role in our lives which could not be taken over by neurophysiological statements, it seemed possible to argue that future scientific advances could make it reasonable to redescribe the relevant phenomena in such a way as to eliminate the concept of wanting in favor of a concept of a type of event whose occurrences could be explained by neurophysiological statements. The conceptual interpretation, however, also provides the most satisfactory account of the connection between intending and wanting, and a consequence of this interpretation is that the elimination of the concept of wanting would involve the elimination of the concept of intending. And since intending is, if not the, at least an essential feature of the concept of a person, the elimination of the concept of the former would result in the elimination of the latter. In view of this result, I think that we would be justified in rejecting the claim

that it would be reasonable to revise our conceptual scheme in the way suggested. I do not propose to argue here that intending is an essential feature of the concept of a person, although I believe that compelling arguments can be given. What would have to be shown is that in the same way in which the concepts of pain, wanting, and intending are connected the concept of intending is connected with the vast majority of concepts of the states of consciousness which would be regarded--if not severally, then at least some group or another of them--as essential features of the concept of a person. Such arguments would be lengthy but would not involve anything essentially different from the arguments I have presented in my discussion of pain, wanting, and intending. Should it be thought possible to analyze the concept of a person into purely "theoretical" concepts, my reply is that mental concepts do not fall into two mutually exclusive classes called "theoretical" and "practical." This is a fact which has often been pointed out. Wittgenstein's notions of a "language-game" and "form of life" are both expressions of and meant to give expression to this fact. And the centrality of the concept of intention among mental concepts applicable to persons has received recognition

in a growing literature already too extensive to cite.

I want to consider one final suggestion for avoiding the conclusion that sensation-statements play an important role in our lives which could not be taken over by neurophysiological statements. It may be suggested that future scientific advances could make it reasonable to redescribe the relevant phenomena in such a way as to eliminate the concept of a person with its essential feature of intention in favor of a concept of a type of entity lacking features not amenable to neurophysiological explanation. I think, however, that an argument can be constructed on the basis of the centrality of the concept of intention among concepts applicable to persons which shows that this suggestion is logically absurd.¹ First, describing (and hence redescribing), whether "public" or "private," is an intentional act. Generally, all of thinking except the mere entertaining of a proposition and the free constructions of the imagination in reverie and dreaming (I rule out the thinking of an intuitive understanding such as God is often said to have) involves judging, and

¹I am indebted to Norman Malcolm's article, "The Conceivability of Mechanism" (Philosophical Review, LXXVII (1968), 45-72, especially pp. 67-71) for the idea of such an argument.

this in turn, since judging is either affirming or denying, is an intentional act. If the concept of intention did not exist, accordingly, no marks or sounds would count as "descriptions." Second, having rational grounds for saying or doing anything implies that the saying or doing is intentional. Without the concept of intention, accordingly, nothing would count as "rational grounds" for anything. The logical absurdity of the suggestion under discussion, then, is that the elimination of the concept of a person with its essential feature of intention is inconsistent with, first, the existence of redescriptions in virtue of which it has been eliminated and, second, the existence of scientific advances which make reasonable such redescriptions.

CHAPTER IV

INCORRIGIBILITY AND MIND-BODY IDENTITY

In the controversy between partisans of the Mind-Body Identity Theory and their opponents, logical similarities and differences between explanation-types have received little attention. My discussion so far has focused on logical differences between sensation-explanation and physiological explanation which must be taken account of in deciding this controversy. I want to consider now an issue which has received considerable attention of late, the so-called issue of "in corrigibility." Opponents of the Mind-Body Identity Theory claim that incorrigibility (in the sense defined below) is an important feature of sensation-statements which could not be a feature of physiological statements. Identity Theorists have replied either by challenging the significance of incorrigibility in the sense in which it cannot be a feature of physiological statements or by trying to show that it would be reasonable in the light of future scientific advances to eliminate in principle this feature. I do not propose in this chapter to take up the question of whether the incorrigibility of sensation-statements can be eliminated. This question will occupy the whole of Chapter

V. What I will try to show there is that the same logical differences between explanation-types which make doubtful the possibility of eliminating sensation-statements also make doubtful the possibility of eliminating incorrigibility as a feature of these statements. The discussion of the present chapter will focus on the significance of incorrigibility for the Mind-Body Identity Theory.

The issue of incorrigibility has recently focused on attempts to distinguish various possible senses in which persons might be said to be in a specially favorable epistemic position vis-à-vis their own current mental states.¹ The senses which are of most concern to the present discussion have been conveniently summarized by Norman Malcolm in a recent essay.

I think the facts that give rise to the illusion of privacy would be the following: (a) you can be in doubt as to whether I am in pain, but I cannot; (b) you can find out whether I am in pain, but I cannot; and (c) you can be mistaken as to whether I am in pain, but I cannot.²

¹The most notable of such attempts are William Alston's "Varieties of Privileged Access," American Philosophical Quarterly, vol. 8 (1971), pp. 223-241, and A. J. Ayer's "Privacy," in his The Concept of a Person and Other Essays (New York: St. Martin's Press, 1963).

²"The Privacy of Experience," in A. Stroll, ed., Epistemology, New Essays in the Theory of Knowledge (New York: Harper & Row, 1967), p. 146.

I shall hereafter refer to the conjunction of propositions which results from generalizing these three claims to cover all phenomenal states (as opposed to dispositional states, such as beliefs, desires, and attitudes) as the "thesis of incorrigibility." Descartes was the first philosopher, I believe, to have realized the metaphysical significance of the thesis of incorrigibility. He saw it as providing the criterion by which to distinguish mind from matter.¹ With the steadily increasing sophistication of materialist theories of the mind since Descartes, there has been a corresponding increase in the sophistication of the metaphysical employment of this thesis. The thesis of incorrigibility has been strengthened and refined in various ways until it has developed into a powerful objection to such theories. Among the most vigorous recent proponents of this thesis is Wittgenstein and many of the philosophers who claim to have been influenced by him. All three components of the thesis are expressed in the following well-known passages from the

¹Descartes held that persons are in a specially favorable epistemic position in the three senses stated above even vis-à-vis their own dispositional states (see Meditations, II and III (HR, I, 153 and 157)). Most philosophers today would not accept this view without considerable qualification.

Philosophical Investigations:

Other people cannot be said to learn of my sensations only from my behavior,--for I cannot be said to learn of them. I have them. The truth is: it makes sense to say about other people that they doubt whether I am in pain; but not to say it about myself.¹

My temptation to say that one might take a sensation for something other than what it is arises from this: if I assume the abrogation of the normal language-game with the expression of a sensation, I need a criterion of identity for the sensation; and then the possibility of error also exists.²

G. E. M. Anscombe in Intention claims that pain belongs to a class of things about which a person has the right to speak without observation, which is not to say that they are known without observation. It makes no sense to speak here of knowledge at all, she holds, since there is no possibility of being wrong.³ John Cook, in an important article on the "private language argument," argues that it makes sense to speak of a person's finding out that he is in pain only if it makes sense to speak of a person as not being "in as good a position as one could want" for correctly answering a certain question or making a certain statement about his

¹Trans. G. E. M. Anscombe (New York: Macmillan, 1953), sec. 246.

²Ibid., sec. 288. See also secs. 289, 408, and pp. 221-222.

³(Ithaca: Cornell University Press, 1966), pp. 13-14.

pain, and that this last does not make sense.¹ Again, he suggests that "where [someone] thinks we can (or do) doubt or make mistakes about our sensations, he has merely oddly described something else."² Two components of the thesis of incorrigibility are expressed in the following passage from Sydney Shoemaker's book, Self-Knowledge and Self-

Identity:

[First-person present-tense phenomenal statements] are incorrigible in the sense that if a person sincerely asserts such a statement it does not make sense to suppose, and nothing could be accepted as showing, that he is mistaken, i.e., that what he says is false [I]t is characteristic of [first-person present-tense phenomenal statements] that being entitled to assert such a statement does not consist in having established that the statement is true, i.e., in having good evidence that it is true or having observed that it is true,³ but consists simply in the statement's being true.

And Shoemaker makes it clear that he thinks the third component to follow from the conjunction of these two. It is because this conjunction is true that "the sentence 'I seem to be in pain' is, if not senseless, without a role

¹"Wittgenstein on Privacy," Philosophical Review, vol. 74 (1965), pp. 285-86.

²Ibid., p. 288.

³(Ithaca: Cornell University Press, 1963), p. 216.

to play in our language."¹ Finally, Robert Coburn, in the course of defending Strawson's claim that it is essential to the character of phenomenal predicates "that they are both self-ascribable otherwise than on the basis of observation of the behavior of the subject of them, and other-ascribable on the basis of behavior criteria,"² details a number of "absurdities" which he claims to follow from denying any one of the three components.³

The metaphysical employment of the thesis of incorrigibility (METI) has naturally been subjected to heavy criticism by philosophers interested in defending one form or another of materialism. It has also been criticized by philosophers who have no special interest in defending materialism but who are merely dissatisfied with the arguments offered in its favor. On canvassing the recent literature, I think it is possible to discern six general lines of reply which this criticism has taken. These may conveniently be set out as

¹Ibid., p. 226, n.

²Individuals (New York: Doubleday, 1963), pp. 104-105.

³"Persons and Psychological Concepts," American Philosophical Quarterly, vol. 4 (1967), pp. 214-20.

follows:

- (R₁) Incorrighibility is not a feature of our phenomenal concepts.
- (R₂) Although incorrighibility is a feature of our phenomenal concepts, the empirical facts could have been such that our phenomenal concepts lacked this feature.
- (R₃) Although incorrighibility is a feature of our phenomenal concepts, the empirical facts could have been such that we lacked concepts with this feature.
- (R₄) Although incorrighibility is a feature of phenomenal concepts, it does not follow that sincere first-person present-tense reports of phenomenal states cannot justifiably be over-ridden.
- (R₅) Although incorrighibility is a feature of our phenomenal concepts, additions to our stock of empirical knowledge together with advances in technology could lead to the elimination of this feature from our phenomenal concepts.
- (R₆) Although incorrighibility is a feature of our phenomenal concepts, additions to our stock of

empirical knowledge together with advances in technology could lead to the elimination of concepts with this feature.

The strategy behind (R₁)-(R₆) is to show not that the thesis of incorrigibility is false but that the only sense in which it may be true is one that is not damaging to any materialist theory of the mind. The force of (R₁) is that the thesis of incorrigibility is plausible only as an empirical thesis and so cannot be used to show that there is some reason in logic preventing the reduction of phenomenal concepts to neurophysiological concepts. The force of (R₂)-(R₆) is that although it is not an empirical thesis, it does nothing to show that persons are in a specially favorable epistemic position vis-à-vis their own current phenomenal states, since some or all of its components are either bare tautologies or true in virtue of linguistic conventions modifiable in the light of future empirical inquiry.

Common to (R₅) and (R₆) are the claims (1) that incorrigibility is a feature of our phenomenal concepts and (2) that additions to our stock of empirical knowledge together with advances in technology could lead to the elimination in principle of this feature. Each of these positions makes

yet a third claim, however, which the other does not. This difference is to be attributed to the way in which claim (1) is construed in (R_5) and (R_6) respectively. In (R_5) , as in (R_2) , incorrigibility is held to be a synthetic feature of phenomenal concepts, while in (R_6) , as in (R_3) , it is held to be a necessary feature of these concepts. Now to eliminate from a concept one of its synthetic features would leave that concept intact, although indeed it is doubtful whether a concept whose great majority of synthetic features had been eliminated would survive. To eliminate from a concept one of its necessary features, however, would be to eliminate that concept. It is a consequence of the way in which claim (1) is construed in (R_5) , accordingly, that the sort of empirical results postulated by claim (2) need not lead to the elimination of our phenomenal concepts, while it is a consequence of the way in which claim (1) is construed in (R_6) that such results must lead to the elimination of these concepts. The sets of claims in which (R_5) and (R_6) respectively consist are completed by statements that give expression to these consequences.¹

¹Versions of (R_1) can be found in the following writings: J. J. C. Smart, Philosophy and Scientific Realism

(New York: Humanities Press, 1963), pp. 92-105; C. S. Chihara and J.A. Fodor, "Operationalism and Ordinary Language: A Critique of Wittgenstein," American Philosophical Quarterly, vol. 2 (1965), pp. 281-295; P. E. Meehl, "The Compleat Autocerebroscopist: A thought-Experiment on Professor Feigl's Mind-Body Identity Thesis," in P. Feysabend and G. Maxwell (eds.), Mind, Matter, and Method (Minneapolis, University of Minnesota Press, 1966); B. Aune, Knowledge, Mind, and Nature (New York: Random House, 1967), pp. 31-38; D. M. Armstrong, A Materialist Theory of the Mind (New York: Humanities Press, 1968). G. Sheridan, in "The Electroencephelogram Argument against Incommensurability," American Philosophical Quarterly, vol. 6 (1969), pp. 62-70, defends a version of (R₁) in which the latter is restricted to the application of phenomenal concepts only in some circumstances but not in others; Sheridan suggests, however, that the thesis of incorrigibility is not even true when restricted to circumstances in which (R₁) does not hold, since (R₅) holds in such circumstances. For a statement and defense of (R₂), see A. J. Ayer, "Can There be a Private Language?," in The Concept of a Person and Other Essays (New York: St. Martin's Press, 1963), pp. 36-51, esp. pp. 49-50. In a later essay, "Privacy," in ibid., pp. 52-81, esp. pp. 68-73, Ayer defends essentially the same position, although he makes a half-hearted gesture in the direction of (R₁). Ayer, of course, is not a materialist, but some materialists have thought his method of argument sufficient to rebut certain arguments in favor of METI. See, for example, the use to which his argument has been put by R. Rorty, "Wittgenstein, Privileged Access, and Incommunicability," American Philosophical Quarterly, vol. 7 (1970), pp. 192-205, esp. pp. 201-02. (R₃) is not anywhere explicitly defended, although it is obviously a close relative of (R₂). Interestingly, it is P. F. Strawson who in an early article, his review of Philosophical Investigations, Mind, vol. 63 (1954), pp. 70-99, esp. pp. 83-89, came closest to enunciating (R₃) (in conflict, I would argue, with other positions he later took about the mental) in criticism of what he claimed was Wittgenstein's attempt to draw a certain consequence about the logic of sensation-language from the thesis of incorrigibility. I suspect too that supporters of (R₆) rely on a tacit reference to (R₃) for some of the plausibility of their own position; it is a short (though illegitimate) step from "contingent on the empirical facts" to "contingent on the state of empirical inquiry." (R₄) can be found in R. Rorty, "Mind-Body Identity, Privacy, and Categories," Review of Metaphysics, vol. 19 (1965), pp. 24-54, esp. 41-48. H. Morick, in "Is Ultimate Epistemic Authority a Distinguishing Characteristic of the Psychological?," American Philosophical Quarterly, vol. 8 (1971), pp. 292-95, also subscribes to (R₄) by way of showing the "merely tautological" character of the thesis of incorrigibility, but distinguishes from the latter the (significant)

thesis of "ultimate epistemic authority" which, he claims, is consistent with the possibility of justifiably overriding any given avowal of sensation. The clearest statements and most vigorous defense of (R₅) appear in H. Putnam's articles: "Minds and Machines," in Dimensions of Mind, ed. S. Hook (New York: New York University Press, 1960), pp. 148-79; "The Analytic and the Synthetic," Minnesota Studies in the Philosophy of Science, III, ed. H. Feigl and G. Maxwell (Minneapolis: University of Minnesota Press, 1962), pp. 358-97; "Brains and Behavior" in Analytical Philosophy, II, ed. R. J. Butler (Oxford: Basil Blackwell, 1965), pp. 1-19. G. Sheridan, as earlier indicated, is also a supporter of (R₅). (R₆) was first introduced into the literature by P. K. Feyerabend in "Materialism and the Mind-Body Problem," Review of Metaphysics, vol. 17 (1963), pp. 49-66. The general position of which (R₆) was a part was thereafter taken up and elaborated upon by Rorty in "Mind-Body Identity, Privacy, and Categories" and applied by him to other features of phenomenal concepts but, oddly, not to the feature of incorrigibility. In that article, as earlier indicated, Rorty chose to rely on (R₄), apparently thinking that he had discovered an apodictic disproof of the significance of the thesis of incorrigibility. In any case, Rorty seems to have abandoned (R₄) in later articles and now seems to rely exclusively on (R₆). See, for example, his article "Incorrigibility as the Mark of the Mental," Journal of Philosophy, vol. 67 (1970), pp. 399-424, pp. 416-18.

W. Sellars is often classified with the adherents of (R₆). See, for example, R. J. Bernstein, "The Challenge of Scientific Materialism," International Philosophical Quarterly, vol. 8 (1968), pp. 252-75. Rorty's repeated acknowledgement of debt to Sellars also carries the same suggestion. I think, however, that this is mistaken. Rorty acknowledges his debt to Sellars in a footnote to the following sentence: "There is an obvious sense of 'same' in which what used to be called 'a quantity of calorific fluid' is the same thing as what is now called a certain mean kinetic energy of molecules, but there is no reason to think that all features truly predicated of the one may be sensibly predicated of the other." ("Mind-Body Identity, Privacy, and Categories," pp. 26-27). Rorty evidently understands Sellars here as holding a view from which it follows that what people now call "sensations" is in an obvious sense identical with what will be referred to by some future expression whose logic would permit or require the elimination of embarrassing features demanded by the logic of present sensation-language. In view, however, of Sellars's insistence on the "relocation" of raw feel universals, the "categorical transformation, but not substantive reduction, of raw feel predicates," the "transposing" of the "logical space" of raw feels ("The Identity Approach to the Mind-Body Problem," Review of Metaphysics, vol. 18 (1965), pp. 447-49), a more likely interpretation of his position would be that the features expressed

Defenders of METI often proceed in a way which leaves them particularly vulnerable to--indeed, almost invites--criticism along the lines of (R₂)-(R₆). The three facts mentioned in the passage I have used to introduce the thesis of incorrigibility, for example, are said by Malcolm to be "features of the 'grammar' of the word ['pain'], or of the 'language-game,' with the word."¹ Again, in his monograph Dreaming, Malcolm claims the incorrigibility of first-person dream reports to be "a matter of definition."² And in an earlier essay restricted to a discussion of after-images, he says that after-image reports are incorrigible in the sense that under certain conditions

. . . there cannot be a question of his being in error when he says 'I see an after-image.' There cannot be

by predicates true of what people now call "sensations" could be (that is to say, it makes sense to suppose that they are) identical with features which will be expressed by predicates of an as yet to be elaborated theory of brain activity. The sense of "identical" in question is the trivial sense in which the features expressed by predicates true of what people now call "sensations" are identical with currently unreduced psychic features of persons. This interpretation is supported by Sellars' admission that he accepts only a non-controversial and unexciting version of the identity theory according to which raw feel universals are trivially identical with certain brain state universals. Sellars' position, so interpreted, however, seems to be at odds with (R₆).

¹ Op. cit., p. 46.

² (London: Routledge & Kegan Paul, 1959), p. 81.

a question of whether he 'takes' something to be an after-image that is really not one.¹

The reason why there cannot be a question under certain conditions as to whether the assertion, "I see an after-image" is mistaken, he argues, is that there is no way of finding out the answer--not in the sense that there is some move which it is possible to make, but which we for some reason cannot make. There is nothing we would call "finding out the answer to such a question." This assertion "might be called 'self-confirming,' implying by this that really they have no confirmation."² Cook offers a similar defense of his claim that it makes no sense to speak of a person as not being in as good a position as one could want for correctly answering a certain question or making a certain statement about his pain.

What is said to be senseless is not merely a combination of words but rather an attempt, by means of a combination of words, to make in one language game a move that belongs only to the other language game.³

Arguments of this type appear also in Shoemaker's defense of the thesis of incorrigibility.

That a [first-person present-tense phenomenal statement] is a sincere assertion is itself a logically

¹"Direct Perception," in his Knowledge and Certainty (Englewood Cliffs: Prentice-Hall, 1963), p. 85.

²Ibid., p. 81.

³Op. cit., p. 300.

sufficient condition of its being true, and that such a statement has been asserted with apparent sincerity is itself criterial evidence that it is true.¹

By "criterial evidence" for the truth of a judgment, Shoemaker means evidence which is such that "the assertion that it is evidence in favor of the truth of the judgment is necessarily (logically) rather than contingently (empirically) true."²

I think that critics of METI have been right to point out the inadequacy of this type of argument; but I also think that they have made things too easy for themselves by ignoring other aspects of the incorrigibilist defense. There are clear indications in the writings of many incorrigibilists that they are themselves dissatisfied with such arguments and regard them as effective only within the context of a more extended defense. That incorrigibilists have in mind a more extended defense is even suggested, apart from such clear indications, by their position with respect to (R₂), (R₃), and the arguments by analogy often given in support of (R₅) and (R₆). So far from holding (R₂) and (R₃) to be incompatible with METI, they are among their most

¹Op. cit., p. 216. ²Ibid., p. 4.

vigorous advocates. (R₂) and (R₃) find expression, for example, in Wittgenstein's remark

What we have to mention in order to explain the significance, I mean the importance, of a concept, are often extremely general facts of nature: such facts as are hardly ever mentioned because of their great generality.¹

Shoemaker appeals to analogues of (R₂) and (R₃) dealing with the concepts of perception and memory in the course of defending a thesis of (general) incorrigibility with respect to first-person present-tense perceptual and memory statements.² And Malcolm shows his firm commitment to (R₂) and (R₃) in the following passages:

Our concepts of sensation and emotion, of belief and doubt, grow out of certain regular patterns of behavior and circumstances that are frequently repeated in human life.³

[Dream-telling is] a remarkable human phenomenon, a part of the natural history of man, something given, the foundation for the concept of dreaming.⁴

The plausibility of (R₅) and (R₆) rests, to a large extent,

¹Op. cit., p. 56. See also sec. 142 and p. 230.

²See op. cit., pp. 236-243.

³"The Privacy of Experience," p. 153.

⁴Dreaming, p. 87.

on arguments by analogy. Putnam, for example, who is a vigorous partisan of (R₅), argues by analogy with names of diseases that "'pain' is a cluster-concept," by which he means that "the application of the word 'pain' is controlled by a whole cluster of criteria, all of which can be regarded as synthetic."¹ And Rorty, who could with reason be counted among the chief architects of (R₆),² argues by analogy with demons, devils, Zeus's thunderbolts, and caloric fluid that the future progress of psycho-physiology may lead to the replacement of sensation discourse with brain-discourse.³ Now none of the incorrigibilists mentioned above have ever denied the facts of which these analogies are based. Such facts, indeed, are underscored in the philosophy of Wittgenstein. In an effort to distinguish clearly

¹"Brains and Behaviour," in R. J. Butler, ed., Analytical Philosophy (Second Series) (Oxford: Basil Blackwell, 1965), p. 5. See also "Dreaming and 'Depth Grammar,'" in R. J. Butler, ed., Analytical Philosophy (First Series) (Oxford: Basil Blackwell, 1962), pp. 218-221.

²Although Rorty explicitly uses (R₄) to criticize METI, there is a strong suggestion in his writings of criticism along the lines of (R₆). And the latter sort of criticism seems to be a more natural development of his general position on the Mind-Body Identity Theory. Were he convinced that (R₄) was impotent against METI, I think it clear that he would fall back on (R₆).

³See "Mind-Body Identity, Privacy, and Categories."

his own position from what he calls "logical behaviourism" (under which title he seems to lump together the in many ways widely divergent positions of Ryle, Strawson, Wittgenstein, Wisdom, and Malcolm), Putnam further explains his notion of a "cluster-concept" as follows:

I mean not only that each criterion can be regarded as synthetic, but also that the cluster is collectively synthetic, in the sense that we are free in certain cases to say (for reason of inductive simplicity and theoretical economy) that the term applies although the whole cluster is missing. This is completely compatible with saying that the cluster serves to fix the meaning of the word.¹

How different is such a notion from Wittgenstein's doctrine of family resemblances--so central to his philosophy--according to which a term can be extended gradually to cover cases which have nothing at all in common with those it originally covered?² Putnam uses his notion of a "cluster-concept" against what he calls "the 'change of meaning' account," according to which any change of criteria involves a change of meaning. On such an account criteria for the presence and character of phenomenal states which make it senseless to suppose that a given person does not enjoy incorrigibility vis-à-vis these states cannot be replaced by criteria

¹"Brains and Behaviour," p. 5, n.

²See op. cit., secs. 65-67.

which do not make this supposition senseless without changing the meaning of phenomenal state terms. The "change of meaning" account is a straw man. Neither Wittgenstein nor Malcolm (against both of whom Putnam seems to be arguing) held such a view. Wittgenstein, on the contrary, explicitly rejected the "change of meaning" account. Arguing against Russell, whom he takes to have held this view, he rejects saying in certain cases that a term acquires a different sense according as we assume one definition or another in favor of saying in these cases that we use the term "without a fixed meaning."¹ By using terms "without a fixed meaning," he is not talking about the ambiguity of terms but about indeterminateness of meaning. Changes in meaning can occur without change of meaning in cases where meaning is indeterminate. That he intends to include scientific terms, such as Putnam's names of diseases, among such cases is clear from the remark he appends to his argument against Russell:

The fluctuation of scientific definitions: what today counts as an observed concomitant of a phenomenon

¹Ibid., sec. 79.

A will be tomorrow used to define 'A.'¹

The facts about the use of scientific terms upon which Putnam bases his argument against the thesis of incorrigibility were pointed out by Wittgenstein as early as 1933, when he dictated the "Blue Book." And, it is interesting to note, he employed the very same examples of scientific terms to which Putnam appeals in his support of (R₅).²

Doctors will use names of diseases without ever deciding which phenomena are to be taken as criteria and which as symptoms; and this need not be a deplorable lack of clarity. For remember that in general we don't use language according to strict rules--it hasn't been taught us by means of strict rules, either.³

Putnam cites no passages from Wittgenstein which might serve as evidence that he ever held the "change of meaning" account. Indeed, his attribution of this account to so-called "logical behaviorists" in general seems to be based

¹Ibid. I have given here the literal translation of Wittgenstein's remark, since I think the translator's rendering is misleading. Wittgenstein is talking about the fluctuating definitions of terms, not of objects or events.

²Putnam's use of the same examples may have been deliberate. If so, I think he may have misunderstood the point Wittgenstein was trying to make by means of them, viz. that fluctuating definitions of scientific terms do not lead necessarily to change of meaning. If he did not misunderstand Wittgenstein's point, then, on the assumption that his use of these examples was deliberate, I have certainly misunderstood Putnam.

³The Blue and Brown Books, ed. Rush Rhees (Oxford: Basil Blackwell, 1958), p. 25.

solely on passages taken from Malcolm's monograph Dreaming.

Let us examine very briefly two of these passages.

With adults and older children there are two criteria of behaviour and [sinere] testimony; with animals and human infants there is only the one criterion of behaviour. The concept of sleep is not exactly the same in the two cases.¹

Considering the radical conceptual changes that the adoption of a physiological criterion of dreaming would entail, it is evident that a new concept would have been created that only remotely resembled the old one.²

The point of the first passage, says Putnam, is that "there are two concepts of sleep, because there are two methods of verification."³ Now, the italics notwithstanding, nothing of significance turns on the issue of whether it is more proper to speak of a concept C_1 and a concept C_2 as one concept or as two concepts if C_1 is not exactly the same as C_2 . How we individuate concepts depends upon our criteria of individuation, and our selection of the latter seems to be a matter of arbitrary choice.⁴ What the first passage must

¹p. 23.

²p. 81.

³"Dreaming and 'Depth Grammar,'" p. 211.

⁴Our ordinary concept of a person is indeterminate in the respect that the criteria for its application leave open the question of whether a four-month-old fetus is a person. Suppose we introduce a criterion which further determines our ordinary concept in this respect. Have we now changed our ordinary concept of a person or have we created a new concept of a person? I am not suggesting,

imply, or at least suggest, if it is to be evidence for a "change of meaning" account which can bear the weight of arguments against (R₅) is that in one of the cases the concept for which the word "asleep" is used is not a concept of sleep at all. This is neither implied nor suggested. Indeed, Malcolm implicitly rejects such a suggestion when, a few lines later, he claims that "we should have two totally different senses of 'asleep,'" ¹ if behavior were not a criterion of sleep in the case of adults and older children. The second passage does imply that the word "dreaming" could no longer be used for the concept of dreaming at all if its application came to be determined by a physiological criterion. But this is not because any change of criteria involves a change of meaning but because this particular change would entail "radical" conceptual changes. The other passages Putnam cites from Malcolm ² as evidence that the latter holds--or, at least, at one time held--the "change of meaning" account are even less conclusive. I do

of course, that the introduction of such a criterion would be a matter of arbitrary choice.

¹Dreaming, p. 24.

²"Dreaming and 'Depth Grammar,'" pp. 211-213.

not think there is need to spend much time showing how truly unimpressive to incorrigibilists are the arguments in support of (R_6) based on the history of such concepts as demons, devils, Zeus's thunderbolts, and caloric fluid. Is it to be supposed that incorrigibilists are unacquainted with the phenomenon of the discovery, abandonment, and eventual loss of a concept? Wittgenstein takes note of this phenomenon in the following remark:

. . . new types of language, new language-games, as we may say, come into existence, and others become obsolete and forgotten.¹

That incorrigibilists strongly support some of the views and readily accept the facts on which their critics try to build a case against METI is a difficulty these critics seem to have overlooked. To dismiss this difficulty as just another indication of the inconsistency or, perhaps, of the "consistent delusional system"² of "logical behaviorists" would be a mistake. For it points the way to an understanding of the extended defense of METI. Generally, the extended defense consists in arguments designed to show that neither (R_2) nor (R_3) is incompatible with METI and

¹Op. cit., sec. 23.

²See Putnam, "Brains and Behaviour," p. 19.

that (R_5) and (R_6) are both false.¹ More specifically, it consists in arguments for two claims: (1) neither (R_2) nor (R_3) implies that any component of the thesis of incorrigibility is either a bare tautology or true in virtue of linguistic conventions modifiable in the light of future empirical inquiry, and (2) no component of the thesis of incorrigibility is either a bare tautology or true in virtue of linguistic conventions modifiable in the light of future empirical inquiry. The general argument for (1) proceeds in two stages: it (a) tries to make a case for distinguishing non-empirical truths which do not express a necessary condition for the existence of our conceptual scheme, such as "Demons are intangible" and "Epileptics are under the spell of a demon," from non-empirical truths which do express such a condition and (b) contends that the latter are neither bare tautologies nor true in virtue of linguistic conventions modifiable in the light of future empirical inquiry. The general argument for (2) is that the thesis of incorrigibility expresses a necessary condition

¹That (R_1) is false and that (R_4) is not incompatible with METI are generally assumed without argument. I do not think it difficult, however, to construct good arguments for both of these claims. I try to do so below.

for the existence of our conceptual scheme. This argument also proceeds in two stages: it contends (a) that the concept of a person is basic to our conceptual scheme and (b) that the thesis of incorrigibility expresses a necessary condition for the existence of this concept.

Allusions to the extended defense are easily discernible in the writings of most of the incorrigibilists mentioned above. Anscombe makes it clear, for example, how she would defend her claim that there is no possibility of a person's being wrong in what he says about his pain against the objection that a conflict between the two criteria of behavior and testimony is possible, and such a conflict, since the former criterion has greater weight than the latter, would show a person to be wrong in what he says about his pain. She suggests that such a conflict would not show the falsehood of what is said but would make what is said unintelligible.¹ This sort of reply, if intended to stand alone, is of course vulnerable to criticism along the lines of (R₂)-(R₆). That Anscombe intends this reply to be understood within the context of the extended defense is

¹Op. cit., p. 14.

indicated by the distinction she is careful to draw between the unintelligibility of what is said and the unintelligibility of the person who says it. In the case of such a conflict, she explains while discussing an analogous situation, the person's words would be unintelligible, "not because one did not know what they meant, but because one could not make out what the man meant by saying them [W]e cannot understand such a man."¹ That there is no possibility of a person's being wrong in what he says about his pain, accordingly, is to be explained not by reference to what pain is but by reference to what a person is. One of the clearest indications of the extended defense appears in the recent essay by Malcolm from which I have taken the passage used at the outset of this chapter to introduce the thesis of incorrigibility. Malcolm writes:

What I conceive to be a kind of explanation [of the grammar of sensation], and one that satisfies me, is to see what the consequences would be for our concept of a person, if the grammar of sensation (or of thinking or intending) were different in the respect that the expression of doubt, which has no₂ place in the language-game, were to have a place in it.

¹Ibid., pp. 26-27.

²"The Privacy of Experience," pp. 151-152.

The explanation [of why the expression of doubt is excluded from language] is that [it] could not fit coherently into the structure of our concept of a person. The excluding it from working language is no superficial point of grammar or semantics, but a matter of deep philosophical importance.¹

Malcolm's appeal to the extended defense is not, as it may perhaps appear to be, a recent development grafted onto his former arguments for the sake of answering his critics.

There is evidence of such an appeal in his monograph Dreaming, apparently overlooked by Putnam in his severe criticism of that work. Malcolm's explanation of why the word "dreaming" could no longer be used for the concept of dreaming at all if its application came to be determined by a physiological criterion is that

. . . what were then called 'dreams' would no longer be of interest to poets, psychoanalysts, philosophers, and to all of us, children and adults, who like a strange tale.²

What are now called "dreams" are of interest to poets, psychoanalysts, etc., because a first-person dream-report reveals important information about the person making the report, beyond what it reveals about his behavior--his

¹Ibid., pp. 154-155.

²P. 81.

"dreaming habits." The consequence of adopting a physiological criterion of dreaming is that a first-person dream-report would not reveal information about anything beyond the behavior of the person making the report. For the concept of a person, according to Malcolm, is such that to reveal information about persons which go beyond information about their behavior, first-person dream-reports must be incorrigible, and the adoption of a physiological criterion of dreaming would entail their corrigibility. Thus the presupposition of Malcolm's explanation is that the corrigibility of first-person dream-reports which report what are now called "dreams" could not fit coherently into the structure of our concept of a person. The elements of this explanation already appear in Wittgenstein's brief remarks on the concept of dreaming. Wittgenstein says:

Assuming that dreams can yield important information about the dreamer, what yielded the information would be truthful accounts of dreams. The question whether the dreamer's memory deceives him when he reports the dream after waking cannot arise, unless indeed we introduce a completely new criterion for the reports 'agreeing' with the dream, a criterion which gives us a concept of 'truth' as distinct from 'truthfulness' here.¹

¹Op. cit., pp. 222-223.

That it would be possible to adopt a physiological criterion of dreaming seems to be suggested by this passage. But the criterion could be employed only where the question whether the dreamer's memory deceives him when he reports a dream after waking can arise. And this question cannot arise where our interest is in the dreamer rather than in his behavior.¹ Allusions to the extended defense appear also in Shoemaker's writings. That Shoemaker intends his arguments for the thesis of (general) incorrigibility with respect to first-person present-tense perceptual and memory statements to be understood within the context of the extended defense is indicated by the following passages, in which he suggests that this thesis expresses a necessary condition both for the existence of our conceptual scheme and for the existence of the concept of a person.

It seems to me . . . that it follows from the logical possibility of anyone's knowing anything about the world that² perceptual and memory beliefs are generally true. . . . [If human beings lacked the capacity generally to make true perceptual and memory statements] they would not make perceptual and memory statements at all, could not be taught to make them, and could not be said to have beliefs that are expressible in such statements.³

¹Cf. ibid., pp. 179, 184, and 190.

²Op. cit., p. 235. ³Ibid., p. 241.

If a group of human beings could not be taught, or trained, generally to make true perceptual and memory statements, they could not be said to perceive or remember at all in the sense in which persons do, and could hardly be said to be persons at all.¹

Shoemaker's arguments for the claims suggested in these passages are taken up and further developed by Coburn in his article cited earlier in which he defends the thesis of incorrigibility by calling attention to a number of absurdities that are supposed to follow from denying it.

Failure to take account of the extended defense of METI has vitiated much of the criticism directed against it. Arguments in favor of (R₂) or (R₃) do nothing to show that METI is incorrect unless they are accompanied by arguments which show that either the general argument for claim (1) of the extended defense or the general argument for claim (2) of the extended defense is incorrect. Now cogent arguments in favor of (R₅) or (R₆) would show that either one or both of these general arguments are incorrect, since METI is incompatible with both (R₅) and (R₆). There are serious objections to the arguments which have so far been given in favor of (R₅) and (R₆), however, arising from considerations relevant to the extended defense of METI. These

¹Ibid., p. 239.

are serious objections to the arguments which have so far been given in favor of (R₅) and (R₆), however, arising from considerations relevant to the extended defense of METI.

These objections seem to have been overlooked by most proponents of (R₅) and (R₆); at least, few attempts have been made to answer them. I will take up (R₅) and (R₆) in Chapter 5. In the remainder of this chapter, I want to consider (R₁) and (R₄).

Arguments in favor of (R₁) are either based on Hume's dictum that what is distinguishable is separable or consist in bringing forward alleged counterexamples to the thesis of incorrigibility, interpreted as a logical thesis. All arguments of the first sort take the following general form:

- (H)
- (1) What is distinguishable is separable.
 - (2) Phenomenal state x and the awareness of phenomenal state x are distinguishable.
 - (3) Therefore, phenomenal state x and the awareness of phenomenal state x are separable.

It follows from (3) that (at least) conjunct (c) of the thesis of incorrigibility is false, since phenomenal state x could exist in the absence of the awareness of phenomenal

state x, and conversely.¹ J. J. C. Smart and D. M. Armstrong, two prominent recent advocates of (R₁), rely almost exclusively on arguments of this sort to make out their case for (R₁).² The trouble with such arguments is that, as they stand, they are all subject to reversal. If, as incorrigibilists maintain, the thesis of incorrigibility is true, then either premise (1) of (H) is false or premise (2) of (H) takes only false propositions as values. What is needed are further arguments for premises (1) and (2), but these are not provided. Clearly, premises (1) and (2) are not self-evident. Consider premise (1). A grin, if I may be permitted to use one of Smart's own examples³ (borrowed from Lewis Carroll) against him, is distinguishable but not separable from the mouth in the shape of a grin. If they were not distinguishable, then whatever is true of the grin would be true of the mouth in the shape of a grin. But, first, it

¹For convenience, I am here departing from the ordinary use of the term "awareness," according to which "A is aware of x" implies "There is an x of which A is aware." Nothing in my argument turns on this extended use of the term.

²See Smart, Philosophy and Scientific Realism (London: Routledge & Kegan Paul, 1963), p. 100, and Armstrong, A Materialist Theory of the Mind (London: Routledge & Kegan Paul, 1968), pp. 106-107.

³Ibid., p. 4.

is true of the grin that it could only have been ascribed to a living being, but this is not true of the mouth in the shape of a grin.¹ Second, it is true of the grin that it could only have been ascribed to a person, but this is not true of the mouth in the shape of a grin. The obvious objection here is that in a fairy tale the cat too can grin. There is an obvious reply, however; it is that the grinning cat can also talk. The cat becomes a possible subject of grin-ascriptions by being made to resemble a person.² Third, it is true of the grin that it requires "surroundings," but this is not true of the mouth in the shape of a grin. Thus, the mouth of a grimacing person could be correctly said to be in the shape of a grin in surroundings which rule out the possibility of a grin-ascription.³

Armstrong supports (R₁) by an argument which, he says, "is meant to be an apodeictic disproof of the thesis of indubitability."⁴ The "thesis of indubitability" is the

¹Cf. Wittgenstein, op. cit., secs. 284-285.

²Cf. ibid., secs. 281-282.

³Cf. ibid., secs. 536-539, 583, and 652.

⁴Cf. op. cit., p. 106.

title Armstrong gives to the claim that propositions ascribing phenomenal states are logically indubitable for the persons whose phenomenal states they are, where he defines "p is logically indubitable for A" as follows:

(i) A believes p

(ii) (A's belief that p) logically implies (p).¹

Thus Armstrong's thesis of indubitability is a version of what I have called the "thesis of incorrigibility," got from the latter by eliminating conjuncts (a) and (b) and interpreting conjunct (c) as a logical thesis. The following is his "apodeictic disproof" of the thesis of indubitability.

. . . let us consider the mechanical analogue of awareness of our own mental states; the scanning by a mechanism of its own internal states. It is clear here that the operation of scanning and the situation scanned must be 'distinct existences.' . . . Consider an eye (taken solely as a mechanism) scanning itself by means of a mirror. Certain features of the eye, such as its color and shape, will register on the eye. But the registering will have to be something logically distinct from the features that are registered Why should the substitution of spiritual for material substance abolish the need for a distinction between object and subject? I must admit that I can see no way to prove that there must be such a parallelism, which is a lacuna in my argument. But it seems clear that the natural view to take is that pain and awareness of pain are 'distinct existences.' If so, a false awareness of pain is at least logically possible.²

Armstrong employs here an argument by analogy with self-

¹Ibid., p. 101.

²Ibid., pp. 106-07.

scanning mechanisms based on the logical distinction between object and subject, between the "being acted on," and the "acting on," to establish the universal generalization of step (2) in (H) above, from which he infers the universal generalization of step (3) in (H). The argument is thus an enthymeme whose suppressed premise is step (1) in (H) and which, when fully stated, would take the form of (H). Armstrong, like Smart, sees no need to support step (1). Yet, Armstrong's own example, like Smart's example earlier, seems under closer scrutiny to indicate that step (1) is far from self-evident. If step (1) were self-evident, then granting that the "registering" of color and shape on the eye is logically distinct from the color and shape registered, it would also be self-evident that they were "distinct existences." But much of the philosophy of perception consists in arguments designed to show either that they are "distinct existences" or that they are not. And there are philosophers on opposite sides of the issue who share in common the view that there is some sort of logical distinction between them.

Armstrong's example of an eye scanning itself casts doubt even on the self-evidence of step (2). Clearly, it

is not self-evident in the case of a self-scanning eye--or even in the case of an eye scanning something other than itself--that the color "registered" is related to the activity of the eye as object to subject, as "being acted on" to "acting on." To maintain that it is self-evident is to beg the question against the adverbial view of perception, according to which colors, sounds, smells, etc., are not objects of perception at all but only ways of perceiving. This view is not lightly to be dismissed; its credentials are impressive and go back at least as far as Descartes, who considered colors, sounds, smells, etc. to be no more than "modes" of thought.¹ Now sentences ascribing phenomenal states such as pain are even more plausible candidates for adverbial analysis; indeed, it is part of the adverbial view of perception that feeling pain is just another way of perceiving. If the adverbial view is correct, step (2) of (H) takes only false propositions as values. There is no need on this view for a distinction between, for example, feeling a pain and the pain felt, just as there is no need for a distinction between dancing a jig and the jig

¹See Meditations, III, and Principles of Philosophy, I, 65 (HR, I, 160, 161 and 246).

danced.

The arguments in favor of (R_1) which consist in bringing forward alleged counterexamples to the thesis of incorrigibility, interpreted as a logical thesis, seem to me to suffer from one or the other of two general failings. In the guise of "inferences" from a set of propositions describing conceivable or actual situations, exponents of these arguments are in fact proposing either re-descriptions of phenomena which are presently described in some other way or descriptions for phenomena which are so abnormal that it is at present doubtful how they are to be described. To defend this claim satisfactorily, it would be necessary to examine all such arguments individually. This, of course, would be an impossible task. It will be useful, however, to glance at a few such arguments, to see how they are affected by these failings. Consider the three alleged counterexamples offered by Bruce Aune, a prominent advocate of (R_1) who does not rely on Humean-type arguments, in his recent book, Knowledge, Mind, and Nature.¹ Aune takes his three counterexamples to establish not merely that the thesis of incorrigibility

¹(New York: Random House, 1967), pp. 31-38.

is false if interpreted as a logical thesis, but false even if interpreted as expressing a de facto universality. They establish, he claims, that this thesis is false insofar as it is meant to express anything more than an extremely high probability restricted to normal conditions.¹ Aune bases his first counterexample on "the verbal behavior of hebephrenic schizophrenics," people who "pour out chaotic jumbles of words, which often appear to be utterly unrelated."² He regards the total confusion of such behavior as evidence that the conscious thinking of these people is also totally confused and this, in turn, as evidence that they make mistaken identifications of even their own feelings and mental images. It should be clear, I think, that this argument suffers from the first general failing mentioned above. Chaotic jumbles of thoughts are no more to be counted as the making of identifications, much less the making of mistaking identifications, than are chaotic jumbles of words. In his second argument, Aune describes a possible experiment in which a man gives a self-contradictory report of his present mental image. Upon being asked to read off the image he claims to have from his recent scrutiny of

¹See ibid., p. 37.

²Ibid., p. 33.

the letter-square

e m f

r z a

o w p

the man makes, sincerely and with confidence, the following three assertions:

1. The image did not change during the experiment.
2. From left to right, top to bottom, the letters were: e, m, f, r, z, a, o, w, p.
3. From right to left, bottom to top, the letters were: p, w, o, r, a, s, f, m, e.¹

Aune infers from the inconsistency of this sequence of assertions that the man must be mistaken either about the existence of the image he claims to have or about its character.

Now, if he has the image he claims to have, not all three of these assertions can be true of it; at least one must be mistaken. Whichever it is, we know that he has made a mistake about the character of his experience.²

This argument appears to suffer from the second general failing mentioned above. Clearly, Aune's "inference" is not the only possible one in the circumstances. One could just as well infer that, although each assertion individually can serve as the report of a mental image, the whole sequence of assertions does not constitute the report of anything, much

¹Ibid., pp. 34-35. ²Ibid.

less the mistaken report of a mental image. Any preference for one inference over the other would be entirely arbitrary. The counterexample Aune offers in his third argument is based on some of the experiments actually performed on congenitally blind adults whose vision has been restored by surgery. These patients, it was found, often have an extremely difficult time visually discriminating squares from triangles and even spheres from cubes. Aune infers from the mistakes they make in visually discriminating physical shapes that they must be making mistakes in discriminating their visual impressions.

Though the shapes in point were physical ones, an empiricist could scarcely deny that the patients had the appropriate visual impressions.¹

The failing of this argument is harder to classify. It falls somewhere between the first and second general failings. On the one hand, the reply might be that one could just as well infer that these patients are experiencing chaotic jumbles of visual impressions which only after a period of time first become "appropriate" for discriminating physical shapes. On the other hand, this "inference"

¹Ibid., p. 36.

would seem to be the more natural one to make, in view of the similarity of such phenomena to the symptoms of vertigo.

Consider now the criticism of METI which takes the form of (R_4) . The force of (R_4) is that the thesis of incorrigibility does nothing to show that persons are in a specially favorable epistemic position vis-à-vis their own current phenomenal states, because one of its crucial components, conjunct c, is a bare tautology. (R_4) alone, however, does not imply the lack of epistemic and, hence, metaphysical significance of the thesis of incorrigibility. That sincere reports of something about which it is impossible to be mistaken can nevertheless justifiably be over-ridden does not imply that the statement expressing this impossibility is a bare tautology. Indeed, incorrigibilists have not denied that one's sincere reports of one's own current phenomenal states can justifiably be over-ridden but, on the contrary, have tried to show that they can. Anscombe suggests, as I noted earlier,¹ that although a conflict between the two criteria of behavior and testimony would not show a person to be wrong in what he says about his pain, it would make

¹See supra, pp. 129-30.

what he says unintelligible. This is a point Malcolm makes also.¹ The clearest indication of Malcolm's position regarding (R₄) appears in his essay on after-images in which (R₄) is explicitly affirmed and defended by appeal to several kinds of cases in which it would be justifiable to over-ride first-person present-tense after-image reports.

When I say that [after-image descriptions and reports] are 'incurrigible' I do not imply . . . that anything whatever that is offered as a description of an after-image should be accepted without question. If it were self-contradictory we should not regard it as a perfectly good, although unusual, description of an after-image. If we found that someone constantly misused certain color adjectives in his descriptions of physical realities, then we should not accept at face value his after-image descriptions containing those adjectives. If there was any language at all whose use in relation to physical realities he had not mastered, then we should disallow his use of it in an alleged after-image report.²

That Wittgenstein would also subscribe to (R₄) is clear from his remark that if someone said "'Oh, I know what 'pain' means; what I don't know is whether this, that I have now, is pain'--we should merely shake our heads and be forced to regard his words as a queer reaction which we have no idea to do with."³ The critics of METI who rely on (R₄), however,

¹See "The Privacy of Experience," p. 136.

²"Direct Perception," p. 83.

³Op. cit., sec. 288.

do not intend that (R_4) be considered alone. (R_4) must be considered together with a certain argument in its favor. In the light of this argument, it is maintained, (R_4) does imply the lack of epistemic and, hence, metaphysical significance of the thesis of incorrigibility. The argument in question is a version of what has come to be called the "electroencephalogram (EEG) argument." This version of the EEG argument was originally formulated by Rorty¹ and has recently been revived by Harold Morick.² Since Morick's formulation of the argument adds nothing to Rorty's but on the contrary is only a highly compressed version of it, I will focus my remarks on Rorty's original formulation.

The argument begins with the claim that advances in physiology and electroencephalography which resulted in a well-confirmed theory correlating brain-processes with phenomenal states would lead to a situation in which the

¹"Mind-Body Identity, Privacy, and Categories," pp. 41-48.

²"Is Ultimate Epistemic Authority a Distinguishing Characteristic of the Psychological?," American Philosophical Quarterly, vol. 8 (1971), pp. 292-293. Although Morick tries to show in this article that the EEG argument establishes less than its proponents think it does, he takes it to establish the "merely tautological" character of conjunct c of the thesis of incorrigibility.

"'final epistemological authority' [of sincere first-person present-tense reports of phenomenal states] would be gone, for there would be a standard procedure for over-riding our reports."¹ The next step is a reply to a certain objection which incorrigibilists are expected to raise against this claim. This objection consists in asserting conjunct c of the thesis of incorrigibility together with the claim that conjunct c is incompatible with the possibility of justifiably over-riding sincere first-person present-tense reports of phenomenal states.² The reply to this objection is that a person's sincere reports can justifiably be over-ridden not only if there is good reason to think that they are false but also if there is good reason to think that he does not understand the words he is using, and in the situation envisaged there could be good reason to think that someone does not know how to use phenomenal state terms correctly.³ It is in the course of this reply that an

¹Op. cit., p. 47.

²To expect incorrigibilists to make such an objection, as I have just indicated, shows a fundamental misunderstanding of their position.

³That such a truism should be thought to be an effective reply to the incorrigibilist position I find most puzzling. Incorrigibilists would simply grant it and add that even in the present situation there could be good reason to think that someone does not know how to use phenomenal state terms correctly.

argument appears for the merely tautological character of conjunct c. The argument is restricted to the case of pains but is meant to apply generally to all sensations. The heart of this argument is contained in the following passage:

. . . the claim that 'such a mistake is inconceivable' is an ellipsis for the claim that a mistake, made by one who knows what pain is, is inconceivable, for only this expanded form will entail that when Jones and the encephalograph disagree, Jones is always right. But when formulated in this way out infallibility about our pains can be seen to be empty. Being infallible about something would be useful only if we could draw the usual distinction between misnaming and misjudging, and, having ascertained that we were not misnaming, know that we were not misjudging. But where there are no criteria for misjudging (or to put it more accurately, where in the crucial cases the criteria for misjudging turn out to be the same as the criteria for misnaming), then to say that we are infallible is to pay ourselves an empty compliment.¹

The argument in this passage may be set out as follows:

The statement

S₁: P (a person) cannot be mistaken in thinking that
he has a pain

is elliptical for

S₂: If P knows what "pain" means, P cannot be mistaken in thinking that he has a pain.

Rorty calls S₂ "empty" and implies that it is "useless."

Morick speaks of S₂ as "merely tautological," "analytic,"

¹Ibid., p. 45.

and "trivial." Let

(1) S_2 is trivially true

represent these claims. Now Rorty claims that (1) follows from

(2) There is no distinction between misnaming and misjudging in the case of "I am in pain,"

which, in turn, follows from either

(3) There are no criteria for misjudging in the case of "I am in pain"

or

(4) The criteria for misjudging are the same as the criteria for misnaming in the case of "I am in pain."

(4), Rorty says, is only a "more accurate" version of (3). To support (3) and (4), Rorty describes a case which, he claims, shows that (3) and (4) are both true. In this case someone who has not been burned before exhibits pain-behavior while being burned but sincerely denies that he feels pain. "But, now as in the past, he both exhibits pain-behavior and thinks that he feels pain when he is frozen, stuck, struck, racked, etc."¹ When he is told

¹Ibid., p. 44.

that what he feels when he is burned is also called "pain," he "admits that he does feel something, but insists that what he feels is quite different from what he feels when he is stuck, struck, etc."¹

I want to examine this argument. I will begin by calling into question Rorty's claim that S_1 is elliptical for S_2 . Rorty supports this claim, as the quoted passage shows, by maintaining that S_1 is held because it is thought to entail

S_3 : Whenever P thinks that he has a pain, P is right. S_1 , however, does not as it stands entail S_3 . Only when S_1 is expanded to S_2 is this entailment secured. Rorty is partly right and partly wrong. It is true that S_1 does not entail S_3 , but neither does S_2 entail S_3 . S_2 is clearly consistent with

S_4 : P is wrong in thinking that he has a pain.

For P may not know what "pain" means. Rorty's mistake is one

¹Ibid. Rorty supposes this case to occur at a time when there is a well-confirmed theory correlating brain-processes with pains. The existence of such a theory, however, is an inessential feature of the case. For Rorty, like Morick, does not hold that our inability to be mistaken about our pains is relative to a given stage of empirical inquiry, although, unlike Morick, he holds that our "final epistemological authority" is so relative (see ibid., p. 47).

of bracketing incorrectly. \underline{S}_1 does entail \underline{S}_3 provided that

\underline{S}_5 : \underline{P} knows what "pain" means.

But this should not be construed as asserting

$$(\underline{S}_5 \supset \underline{S}_1) \supset \underline{S}_3$$

but rather

$$\underline{S}_5 \supset (\underline{S}_1 \supset \underline{S}_3).$$

Now the second expression is indeed equivalent to

$$(\underline{S}_5 \cdot \underline{S}_1) \supset \underline{S}_3.$$

So we could say, were we to grant that \underline{S}_1 is held because it is thought to entail \underline{S}_3 , that \underline{S}_1 is elliptical for $(\underline{S}_5 \cdot \underline{S}_1)$. But since $(\underline{S}_5 \cdot \underline{S}_1)$ is clearly not trivial, we could not then go to say that \underline{S}_1 is elliptical for a trivial assertion.

There are other difficulties with Rorty's argument.

It is important to notice that (2) is ambiguous. (2) can be understood to mean either

- (5) The distinction between misnaming and misjudging does not apply to "I am in pain"¹

or

¹Morick understands (2) in this sense. See op. cit., p. 292.

(6) To misname is to misjudge in the case of "I am in pain."¹

Consider (5). Although (5) does follow from either (3) or (4), (1) does not follow from (5). If (1) did follow from (5), one could argue analogously that

(7) "If N (a number) is not red, N cannot be even" is trivially true

follows from

(8) The distinction between being red and being even does not apply to numbers.

Replacing (5) by (6) does not help the argument in this regard and indeed adds to its difficulties. For neither does (1) follow from (6). Otherwise, one could argue that

(9) "If X is not the Evening Star, X cannot be the Morning Star" is trivially true

follows from

(10) The Evening Star is the Morning Star.

Further, (6) does not follow from (3). This form of argument

¹Rorty seems to vacillate between these two ways of understanding (2). His asserting (3) suggests that he understands (2) in the sense of (5), but his asserting (4) suggests that he understands (2) in the sense of (6).

permits the inference from

- (11) There are no criteria for being red in the case of (the number) "2"

to

- (12) To be even is to be red in the case of (the number) "2."

Now (6) does follow from (4). But Rorty's argument in favor of (4) is at best a weak one and at worst irrelevant. One cannot establish that the criteria for the application of a description \underline{d}_1 are the same as the criteria for the application of a description \underline{d}_2 by citing cases in which the criteria do not determine which of \underline{d}_1 or \underline{d}_2 is to be preferred. The case Rorty presents us with, however, does not even seem to be of this sort but rather one in which the phenomena are so abnormal that it is doubtful whether either \underline{d}_1 or \underline{d}_2 has an application or one in which, since there is a conflict of criteria (for being in pain), no description has an application other than "the phenomena are unintelligible."

CHAPTER V

INCORRIGIBILITY AND THE EEG ARGUMENT

One of the ways in which proponents of the Mind-Body Identity Theory have replied to objections based on the thesis of incorrigibility is to challenge the significance of this thesis. The four general positions I have called R_1 - R_4 are different forms this challenge has taken. R_1 - R_3 are supported by Humean arguments, arguments by analogy, and alleged counterexamples to the thesis of incorrigibility and R_4 by the electroencephalogram (EEG) argument. In Chapter IV, I examined these arguments and alleged counterexamples and concluded that they fell short of establishing the positions they support. A second line of reply to objections based on the thesis of incorrigibility is that it would be reasonable in the light of future scientific advances to eliminate in principle incorrigibility in the sense in which it cannot be a feature of physiological statements. The general positions I have called R_5 and R_6 are different forms of this reply. Identity Theorists rely almost entirely on the EEG argument to make out their case for R_5 and R_6 . In the present chapter, I will try to show that the EEG argument does not establish either of these positions or even

provide some reason for thinking that one of them is true. There will be no need to consider these positions separately, for my argument is intended to show that the EEG argument does not establish or even provide some reason for thinking true one of the two claims essential to both positions, viz., that additions to our stock of empirical knowledge together with advances in technology could lead to the elimination in principle of the feature of incorrigibility.¹ I will conclude my discussion by considering whether any argument could establish or provide some reason for thinking that this claim is true. Respecting this question, I will argue that the logical differences between explanation-types pointed out in Chapters II and III which make doubtful the possibility of eliminating sensation-statements also make doubtful the possibility of eliminating incorrigibility as a feature of these statements. Heretofore the issue of incorrigibility has been debated as if it were decidable independently of considerations pertaining to the logical similarities and differences between explanation-types. No really important philosophical issue can be decided, however, independently

¹For an account of the claims common to both R_5 and R_6 and peculiar to each, see supra, Chapter IV, pp. 113-14. ⁶

of the way in which a great many other philosophical issues are decided. One of my aims is to show that incorrigibility is such an issue.

As things are, say proponents of the EEG argument, no one can be in a position to override a person's honestly avowed reports about the existence and character of his own current sensations. That no one can be in such a position, however, is merely contingent on the present state of empirical inquiry. There are no accepted procedures by applying which it would be rational to override a person's honest avowal of sensation. Advances in physiology and technology, however, could lead to a well-confirmed theory correlating sensations with brain processes. In such circumstances, it is claimed, it would be rational to override a person's honest avowal of sensation on the basis of a conflicting EEG report (provided there was not a sizable accumulation of conflicts between similar avowals of other people and EEG reports). To say that it would be rational to override such an avowal, of course, does not rule out the possibility that the avowal might be correct after all.

There are serious difficulties with the foregoing position. I want to discuss two such difficulties. The first

is not serious in itself but nevertheless deserves some attention, since it is connected with what is, I think, an obviously mistaken view of our epistemic position vis-à-vis our own current sensations. The second deserves a great deal more attention and will occupy most of the discussion in this chapter. The first difficulty is this. It is not true that as things are no one can be in a position to override a person's honestly avowed reports about the existence and character of his own current sensations. There are criteria whose satisfaction establishes that a word is being used incorrectly, and anyone who is in a position to observe that these criteria have been satisfied on a certain occasion is in a position to override the report made by using the word on that occasion. It might be thought that this difficulty can be easily removed by adding a proviso to the original assertion: As things are no one can be in a position to override a person's honestly avowed reports about the existence or character of his own current sensations, if that person is granted mastery of the words used to make these reports. The trouble with this suggestion, however, is that some proponents of the EEG argument hold the view that anyone who grants a person mastery of the words he uses

to make honestly avowed reports of his own current sensations is logically precluded from rejecting or even doubting these reports, because it is "merely tautological" that a mistake about one's own current sensations made by one who knows the meaning of the words for these sensations is inconceivable.

. . . 'such a mistake is inconceivable' is elliptical for 'such a mistake made by one who knows what 'pain' means is inconceivable.' And that assertion is merely tautological¹

If these proponents of the EEG argument are right, the assertion resulting from adding the proviso "if that person were granted mastery of the words used to make these reports" expresses not only how things are but how they always were and always will be. For surely if our present epistemic position vis-à-vis our own current sensations is grounded in a mere tautology, no changes in the world whatever could lead to any change in this position. But the purpose of the EEG argument is to show how some changes in the world could

¹Harold Morick, "Is Ultimate Epistemic Authority a Distinguishing Characteristic of the Psychological?," American Philosophical Quarterly, vol. 8 (1971), p. 292. See also Richard Rorty, "Mind-Body Identity, Privacy, and Categories," Review of Metaphysics, vol. 19 (1965), pp. 45-46. Although Morick claims not to be a proponent of the EEG argument, he accepts the version of this argument sketched above.

lead to a change in this position. So the proponents of the EEG argument who hold it to be merely tautological that a mistake about one's own current sensations made by one who knows the meaning of the words for these sensations is inconceivable seem to be caught in a dilemma. Either the EEG argument is intended to show how someone could be in a position to override a person's honestly avowed reports about his own current sensations, in which case it is unnecessary. For as things are someone could be in such a position. Or it is intended to show how someone could be in a position to override a person's honestly avowed reports about his own current sensations even when that person is granted mastery of the words used to make these reports, in which case the argument is useless. For such a position is logically impossible in the sense that the statement describing it is self-contradictory.

The argument against the second horn of the dilemma has evidently appeared so compelling as to outweigh the considerations against the first, for the philosophers in question are content to put forward the EEG argument as one intended to show how someone could be in a position to override a person's honestly avowed reports about his own

current sensations. Now it seems on the face of it not self-contradictory to assert that someone could be in a position to override a person's honest avowal of his own current sensations even when that person is granted mastery of the language. And, indeed, I think it is this view which must be rejected to preserve the significance of the EEG argument. This view, as we have seen, rests on the claim that it is merely tautological that a mistake about one's own current sensations made by one who knows the meaning of the words for these sensations is inconceivable. What is the argument for this claim? It is that the distinction between verbal and factual error does not apply to a person's honest avowals of his own current sensations.

. . . since the distinction between verbal and factual error, between misnaming and misjudging, doesn't apply to 'I am in pain,' it follows analytically that if a person knows what 'pain' means he cannot mistake pain for something else nor something else for pain.¹

Now this by itself does not establish the correctness of the claim. The argument would be valid only if it were also merely tautological that the distinction between verbal and factual error does not apply to a person's honest avowals of his own current sensations. But I do not see

¹Morick, op. cit., p. 292. See also Rorty, op. cit. pp. 43-44.

that it is. The argument given for the lack of such a distinction is that there are no criteria whose satisfaction would establish that someone using a sensation-word correctly on a certain occasion to make an honest report about his own current sensation is mistaken or, as it is sometimes put, that the criteria whose satisfaction would establish that someone who makes a report about his own current sensation on a certain occasion is mistaken are the same as the criteria whose satisfaction would establish that the relevant sensation-word is being used incorrectly on that occasion.

. . . where there are no criteria for misjudging (or to put it more accurately, where in the crucial cases the criteria for misjudging turn out to be the same as the criteria for misnaming), then to say that we are infallible is to pay ourselves an empty compliment.¹

But clearly it is not merely tautological that there are no criteria for factual error in the case of honest reports of one's own current sensations or that the criteria for factual error turn out in this case to be the same as the criteria for verbal error. These are contingent facts. That the distinction between verbal and factual error does

¹Rorty, ibid., p. 45.

not apply to a person's honest avowals of his own current sensations, therefore, is also a contingent fact and so cannot legitimately be used to support the claim that it is merely tautological that a mistake about one's own current sensations made by one who knows the meaning of the words for these sensations is inconceivable.

It may appear that the purpose of the foregoing argument is to show that it is conceivable that a mistake about one's own current sensations made by one who knows the meaning of the words for these sensations is conceivable. In a way this is right. But it should be noticed that in this reformulation of the intended conclusion, the word "conceivable" is being used to express different notions in its two occurrences. The intended conclusion could be restated in the following way so as to bring out the ambiguity: It is not self-contradictory for someone to be in a position to override a person's honestly avowed reports about his own current sensations even when that person is granted mastery of the words used to make these reports. That no one can be in such a position is based on the contingent fact that there are no criteria whose satisfaction would establish that someone using a sensation word correctly on a

certain occasion to make an honest report about his own current sensation is mistaken.

With this shift from "merely tautological" to "merely based on a contingent fact," the EEG argument acquires not only significance but also force and persuasiveness. If it is a contingent fact that there are no criteria of the kind in question, then surely there could come to be such criteria. And it is the purpose of the EEG argument to describe circumstances in which it would be reasonable to introduce them. I do not think, however, that the circumstances described by the EEG argument are ones in which it would be any more reasonable to introduce such criteria than the present circumstances. That our epistemic position vis-à-vis our own current sensations is merely based on a contingent fact does not imply that this position is merely contingent on the present state of empirical inquiry. And there is a serious difficulty with the view that our epistemic position vis-à-vis our own current sensations is merely contingent on the present state of empirical enquiry, which becomes apparent upon an examination of the EEG argument. The difficulty is this. According to the EEG argument, in circumstances in which there is a well-confirmed theory

correlating sensations with brain processes, it would be rational to override a person's honest avowal of sensation on the basis of a conflicting EEG report (provided there was not a sizeable accumulation of conflicts between similar avowals of other people and EEG reports). The question immediately arises: Why should we not rather conclude that the theory is defective? Even a well-confirmed set of psychophysical correlation laws must be open to empirical disconfirmation. It is indeed the recognition of this point that lies behind the proviso that "there was not a sizeable accumulation of conflicts between similar avowals of other people and EEG reports." The requirements that correlation laws be open to empirical disconfirmation seems also to underlie Morick's thesis that a person's honest avowal of his own current psychological state cannot possibly be overridden by any evidence which is not supported by other such avowals from a sufficient number of overall competent speakers of the language in the same circumstances.¹ But I do not see why nothing can count as a disconfirmatory instance of a correlation law unless it is accompanied by other disconfirmatory instances. If a certain piece of metal were discovered

¹See op. cit., pp. 294-95.

to have all the known properties of copper except high electrical conductivity, it would surely not be irrational to count it as a disconfirmatory instance of the law "Copper is a good electrical conductor." It would be possible, of course, to decide not to count the piece of metal as copper, but there seems to be no good reason for holding such a decision to be more rational than, say, the decision to count it as a new kind of copper. But, it may be objected, the psycho-physical correlations of the EEG argument are not mere empirical generalizations like, "Copper is a good electrical conductor." They are part of a well-confirmed theory which explains and exhibits systematic connections between a wide variety of empirical generalizations about inner states and thus deals with a more extensive range of phenomena than does any given empirical generalization. The theory, moreover, is one from which more precise predictions concerning inner states may be obtained and continually suggests new empirical laws which subsequently prove to be correct. Now there is good reason, the objection concludes, to hold that a theory so central to current scientific explanations cannot be broken down at one blow. But why should it be thought to follow from this that a correlation which figures as part of the theory cannot be broken

down at one blow? That a theory is central to current scientific explanations does not imply that every part of the theory is also central to current scientific explanations. And a theory of such great generality as is here envisaged has little to fear from the acknowledgement that it is defective in one of its parts. It would have something to fear, of course, from the acknowledgement of a defect which involved the breakdown of a fundamental assumption of the theory. But clearly a law correlating, say, pain with a certain kind of brain process would not be a fundamental assumption of this theory. Nor need the falsification of such a law involve the breakdown of a fundamental assumption. A law correlating pain with a certain kind of brain process could be false if the set of initial or boundary conditions mentioned in its antecedent was not sufficient to exclude feelings closely resembling pain. And it might in fact be the case that the person whose honest avowal conflicts with an EEG report is the only person who ever had such feelings. The recognition of such a case need not lead to the breakdown of the theory or one of its fundamental assumptions, although it would prompt a search after the relevant difference or differences between the anomalous case and all

other cases so far observed.

It may be objected that in any isolated case of conflict between a person's honest avowal and an EEG report in the circumstances envisaged by the EEG argument, we could conclude that the person is making a verbal mistake. Regarding the expressions "pain" and "seems to me as if I were seeing something red," Rorty writes:

. . . the device which we should use to justify ourselves [in overriding sincere reports using these predicates in certain imaginable situations]--is one which can apply in all [Rorty's italics] proposed cases this escape-hatch is always available, and . . . the question of whether the reporter does know how to use the word or does not is probably not itself a question which could ever be settled by recourse to any absolute epistemological authority. . . .¹

This objection seems confused. In Rorty's statement of the objection, I think it important to ask what he means by "this escape-hatch is always available." Does he mean simply that the device to which he alludes is one which can apply in all proposed cases of introspective reports? If so, then he is surely right. But this does not imply that it would be rational to apply the device in any proposed case of this kind. Does he mean then that it would be

¹Op. cit., p. 51.

rational to apply the device in any proposed case of this kind? If so, then he is wrong. For clearly there are some cases of introspective reports in which it would be not merely less rational to apply the device than not to apply it, but not rational to apply the device at all. Such cases are ones in which it has been established with certainty that the reporter does know how to use the relevant words. And surely there are such cases. I do not know what Rorty means by an "absolute epistemological authority" by recourse to which the question of whether the reporter does know how to use the relevant words or does not could be settled. But there are criteria whose satisfaction would establish with certainty that a person does understand the words he is using to make an introspective report. The satisfaction of these criteria, of course, does not rule out the logical possibility that the reporter does not know how to use the relevant words, and it may even be established with certainty later that the reporter did not in fact know how to use the relevant words. If by "absolute epistemological authority" Rorty means an authority to which one could appeal to rule out these possibilities, I agree that there is none. But such possibilities provide no justification

whatever for the claim in any given case that someone making an introspective report does not know how to use the relevant words. Even if we were to regard Rorty's statement "this escape-hatch is always available" as an ellipsis for "this escape-hatch is always available in any isolated case of conflict between an introspective report and an EEG report in the circumstances envisaged by the EEG argument," the above criticism would still apply. For there are some imaginable cases of this kind in which there would be no justification whatever for the claim that the reporter does not know how to use the relevant words. Such a case, for example, would be one in which someone who does not exhibit pain-behavior and has a long history of the correct use of the word "pain" reports that he has no pain, but an EEG report says that the brain process correlated with pain did occur. To reply that in any such case the "a priori improbability" of "the body of current scientific theory foundering upon the rock of a single over-riding report"¹ gives some justification to the claim that the reporter does not know how to use the relevant words would be, as I have argued, to appeal

¹Rorty, ibid., p. 51.

to a highly exaggerated picture of the situation.

The circumstances envisaged by the EEG argument, accordingly, do not preclude the occurrence of isolated cases of conflict between an honest first-person present-tense sensation-report and an EEG report in which it would be rational to conclude that the reporter does know how to use the relevant words. This does not imply, however, that the existence of a well-confirmed theory correlating sensations with brain processes would provide any special reason for introducing criteria whose satisfaction would establish that someone using a sensation-word correctly on a certain occasion to make an honest report about his own current sensation is mistaken. And, if the foregoing argument is correct, such a theory would not provide any special reason for introducing these criteria. A given kind of sensation is correlated with a given kind of brain process only under certain conditions. A conflict between an honest avowal of sensation by someone who is granted mastery of the language and an EEG report would prompt a search after the relevant difference or differences between the anomalous case and all other cases so far observed. And in fact no relevant difference may be found. But why should this be thought to show or

provide any reason for saying that the person whose avowal it is has made a mistake? As things are, our criteria for the existence and character of sensations consist in the occurrence of certain kinds of behavior. But the occurrence of a given kind of behavior is our criterion for the existence of a given kind of sensation only in certain circumstances. A conflict between an honest avowal of sensation by someone who is granted mastery of the language and his (non-verbal) behavior would prompt a search after something in the circumstances which made them abnormal. Has he been hypnotized,¹ drugged, electrified, etc.? Is his body subject periodically to unexpected and uncontrollable convulsions, etc.? And in fact nothing unusual may be found. But, as things are, this would not show or provide any

¹Morick argues that if we judge that a hypnotized man who says, "I am in pain" is in fact not in pain, "then ipso facto we judge that his hypnotic 'I am in pain' fails to display knowledge of the correct use of this sentence for avowing pain" (op. cit., p. 294). For reasons I shall not enter into here, I think his argument incorrect, but even if it were correct, it would not follow that if we judge that a hypnotized man who says, "I am in pain" is in fact not in pain, then ipso facto we judge that he does not know how to use the word "pain." For the judgment that a man fails to display knowledge does not carry the implication that he lacks the knowledge he fails to display. And there are criteria whose satisfaction would establish with certainty that a hypnotized man whom we judge not in fact to be in pain when he says, "I am in pain" knows how to use the word "pain."

reason for saying that the person whose avowal it is has made a mistake. The failure to find something in the circumstances which made them abnormal would be a reason either for looking further or for throwing up one's hands in despair of ever making sense of this piece of behavior. The analogy between the way things are and the way things would be given the existence of a well-confirmed theory correlating sensations with brain processes partially breaks down here. To throw up one's hands in despair when faced with an apparent disconfirmatory instance of a well-confirmed law does not manifest a scientific attitude. A scientifically respectable response, however, whose only purpose is to forestall an open declaration of despair is that of protecting the law with ad hoc assumptions. It is true that such "protection" very quickly approaches the limits of scientific respectability.¹ Even when these limits are reached, however, one need not abandon the law. One can simply cease to protect it. If the law is fundamental to current scientific explanations, it would not be irrational to continue to employ it

¹Cf. D. M. Armstrong, A Materialist Theory of the Mind (London: Routledge & Kegan Paul, 1968), p. 110.

for explanation and prediction while acknowledging the existence of evidence against it. Such an attitude toward the law would resemble the present attitude of a large part of the scientific community toward the wave and particle interpretations of quantum mechanics. On the other hand, if the law is not fundamental to current scientific explanations, these explanations would have little to fear from its abandonment.

That the circumstances envisaged by the EEG argument are circumstances in which it would be reasonable to introduce criteria whose satisfaction would establish that someone using a sensation-word correctly on a certain occasion to make an honest report about his own current sensation is mistaken, accordingly, cannot be shown by the EEG argument. If this cannot be shown by the EEG argument, however, there is good reason to think that this cannot be shown at all. For it would be reasonable to introduce such criteria only if it already made sense for a person to be mistaken, say, as to whether he is in pain. But this makes sense only if it also makes sense for a person to be in doubt as to whether he is in pain. And it is senseless for a person to be in doubt as to whether he is in pain where it is senseless

for a person to find out whether he is in pain. But it is senseless for a person to find out whether he is in pain. For consider the possible ways in which, it might be said, a person could find out whether he is in pain. It might be said that a person could find out whether he is in pain by applying to himself a physiological criterion of pain. But it would make sense for a person to apply to himself a physiological criterion of pain only if such a criterion had been introduced. Since ex hypothesi it would not be reasonable to introduce such a criterion unless it already made sense for a person to be mistaken as to whether he is in pain, however, it would not be reasonable to introduce one unless it already made sense for a person to find out whether he is in pain. Or it might be said that a person could find out whether he is in pain by observing his pain. But there is no difference between observing one's own pain and having it. What this proposal comes to then is that a person could find out whether he is in pain from his pain, that is, simply by having or not having it.¹ Now there is a (possible)

¹This proposal is defended by Gregory Sheridan. See his article, "The Electroencephalogram Argument against Incurability," American Philosophical Quarterly, vol. 6 (1969), p. 69.

sense of "find out" in which it is true that a person does find out whether he is in pain by having or not having a pain, but this is not the sense of "find out" in which it must make sense for a person to find out whether he is in pain if it is to make sense for a person to be in doubt as to whether he is in pain. A person finds out whether he is in pain by having or not having a pain in the sense that having or not having a pain puts him in a position and gives him the right to say that he is in pain or that he is not in pain.¹ It is senseless, however, for a person to be in doubt about what he can find out in this sense. If it is to make sense for a person to be in doubt as to whether he is in pain, it must make sense for him to find out whether he is in pain in the sense in which he can find out, say, the position of his limbs, that is, by a kind of observing which goes beyond the mere having of what is being observed. A person does not find out that his arm is straight in the sense that his arm's being straight puts him in a position and gives him

¹Cf. A. J. Ayer, "Privacy," in his The Concept of a Person (New York: St. Martin's Press, 1963), p. 64; Sydney Shoemaker, Self-Knowledge and Self-Identity (Ithaca, N.Y.: Cornell University Press, 1963), p. 216; William Alston, "Varieties of Privileged Access," American Philosophical Quarterly, vol. 8 (1971), p. 234.

the right to say that his arm is straight. For it is not senseless for someone honestly to say that his arm is bent when it is straight. Normally, a person knows whether his arm is straight without finding out whether his arm is straight. But it does make sense for a person to find out whether his arm is straight (and indeed it must make sense, since a person can be mistaken as to whether his arm is straight)--not by observing his kinesthetic sensations, which is a kind of observing, if it may be called an "observing" at all, that does not go beyond the mere having of what is being observed, but by looking.¹

Consider, then, whether it does make sense for a person to find out whether he is in pain by observing whether he is in pain in the sense of "observing" which goes beyond the mere having of what is being observed. The only plausible

¹This seems to be Malcolm's point in the passage from "The Privacy of Experience" (in Avrum Stroll (ed.), Epistemology (New York, Harper and Row, 1967), pp. 147-151) which Sheridan is criticizing when he claims that a person finds out whether he is in pain by having or not having a pain. Sheridan asks: "Why would I have to observe my own behavior and listen to my words in order not to be in doubt about my sensations? Would I not still know my sensations as a result of directly experiencing them?" (op. cit., p. 69). The point, however, is not that the introduction of first-person doubts into the language of sensations commits us to the view that a person would have to observe

object of observation in this sense, I believe, is a person's pain-behavior. The proposal then is that a person finds out whether he is in pain by observing his own pain-behavior.¹ Now how could a person find out whether he is in pain in this sense of "find out?" As Wittgenstein remarks it would be necessary for him to take notion of himself as others do, to listen to himself talking, to be able to draw conclusions from what he says.² But if this were possible

his own behavior and listen to his own words in order not to be in doubt about his sensations, but rather that it would have to make sense for a person to observe his own behavior and listen to his own words in order to find out what his sensations are.

¹Although it is not really clear, this proposal seems to be defended by Rorty in "Wittgenstein, Privileged Access, and Incommunicability," American Philosophical Quarterly, vol. 7 (1970), p. 202. Rorty speaks of a person's utterances and inclinations or dispositions to utter certain first-person reports such as, "I am in pain" as his evidence for the truth of the proposition that he is in pain. He also speaks, however, of a person's belief that he is in pain as his evidence for the truth of the proposition that he is in pain. This suggests that it is not by observing one's own pain-behavior that one finds out whether one is in pain. Yet, by "belief" he seems to mean "the expression of belief." This is suggested by his remark ". . . other people certainly take my beliefs about my mental states as evidence for their own beliefs about my mental states. So why shouldn't I?" In any case, his defense of the view set forth in this passage is simply that its denial carries the "paradoxical" implication that there is no genuine use of the verb "to know" as an expression of certainty with first-person present-tense sensation statements.

²See Philosophical Investigations, trans. G. E. M. Anscombe (New York: Macmillan, 1953), p. 192.

it would make sense to say such things as "Judging from what I say, I must be in pain" and "It seems to me that I am not in pain, but, judging from what I say, this isn't true." "Monologues" of the sort imagined by Malcolm would also make sense: "I shall sometimes ask myself, 'Does it hurt?' If I reply, 'Yes, it does,' then I shall know that I have pain, my evidence being that I myself said so!"¹ But, further, others notice and draw conclusions from not only what a person says but also what a person does. Pain-behavior is what pain-explanations explain. Declining an invitation, making an appointment to see the dentist, hiring someone to do the heavy work involved in running one's business, even a person's selling his business and marrying someone who he knows can be depended on to take care of him are sometimes instances of pain-behavior.² If, then, a person could take notice of himself and draw conclusions from his own behavior as others do, it would make sense to say such things as "Judging from the fact that I am trying to sell my business and have proposed to Mary, I must be in pain" and "It seems to me that I am not in pain, but, judging from the fact that I

¹Op. cit., p. 154.

²Cf. supra, chap. II, pp. 37-41.

am trying to sell my business and have proposed to Mary, this isn't true." But to say such things does not make sense. They are senseless in the sense that we would not understand the person who said them. Wittgenstein remarks: "If a lion could talk, we could not understand him."¹

The reason why we could not understand him is not that we could not understand what he said. We could not understand him, because his verbal behavior would diverge too far from the normal behavior of beings to whom we ordinarily apply the concept of a person. Whatever the lion said would be senseless in the sense that we would never understand him. It is of course false that we would never be able to understand a person who exhibited the sort of verbal behavior we have been imagining. But this behavior would have the same sort of unintelligibility that would for us be a feature of the verbal behavior of any being not falling under the concept of a person.

¹Op. cit., p. 223. Cf. supra, chap. II, pp. 54-55.

SELECTED BIBLIOGRAPHY

Alston, William, "Varieties of Privileged Access," American Philosophical Quarterly, VIII (1971), pp. 223-41.

Anscombe, G. E. M., Intention (Ithaca, New York: Cornell University Press, 1966).

Aristotle, Metaphysics.

_____, De Anima.

_____, Nicomachean Ethics.

_____, Posterior Analytics.

Armstrong, D. M., A Materialist Theory of the Mind (New York: Humanities Press, 1968).

Aune, Bruce, Knowledge, Mind, and Nature. (New York: Random House, 1967).

Austin, J. L., How to Do Things with Words (New York: Oxford University Press, 1965).

Ayer, A. J., "Can There Be a Private Language?," in his The Concept of a Person and Other Essays (New York: St. Martin's Press, 1963), pp. 36-51.

_____, "Privacy," ibid., pp. 52-81.

Baier, Kurt, "Smart on Sensations," Australasian Journal of Philosophy, XL (1962), pp. 57-68.

Bernstein, Richard J., "The Challenge of Scientific Materialism," International Philosophical Quarterly, VIII (1968), pp. 252-75.

Chihara, C. S. and J. A. Fodor, "Operationalism and Ordinary Language: A Critique of Wittgenstein," American Philosophical Quarterly, II (1965), pp. 281-95.

Coburn, Robert C., "Persons and Psychological Concepts," American Philosophical Quarterly, IV (1967), pp. 208-21.

Cook, John W., "Human Beings," in Peter Winch (ed.), Studies in the Philosophy of Wittgenstein (New York: Humanities Press, 1969), pp. 117-51.

_____, "Wittgenstein on Privacy," Philosophical Review, LXXIV (1965), pp. 281-314.

Cornman, James W., "Mental Terms, Theoretical Terms, and Materialism," Philosophy of Science, XXXV (1968), pp. 45-63.

_____, "On the Elimination of 'Sensations' and Sensations," Review of Metaphysics, XXII (1968), pp. 15-35.

_____, "The Identity of Mind and Body," Journal of Philosophy, LIX (1962), pp. 486-92.

Descartes, Meditations.

_____, Principles of Philosophy.

_____, The Passions of the Soul.

Feyerabend, P. K., "Materialism and the Mind-Body Problem," Review of Metaphysics, XVII (1963), pp. 49-66.

_____, "Mental Events and the Brain," Journal of Philosophy, LX (1963), pp. 295-96.

Hintikka, Jaako, "Cogito, Ergo Sum: Inference or Performance?," Philosophical Review, LXXI (1962), pp. 3-32.

Hume, David, A Treatise of Human Nature.

Malcolm, Norman, "Behaviorism as a Philosophy of Psychology," in T.W. Wann (ed.), Behaviorism and Phenomenology (Chicago: University of Chicago Press, 1964), pp. 141-55.

_____, "Explaining Behavior," Philosophical Review, LXXVI (1967), pp. 97-104.

_____, "Direct Perception," in his Knowledge and Certainty (Englewood Cliffs, New Jersey: Prentice-Hall, 1963), pp. 73-95.

Malcolm, Norman. Dreaming (London: Routledge & Kegan Paul, 1959).

_____, "Scientific Materialism and the Identity Theory," Dialogue, III (1964), pp. 115-25.

_____, "The Conceivability of Mechanism," Philosophical Review, LXXVII (1968), pp. 45-72.

_____, "The Privacy of Experience," in Avrum Stroll (ed.), Epistemology: New Essays in the Theory of Knowledge (New York: Harper and Row, 1967), pp. 129-58.

Meehl, P. E., "The Compleat Autocerebroscopist; A Thought-Experiment on Professor Feigl's Mind-Body Identity Thesis," in P. K. Feyerabend and G. Maxwell (eds.), Mind, Matter, and Method (Minneapolis, University of Minnesota Press, 1966), pp. 103-30.

Morick, Harold, "Is Ultimate Epistemic Authority a Distinguishing Characteristic of the Psychological?," American Philosophical Quarterly, VIII (1971), pp. 292-95.

Place, U. T., "Is Consciousness a Brain Process?," British Journal of Psychology, XLVII (1956), pp. 44-50.

Putnam, Hilary, "Brains and Behavior," in R. J. Butler (ed.), Analytical Philosophy, Second Series (Oxford: Basil Blackwell, 1965), pp. 1-19.

_____, "Dreaming and 'Depth Grammar,'" in R. J. Butler (ed.), Analytical Philosophy, First Series (Oxford: Basil Blackwell, 1962), pp. 211-35.

_____, "Minds and Machines," in Sidney Hook (ed.), Dimensions of Mind (New York: New York University Press, 1969), pp. 138-61.

_____, "The Analytic and the Synthetic," in H. Feigl and G. Maxwell (ed.), Minnesota Studies in the Philosophy of Science, III (Minneapolis: University of Minnesota Press, 1962), pp. 358-97.

Rorty, Richard, "Incorrigibility as the Mark of the Mental," Journal of Philosophy, LXVII (1970), pp. 399-424.

- Rorty, Richard, "In Defense of Eliminative Materialism," Review of Metaphysics, XXIV (1970), pp. 112-21.
- _____, "Mind-Body Identity, Privacy, and Categories," Review of Metaphysics, XIX (1965), pp. 24-54.
- _____, "Wittgenstein, Privileged Access, and Incommunicability," American Philosophical Quarterly, VII (1970), pp. 192-205.
- Ryle, Gilbert, The Concept of Mind (New York: Barnes and Noble, 1949).
- Sellars, Wilfrid S., "The Identity Approach to the Mind-Body Problem," Review of Metaphysics, XVIII (1965), pp. 430-51.
- Sheridan, Gregory, "The Electroencephalogram Argument Against Incorrigibility," American Philosophical Quarterly, VI (1969), pp. 62-70.
- Shoemaker, Sydney, Self-Knowledge and Self-Identity (Ithaca, New York: Cornell University Press, 1963).
- Smart, J. J. C., "Brain Processes and Incommensurability: A Reply to Professor Baier," Australasian Journal of Philosophy, XL (1962), pp. 68-70.
- _____, Philosophy and Scientific Realism (New York: Humanities Press, 1963).
- _____, "Sensations and Brain Processes," Philosophical Review, LXVIII (1959), pp. 141-56.
- Strawson, P. F., Individuals (Garden City, New York: Doubleday Anchor Books, 1963).
- _____, untitled review of Wittgenstein's Philosophical Investigations, Mind, LXIII (1954), pp. 70-99.
- Wittgenstein, Ludwig, Blue and Brown Books (Oxford: Basil Blackwell, 1958).
- _____, Philosophical Investigations, trans. G. E. M. Anscombe (1st ed.: Oxford: Basil Blackwell, 1953).
- _____, Zettel, trans. G. E. M. Anscombe (Oxford: Basil Blackwell, 1967).

APPROVAL SHEET

The thesis/dissertation submitted by Gordon Cohen
has been read and approved by members of the Department
of Philosophy.

The final copies have been examined by the director of
the thesis/dissertation and the signature which appears
below verifies the fact that any necessary changes have
been incorporated and that the thesis/dissertation is
now given final approval with reference to content and
form.

The thesis/dissertation is therefore accepted in partial
fulfillment of the requirements for the degree of Doctor
of Philosophy.

April 13, 1973
DATE

Edward A. Mazion
ADVISOR'S SIGNATURE