USE OF THESES

This copy is supplied for purposes of private study and research only. Passages from the thesis may not be copied or closely paraphrased without the written consent of the author.
This thesis is my own work written partly while I was a research scholar in the Department of Philosophy of the Research School of Social Sciences at The Australian National University.

John Kekes
There is supposed to be a gulf between mind and matter, and a mystery which it is held in some degree impious to dissipate. I believe, for my part, that there is no greater mystery than there is in the transformation by the radio of electromagnetic waves into sounds. I think the mystery is produced by a wrong conception of the physical world and by a Manichean fear of degrading the mental world to the level of the supposedly inferior world of matter

B. Russell
My Philosophical Development, p. 22.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Synopsis</td>
<td></td>
<td>v</td>
</tr>
<tr>
<td>Chapter One</td>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>Chapter Two</td>
<td>The Statement of the Identity Theory</td>
<td>23</td>
</tr>
<tr>
<td>Chapter Three</td>
<td>The Nature of the Identity Theory</td>
<td>39</td>
</tr>
<tr>
<td>Chapter Four</td>
<td>The Corrigibility of Avowals</td>
<td>74</td>
</tr>
<tr>
<td>Chapter Five</td>
<td>The Groundedness of Avowals</td>
<td>118</td>
</tr>
<tr>
<td>Chapter Six</td>
<td>The Spatial Localizability of Sensations</td>
<td>162</td>
</tr>
<tr>
<td>Chapter Seven</td>
<td>The Problem of Identity</td>
<td>200</td>
</tr>
<tr>
<td>Chapter Eight</td>
<td>The Identity Theory, Dualism, and the Criteriological View</td>
<td>230</td>
</tr>
<tr>
<td>Bibliography</td>
<td></td>
<td>263</td>
</tr>
</tbody>
</table>
SYNOPSIS

Chapter One presents one formulation of the problem which different versions of the identity theory attempt to solve. The attempted solutions offered by other versions of the identity theory are evaluated.

Chapter Two is a statement of the identity theory. It restricts the discussion to the identity of sensations and brain processes. It tries to clarify what is to count as a sensation and what as a brain process. Some introductory remarks are made in it about the nature of the concept of identity.

Chapter Three is concerned with distinguishing the factual and logical (in the widest sense of these terms) components of the identity theory. It is argued that they cannot be treated in isolation of each other. The independent identification of sensations through nonverbal behaviour, and of brain processes through brain-readings is discussed.

Chapters Four and Five are devoted to a discussion of avowals. It is argued that sensations can be identified through avowals as well. The nature of avowals is discussed; it is argued that the typical avowal is corrigible, and that
avowals have grounds. Alternative analyses of avowals are discussed.

Chapter Six takes up the question whether or not sensations can be said to have spatial location. It is argued that an affirmative answer can be given if we recognize that sensations, like all processes, can be said to have spatial location only in a derivative sense.

Chapter Seven examines the nature of the identity between sensations and brain processes. Identity is argued to be an empirical, extensional, heterogeneous, reductive identity of properties.

The task of Chapter Eight is twofold: first, to meet the objection that the identity proposed is too weak—and thus compatible with the distinctness of sensations and brain processes; and second, to compare the identity theory with dualism and with the criteriological view.
CHAPTER ONE

INTRODUCTION

(1) Of the many dichotomies used for characterizing what there is, I shall select one for the purpose of stating what the identity theory is trying to do. The dichotomy I have in mind is between whatever can be known immediately, by acquaintance, by introspection, and whatever can be known only mediately by description, by observation. Typical objects\(^1\) said to be knowable in the first way are one’s own thoughts, feelings, sensations, images, and the like, while typical objects supposedly knowable in the second way are material objects, animals, other human beings, and so on.

One of the most important characteristics that objects knowable in the first way are supposed to have is that their existence depends upon being the objects of awareness of a person. Objects knowable in the second way are said to be capable of existing regardless of whether or not anyone is aware of them.

A convenient way of labelling the slice of the

---

\(^1\)I am using "object" in the widest possible sense; the range of the word is meant to cover entities, processes, events, states, dispositions, episodes, and so on.
universe containing objects knowable in the first way is 'mental,' while the section of the universe supposedly containing objects knowable only in the second way may be labeled 'physical.' Thoughts, feelings, sensations, etc., belong to the 'mental' realm, because they are knowable by someone in the first way; while material objects, animals, etc., belong to the 'physical' realm, because they are knowable in the second way.

The distinction between the mental and the physical realms is based, first, on the way in which the objects within these realms come to be known by us, viz., either inferentially or non-inferentially, or, by description or by acquaintance. Second, the distinction is based on the apparent necessity for objects of the mental realm to be dependent for their existence upon being objects of someone's awareness, and the lack of such necessity for objects of the physical realm. It is not suggested that these are the only differences.

Knowledge of objects in the mental realm is obtainable by introspection, self-scrutiny, by paying attention to what goes on in one's own mind. Knowledge of objects in the physical realm is obtainable, on an unsophisticated level, by sense perception, and, on a much more sophisticated level, by the use of scientific methods.

The use of scientific methods leads to the discovery of objects in the physical realm that are unobservable by the
unaided senses, e.g. such microscopically observable objects as viruses or bacteria, and such telescopically observable objects as quasars. Science postulates the existence of very minute and very distant objects that are observable only by complex instruments; but it also postulates the existence of objects that are not observable at all, e.g. theoretical entities like electrons and electric fields.

The question, in the light of which the identity theory can perhaps be best understood, is this: what is the relationship between the mental and the physical realms? I shall consider some of the different answers given to this question by way of the languages that are used to communicate about the two supposedly distinct realms.

The language used for communication about objects of the mental realm is a certain part of ordinary language. In particular, it is that part of ordinary language in terms of which psychological, mentalistic descriptions are given by a person of his own experiences, and perhaps also of the experiences of others. This 'mental language' is subjectivistic, in the sense that when a person uses it for describing his own experiences it is implied that his experiences are more or less private to him, that he is normally in the best possible position to judge the nature of his experiences, and consequently, that he is, as a rule, the final authority on the correctness or appropriateness of the description given in the mental language.
'Physical language,' on the other hand, is either a different part of ordinary language, or else it is a non-ordinary, technical language. The particular objects of the physical realm being described may be given a common-sensical description based normally on unaided sense experience, or given a scientific description based normally on instrumental observation. As opposed to mental language, physical language is used to describe not what is 'in the mind,' but rather what is 'in the world.' Physical language is objective, in the sense that at least in principle everyone is in an equally good position for determining the correctness or appropriateness of a description given in its terms. The speaker, in using physical language, even about himself, enjoys no authority comparable to that which a speaker using mental language has.

Thus the question asked above about the relation between the mental and physical realms is to be interpreted as asking: what is the relation between subjectivistic descriptions, i.e. those given in mental language, and objectivistic descriptions, i.e. those given in physical language?

The identity theory can be understood as an attempt to argue that it is not the case that the objects of the mental realm are distinct from the objects of the physical realm. Thoughts, feeling, sensation, images, and the like are not objects that exist outside of, or apart from the physical realm; they are in fact identical with certain objects of the physical realm. In particular, mental objects
are identical with certain types of neurological processes going on in the brain.

The reason why mental objects appear to be different from physical objects is that normally they come to be known differently. Physical objects, as a rule, come to be known through sense experience and instrumental observation, while mental objects, as a rule, come to be known through introspection. The apparent difference between mental and physical objects, according to the identity theory, is not due to any intrinsic difference between these objects, but rather to the way in which we normally acquire knowledge of them. The identity theory argues also that introspection itself is a physical activity, identical with certain activities of the brain.

The most fundamental question that the identity theory has to face is that of accounting for the presence of the subjectivistic element in mental descriptions. For if the theory is correct in maintaining that mental objects are really physical objects, then mental descriptions of mental objects should be purely objectivistic like physical descriptions of physical objects are. It is obvious, however, that this is not the case.

The identity theory meets this problem by stipulating that what is being argued for is not an identity of descriptions in the mental and the physical languages, but rather an identity of the referents of these descriptions. But this still leaves unexplained why is there a subjectivistic
element in the mental language. Different versions of the
identity theory will be distinguished according to the solu-
tions they offer to this question. There are, of course,
other differences apart from the one that I shall discuss,
between different versions of the identity theory.

(2) The idealistic version of the theory.

The main exponent of this version is H. Feigl.

Feigl's position, stated very briefly, is that while objects
of the mental realm are identical with objects of the physi-
cal realm, mental objects have an irreducibly subjective
aspect; Feigl called this aspect 'raw feels.' In virtue of
the identity, it is possible to explain and predict the
occurrence of thoughts, feelings, sensations, images, and the
like, because they are identical with the occurrence of cer-
tain types of neurological events. So, mental objects can
be accommodated in the physical realm, in the space-time-
causal network, or, as Feigl put it, in the 'nomological net.'
In other words, with the help of science the behaviour of
mental objects can be explained by subsuming it under

---

2H. Feigl, 'Physicalism, Unity of Science and the
Foundations of Psychology,' in The Philosophy of Rudolf
Carnap, ed. P. A. Schilpp, The Library of Living Philosophers,
(New York: Tudor, 1966). The article was written in 1954.
Also, 'The "Mental" and the "Physical,"
' in Minnesota Studies in the Philosophy of Science, Vol. 11, ed. H. Feigl, M. Scriven,
G. Maxwell. (Minneapolis: University of Minnesota Press,
1958). Hereafter abbreviated as 'Feigl.' All references will
be to this work. Also, 'Mind-Body, Not a Pseudoproblem'
scientific laws. Mental objects, according to Feigl, are not 'nomological danglers.'

However, according to Feigl's view, it does not follow from the fact that the behaviour of mental objects can be given a scientific explanation that the raw feel aspect of mental objects is nothing but physical. Just the contrary; Feigl stressing the role of raw feels writes:

If the synthetic (i.e. contingent) element in the relation [between mental and physical objects] that we have stressed throughout is admitted, then there is something which purely physical theory does not and cannot account for. Is there then a kind of 'brute fact' which our monistic theory has to accept but for which there is possibly no explanation, in the same sense as there can be (within naturalistic empiricism) no explanation for the fact that our world is what it is? . . . ?

When Feigl faces the question of just what this brute fact is, and what its supposed existence implies about the world, he adopts a position that is indistinguishable from some forms of idealism. Feigl writes:

Speaking "ontologically" for the moment, the identity theory regards sentience (qualities experienced, and in human beings knowable by acquaintance) and other qualities (unexperienced and knowable only by description) the basic reality. . . . It shares with certain forms of idealistic metaphysics, in a very limited and (I hope) purified way, a conception of reality.

Thus Feigl's answer to the question of why there is a subjectivistic element in the mental language is that it is there to describe the raw feel aspect of mental objects. This aspect of mental objects is the "basic reality." As

3Feigl, pp. 390-391.
4Feigl, p. 474.
Feigl puts it:

According to the identity thesis the directly experienced qualia and configurations are the realities-in-themselves that are denoted by the neurophysiological descriptions.

So, according to Feigl's view, it is not the case that all mental descriptions are really descriptions of objects in the physical realm. It is rather the case that some physical descriptions are really descriptions of objects in the mental realm.

It follows from Feigl's position that there actually do exist objects of the mental sort. Feigl insists on "the efficacy of mental states, events, and processes in the behaviour of human (and also sub-human) organisms." But if this is the case, then Feigl has abandoned the original purpose of the identity theory, viz., to argue that mental objects are in fact objects of the physical realm. It is not clear whether Feigl has opted for some form of idealism, or rather for some form of dualism. It is clear, however, that the position he has ended up with is not one asserting the identity of mental and physical objects.

There is, however, a general objection against all forms of the identity theory. The objection is that all versions of the identity theory imply idealism. Feigl, on this

---

5Feigl, p. 457.
6Feigl, p. 388.
7I shall offer a more detailed criticism of Feigl's position in Chapter Eight.
objection, merely made explicit what was already implicit in the theory. It might be argued that if we emphasize the identity holding in one direction in order to support materialism, then the fact that the identity must hold in the other direction as well, can be used to support something like idealism. Why, it might be asked, should the objectivistic properties of objects in the physical realm be ascribed to mental objects, rather than the subjectivistic properties of mental objects be ascribed to objects of the physical realm? Or, as Stevenson \(^8\) puts it: if it is true that if sensations are not identical with brain processes, then sensations are 'nomological danglers,' it follows, that if sensations are identical with brain processes, then brain processes too are 'nomological danglers.' Stevenson's assumption is that if sensations are identical with brain processes, then all the properties of sensations are also properties of brain processes. If sensations are 'nomological danglers' in virtue of some of their properties, then since brain processes have these properties, they must be 'nomological danglers' as well.

If it were the case that mental objects had properties which rendered them 'nomological danglers,' then of course Stevenson's criticism would be well-founded. One has to ask, however, what are the properties which, if possessed by

mental objects, would cause them to 'dangle?' I take it that the apparent privacy, the agent's apparently privileged access and final authority, the apparent lack of spatial localizability, are some of the properties which are supposed to make it impossible to find a place for mental objects in the space-time-causal network. Naturally, the identity theory has to show either that these properties are not really properties of mental objects, or that while the properties are properties of mental objects they are not such as to render them, and thereby the physical objects with which they are supposedly identical, 'nomological danglers.' Stevenson's criticism rests on the assumption that the identity theory cannot provide the required analysis of the properties in question. I shall attempt later on to do part of the analytical job that needs to be done by showing that the concept of identity used by the identity theory does not involve a symmetrical relation. Stevenson is correct, I think, in putting the onus of argument upon the identity theorist. He is wrong, however, to assume even without argument that the onus cannot be discharged.

Assuming for a moment, then, that the required analysis can be provided by the identity theory, the question still remains: Why does the theory emphasize that mental objects are really physical objects, and not that physical objects are really mental objects? The reason is that the identity theory proceeds on the assumption that the behaviour of mental objects can be explained in scientific terms. If
the subjectivistic elements in the mental realm were irreducible, then the possibility of a complete scientific explanation of the mental realm would be ruled out. The identity is stated in the way that it is, because the mental-to-physical direction allows for the possibility of scientific explanation, whereas the physical-to-mental direction does not.  

(3) The radical version of the theory.

Two contemporary exponents of this view are B. A. Farrell and P. K. Feyerabend. I shall discuss Feyerabend's position. He argues that any statement of mental-physical identity (H) is basically defective, since it necessarily involves some expressions from the mental language and therefore necessarily involves those subjectivistic elements which it wishes to eliminate. In considering a paradigmatic H, i.e. "X is a mental process of a kind A=X is a central process of kind a," Feyerabend writes:

But this hypothesis backfires. It not only implies, as it is intended to imply, that mental events have physical features; it also seems to imply (if read from the right to the left) that some physical events, viz., central processes, have nonphysical

9This will be discussed in detail in Chapter Seven.


11P. K. Feyerabend, 'Materialism and the Mind-Body Problem,' Rev. of Meta., 17 (1963), pp. 49-66. Also, 'Comment: Mental Events and the Brain,' J. of Phil., 60 (1963), pp. 295-296. References will be to the second article.

12Feyerabend, p. 295.

13Feyerabend, p. 295.
features. It thereby replaces a dualism of events by a dualism of features.

Feyerabend, in this passage, accepts the sort of criticism Stevenson has offered. His reply to it, unlike mine, is to change the identity theory in such a way as to escape what is taken to be the force of the criticism. In recommending a way out for the identity theorist, Feyerabend suggests\textsuperscript{14} that

The proper procedure for him to adopt is to develop his theory without any recourse to existent terminology. If he wants to use \textit{H} at all, he ought to use it for redefining 'mental process' (if he intends to perpetuate ancient terminology, that is).

Feyerabend's proposal for overcoming the difficulties of the identity theory takes the heroic course of abandoning mental language altogether. Farrell and Feyerabend alike regard adherence to a subjectivistic mental language as a form of superstition. Farrell claims that the subjectivistic element "can be shown to resemble an occult notion like 'witchcraft' in a primitive society that is in the process of being acculturated to the West."\textsuperscript{15} While Feyerabend argues\textsuperscript{16} that abandoning subjectivistic mental language for a physical language is like abandoning such notions as "possessed by the devil" for a physiological theory of epilepsy.

The obvious criticism of Feyerabend's position is that it is developed in order to solve a problem, viz., to

\begin{itemize}
\item \textsuperscript{14}Feyerabend, p. 296.
\item \textsuperscript{15}Farrell, p. 195.
\item \textsuperscript{16}Feyerabend, p. 296.
\end{itemize}
discover the relation between mental and physical objects. However, in the course of offering a solution, Feyerabend redefines the problem in such a way that the only possible solution is his. If all the indications that mental objects may have a subjective aspect are labelled 'superstitious,' then naturally a purely physical description of mental objects would suffice. So the radical version of the identity theory should be regarded either as tautologically true, or as question-begging.

Feyerabend, however, is aware of this objection, and attempts to meet it. He calls into question the view that the terms of a general point of view of a corresponding language can obtain meaning only by being related to the terms of some other point of view that is familiar and known by all.\(^{17}\)

In other words, the above objection against Feyerabend's position assumes that a neurophysiological description of mental objects would be meaningful only if that description could be related to mental descriptions of mental objects. Feyerabend points out, however, that if the fulfillment of this requirement is demanded of neurophysiological language, then it ought also to be demanded of mental language, as well as of all other languages. But given the justifiability of this demand, it would follow that there could not have been a first language, since descriptions in that language could \textit{ex hypothesi} be related to no other descriptions. And so, Feyerabend concludes, if it is necessary, as it is, to

\(^{17}\)Feyerabend, p. 296.
have a language in terms of which descriptions are given that are unrelated to descriptions given in terms of other languages, then this language might as well be a sophisticated, quantifiable technical language, rather than an unsophisticated, superstition-ridden mental language.

One of the many difficulties in Feyerabend's position is that in the last analysis it failed to provide a method of deciding between alternative theories stated in terms of different languages. Feyerabend might reply that the method is to consider the number of successful predictions that could be made on the basis of the rival theories. But this answer merely postpones the criticism that because the success of a prediction, according to Feyerabend, must be judged in terms of the theory in which it was made, each theory will be judged successful in its own terms. The only case in which this does not seem to hold is that in which one of the two rival theories could be made a sub-theory of the other. The alternative is to postulate some sort of extra-theoretical standard by which theories could be evaluated. And this standard we actually have in the form of ordinary language.

Feyerabend proceeds by accepting Stevenson's criticism. To meet this criticism Feyerabend attempted to redefine the terms in mental description, or even to abandon mental descriptions altogether. This led to the problems discussed above. But the identity theory need not accept Stevenson's criticism, as I have argued, nor is it necessary, if one does accept Stevenson's criticisms, to redefine mental descriptions.
It may be possible to find some way of translating mental descriptions into physical descriptions. Smart's version of the identity theory attempts to do just this.

(4) The translation version of the theory.

The main exponents of this version are Place and Smart. I shall discuss Smart's formulation of the theory.

Smart's position may be interpreted as an attempt to account for the subjectivistic elements in mental language by showing that they are elements of the language used to describe mental objects, and not elements of these objects themselves.

---


Consequently, mental objects may be identified with physical objects. Smart's way of doing this is to argue that mental descriptions can be translated into "topic-neutral" descriptions, i.e. descriptions that are neutral between subjectivistic and objectivistic analysis. Once this is done, the way is open for an investigation into the referents of topic-neutral descriptions. Smart argues that this investigation would reveal that the referents of topic-neutral descriptions are in fact brain processes.

Smart argues in the following way: suppose it is objected that sensations and brain processes cannot be identical because in order to assert their identity we have first to identify them independently of each other. Yet sensations are identified through those subjective elements whose presence distinguishes them from physical processes. So either sensations and brain processes cannot be identical, because they cannot be identified independently, or they cannot be identical, because the way in which they are identified independently reveals that each has properties the other does not have. 21

Smart replies by arguing that this objection establishes only that there is one process having both subjectivistic and objectivistic properties; so a dualism of properties still needs to be refuted by Smart. Let us suppose that

20 Smart, SBP, pp. 147-148.

21 I shall discuss this objection at some length in Chapter Seven.
someone gives a mental description, e.g. 'I see a yellowish-orange after-image.' The question for Smart is how to interpret this description in such a way as to exclude the subjective element, to remove, as it were, the suspicion that there is something irreducibly subjective about the experience of having an after-image.

Smart writes: 22

My suggestion is as follows. When a person says, "I see a yellowish-orange after-image," he is saying something like this: "There is something going on which is like what is going on when I have my eyes open, am awake, and there is an orange illuminated in good light in front of me, that is, when I really see an orange."

Smart then goes on to argue: 23

Notice that the italicized words, namely "there is something going on which is like what is going on when," are all quasi-logical or topic-neutral words. This explains why the ancient Greek peasant's reports about his sensations can be neutral between dualistic metaphysics and my materialistic metaphysics. It explains how sensations can be brain-processes and yet those who report them need know nothing about brain-processes. For he reports them only very abstractly as "something going on which is like what is going on when . . ."

It also explains, if Smart is right, how it is possible to get rid of the subjective element in mental descriptions, viz., by translating them, with the help of the above formula, into topic-neutral descriptions.

The logical status of the topic-neutral formula is

22 Smart, SBP, p. 149.

23 Smart, SBP, p. 150.
much like that of "someone." Smart argues:

A person may say "someone is in the room," thus reporting truly that the doctor is in the room, even though he has never heard of doctors. (There are not two people in the room "someone" and the doctor).

Similarly, when it is reported that "something is going on which is like . . .," and it is after-imaging that is going on, it is not the case that two things are going on, after-imaging and the "something" that may be unpacked as a brain-process; there is only one thing going on, viz., the brain process. This one thing can be given an indefinite description, and thereby eliminating the subjective element, or it can be given a definite description in terms of physical processes.

It seems to me, however, that this procedure involves Smart either in self-refutation or circularity. Suppose that I am having an after-image and I wish to state this with the help of the topic-neutral formula. So I shall say, 'there is something going on which is like what is going on when . . . ' and attempt to complete the expression by filling in the details required by Smart. But how can I do this? If I give anything less than a complete account of

\[\text{24}^{\text{Smart, SBF, p. 150.}}\]

what whatever is going on is like, then I am always open to
the charge that my topic-neutral description does not
uniquely describe having an after-image. For several experi-
ences may fit Smart's topic-neutral description that "There
is something going on which is like what is going on when I
have my eyes open, am awake, and there is an orange illumi-
nated in good light in front of me, when I really see an
orange." For example, I may really be seeing an orange,
rather than having an after-image of one, or I may be having
a daydream, or hallucination, or suffer from an illusion.
In order to achieve unique description I should have to
express, in terms of the topic-neutral formula, my having an
after-image in such a way that that experience would be dis-
tinguished from other similar experiences. In other words,
I should have to mention properties that my present experi-
ence has, but other similar experiences do not have. But
what sort of property could I mention?

It can not be that all the properties I mention are
neutral between being physical and mental. For there will
always be at least one experience from which the present
experience would not be distinguished, viz., the experience
to which I compare my present experience. If what is going
on now is like what is going on when I really see an orange,
then perhaps I really am seeing an orange, and am not having
an after-image of one. Thus, either additional, non-neutral
properties must be mentioned to distinguish the two experi-
ences, or one could never express, if one relied on the
topic-neutral formula, what experience he is having.

The additional properties required must be either physical, e.g. there is involved in the experience I am now having a certain sort of neurophysiological process, or the property must be mental, e.g. I know through introspection that it is an after-image and not a visual sense datum that I am having. If the property is said to be mental, Smart’s case is lost, because non-physical properties have been allowed into his physicalistic universe. If the property is said to be physical, then Smart is involved in circularity. For the topic-neutral formula was introduced to provide a non-mental and non-physical description of an experience. But if description of a physical property is required for the description of the experience, then the topic-neutral formula has not accomplished its purpose. Consequently Smart’s way of excluding the subjectivistic elements fails.

(5) The reductive version of the theory.

This is the version of the theory that I propose to expound, defend, and justify. In agreement with Smart and Feyerabend I hold that the subjectivistic elements in descriptions of mental objects are elements of the language used to describe mental objects, and not elements of the objects themselves. But I differ from Smart and Feyerabend in that I deny both that descriptions given in a mental language can be translated into physical language, and that the mental language should be abandoned or redefined. My
position in respect to the subjective elements lies between Smart's and Feyerabend's.

It should be recognized, I think, that descriptions given in mental language cannot be translated, without change of meaning, into descriptions given in the physical language. The reason for this is, according to my view, that mental descriptions are surrounded by intensional practices that physical descriptions lack, e.g. privacy, intentionality, first-person authority, modal predicates appropriate to introspection, etc. Or, to put it differently, mental and physical descriptions have different sense, as opposed to having different reference. However, it would be a mistake, I think, to conclude from this that if an identity theory were to hold, then the mental language would have to be abandoned or redefined. It is sufficient, according to my view, to recognize that intensionality is not a property of the objects described in the mental language, but rather it is a property of these descriptions themselves. Thus there is nothing to prevent one from arguing that the identity of the referents of mental and physical descriptions is a kind of extensional identity. It follows that, once this is recognized, there is no need for attempting the futile task of translation, nor is there need for causing a linguistic revolution by changing the mental language.

The identity between the referents of mental descriptions and physical descriptions thus holds in virtue of extensional properties, e.g. having the same causes and
effects, having the same spatio-temporal location, etc. It is not the case, however, that all extensional properties that mental objects have are also properties that physical objects have. The identity, I shall argue, is not a symmetrical one. It is rather an asymmetrical, reductive identity. Thus the properties of mental objects are either identical with or are reducible to the properties of physical objects.
(1) The purpose of this chapter is to state, in very general terms, one version of the identity theory.¹ The chapter will be devoted completely to exposition, although, of course, not everything that needs to be said in an expository way will be said here. Subsequent chapters will deal with many of the problems that arise in connection with the identity theory. Further exposition and clarification will be offered in reply to criticisms considered at later stages.

The statement I wish to clarify, justify, and defend against objections is the following:

T: Each occurrence of the process of having a particular type of sensation is identical with the occurrence of a particular specifiable type of brain processes (BP).²

That is, for each and every sensation there is a BP such that the occurrence of the sensation is identical with the occurrence of the BP.

¹Frequent reference will be made to the 'identity theory'; it should be interpreted as 'the version of the identity theory I hold.'

²The phrase 'particular specifiable type of brain processes' will be abbreviated throughout as 'BP.'
In this chapter I shall discuss, sometimes only in an introductory way, the following points: 'Why is the identity restricted to sensations?' 'Are sensations entities or processes?' 'What is included in the class of BP?' 'What is the nature of the alleged identity?'

(2) The identity theory will be applied only to the case of sensations. This self-imposed limitation has certain implications. First, if the identity theory turns out to be correct, i.e. if T is true, then there is still a long way to go before an identity theory of mind can be established. The relation between the identity theory of sensations and an identity theory of mind is that if the former is correct, then the latter may or may not be correct. While if the former is incorrect, then the latter is necessarily incorrect.

Secondly, there seems to be an intimate connection between sensations and bodily processes, e.g. it is my body that hurts, or itches, or is tired. (I shall say more about this shortly.) The apparently intimate connection between sensations and bodily processes makes sensations perhaps the easiest type of experience about which an identity theory can be stated. If the theory fails in the case of sensations, it is highly probable that it will fail also in the case of experiences whose connection with bodily processes appears to be much more tenuous.

Thirdly, since sensations are not typical of all other mental states, events, or processes, the correctness of the
identity theory of sensations will not necessarily provide a model for the explanation of other mental occurrences. That is, from the alleged fact that sensations are identical with BP, it does not follow *mutatis mutandis* that the same logical type of explanation will hold, for instance, of thoughts as well.

Fourthly, the following considerations lead me to say that sensations are not typical of all mental occurrences. The question of whether animals can be said to have a mind does not allow of a straightforward answer. What is usually said is that it depends on what is meant by 'mind.' On the other hand, whether some animals have sensations is not in the least puzzling. Of course some animals have pains and itches, feel fatigue, and the like. We are willing to ascribe the having of sensations to animals, but we are hesitant to do the same of minds. Also, if such typical mental states or events as thinking, feeling, and willing are impaired, it is natural to say that the mind is impaired. Different types of mental illness are sometimes described as 'thought disorder' or 'emotional disorder,' while many neuroses are characterized by the impairment of volition, e.g. obsessive-compulsive neurosis. The impairment of a person's capacity to have a particular sensation does not warrant the conclusion that his mind is impaired. A person who cannot feel pain is said to suffer from 'brain damage,' not from the abnormality of his mind.

Lastly, in spite of these limitations of the identity
theory, its success would drive a not-so-thin wedge into the position of those who deny the possibility of an identity theory of mind. For if this theory were to succeed in identifying one type of mental occurrence with some sort of bodily occurrence, then at least the possibility of doing the same in the case of other mental occurrences would thereby be implied.

(3) In this section I shall offer some general features of sensations, i.e. of pains, aches, itches, tingles, dizziness, nausea, fatigue, and the like. Not all the sensations have all the features I am going to discuss. However, for an experience to be a sensation it has to be describable in terms of, at least, many of these features.

Following Armstrong I shall distinguish transitive and intransitive sensations. Sensations of warmth, pressure, motion, and the like, are transitive, while pains, aches, itches, tingles, dizziness, nausea, fatigue, and the like, are intransitive.

A difference between the two kinds of sensations is that in the case of transitive sensations we can distinguish the sensation of warmth and warmth, the sensation of pressure and pressure, and so on. Warmth, pressure, and motion can and do exist without a person sensing them. The distinction, however, cannot be made in the case of intransitive sensations.

Pains, aches, itches, tingles, and the like are sensations; they could not exist without a person sensing them. I shall be concerned almost exclusively with intransitive sensations.

Sensations, then, seem to have all, or most, of the following characteristics. First, they must be individuated with reference to a particular agent. It makes no sense to speak of the occurrence of a sensation independently of an experiencing agent. Although it seems to be a universally applicable rule that sensations have to be individuated with reference to agents, it is not nearly as clear that a particular sensation has to be individuated with reference to one and only one agent. Cases like pain at the point where Siamese twins are conjoined, or the logical possibility of connecting two nervous systems are sufficiently puzzling not to allow for an unqualified statement of the 'one agent, one sensation' principle. Such cases, however, are deviant; normally the principle holds without need for qualification.

Second, sensations are temporally determinable. It is usually proper to ask 'How long did it last?' 'When did it start?' 'Does it occur frequently?' and so on. In other words, sensations have duration. This, however, does not mean that whenever a person has a particular sensation he can make true, or even any, statements about its duration. It

4 Instead of writing 'intransitive sensation' I shall write 'sensation' from now on. Qualifications, if required, will be noted.
merely means that it is proper to ask temporal questions about sensations.

Third, sensations, unlike other mental states, e.g. anger, belief, thought, etc., either appear to have bodily location, or appear to be intimately connected with bodily conditions.\(^5\) Pains, aches, itches, tinges, and the like can always be located with greater or lesser accuracy at some part of the body. It is possible to 'ache all over,' but this indicates only that many parts of the body can ache simultaneously. A possible test of whether or not a sensation has been correctly located is the application of local anesthesia. If the sensation disappears, it was correctly located; if it persists it was incorrectly located.

On the other hand, nausea and dizziness seem to have no bodily location; only total anesthesia will put an end to them. Both of these, however, appear to be intimately connected with conditions in certain parts of the body: nausea with the digestive tract, dizziness with the inner ear. The test of the intimate connection holding between these sensations and parts of the body is that the sensation can be produced at will by subjecting the required part of the body to appropriate stimulation.

It is quite unimportant for my present purposes that

\(^5\)The problem of the spatial localizability of sensations will be the subject-matter of a subsequent chapter. Much more will be said there about this characteristic of sensations.
involuntary holds with the proviso that they are involuntary once the physical conditions giving rise to them are present.

Fifth, reports of sensations, unlike reports of perceptual observations, memory-claims, statements about one's motives, inclinations, personality traits, are, at least relatively, incorrigible.6 The owner of a sensation seems to have an advantage over other people in knowing about his own sensations. Whether the report of the owner has a necessary or only a contingent authority is a highly controversial question. Regardless of the outcome of the controversy, it seems clear that the person who has the sensation is normally in a far better position to make true statements about it than other people are.

Sixth, in the description of a sensation, e.g. pain, the following considerations may be relevant: location, duration, source or cause, intensity, quality, (e.g. sharp, throbbing, stabbing, etc.) comparison (i.e. how the present sensation is like or unlike other past sensations).

Seventh, in all but the most peculiar cases, a physical-cum-physiological explanation of the occurrence of the sensation is proper. Several different types of scientific explanation may be given in reply to the question: 'Why did the sensation occur?' An explanation may be offered in terms of

6 The problem of the incorrigibility of avowals will be discussed in two subsequent chapters. The claim made here will be greatly amplified later.
external stimulation, e.g. you have a pain because you were hit by a falling rock. This sort of answer explains the onset, but not the continuation of the sensation. The continuation of the sensation, i.e. the transmitting of the stimuli by the pain receptors to the brain, is explained by describing how the human body functions. This explanation includes reference both to the peripheral and the central nervous system. However, it seems possible to produce sensations in a person in the absence of the transmitting activity of the peripheral nervous system. In such cases the explanation is in terms of the structure and function of the central nervous system alone.

(4) The expression 'sensation' carries with it the process-product ambiguity. 'Sensation' may be taken to refer to the process of sensing, or it may be interpreted as referring to the product of the process of sensing. In the first sense of 'sensation,' pains, aches, tickles, and the like should be regarded as processes going on in the agent. In the second sense of 'sensation,' particular sensations are supposed to be objects or entities.

Following Smart, I shall argue for the identity of BP and of sensations as processes, not of sensations as objects. Sensations will be regarded as processes going on

---

7 I shall discuss this possibility, in greater detail, later.

8 Smart, PSR, pp. 37-38.
in an agent. The phrase appearing in T: 'the process of having a particular sensation' serves to emphasize this point.

The process of having any particular sensation, according to the identity theory, has two main components. The first component is a stimulus, the second component is awareness\(^9\) of the stimulus. The presence of these two processes, viz., the stimulus and awareness of it, constitute the necessary and sufficient conditions for the ascription of a sensation to a person. Without some sort of stimulus there would be nothing of which the person could be aware, while without awareness of the stimulus the person would not sense anything.

Different kinds of sensations are distinguished on the basis of differences among stimuli, e.g. whether the external stimulus is transmitted by pain receptors or by heat receptors, and also on the basis of such cognitive attitudes as recognition, comparison, evaluation of the present sensation in relation to other sensations, past and present.

The contention of the identity theory is that the process of having a sensation is actually identical with BP.

(5) The BP with which any particular process of having a sensation is identical is: a stimulus being scanned by the brain. The cash value of T is that both components of sensations, viz., the stimulus and awareness of it, are identical

\(^9\)The concept of awareness will be discussed later at length.
with physiological processes. The stimulus involved in the
having of a sensation is identical either with physiological
processes going on in the peripheral nervous system, or with
direct stimulation of the brain. The awareness of the stimu-
lus, on the other hand, is identical with the scanning activ-
ity of the brain.

By 'stimulus' is not meant such environmental influ-
ences as may impinge on the body; what is meant is rather
environmental influences that impinge on the body and are
transmitted by the peripheral nervous system to the central
nervous system. A person, however, may or may not become
aware of a stimulus. A stimulus having reached the brain is
a necessary condition of awareness of that stimulus. But a
stimulus may have reached the brain without the person becom-
ing aware of it. The person becomes aware of a stimulus only
if it is scanned by the brain.

Thus there are three stages to be distinguished:
(i) The presence of an environmental influence that may or
may not become a stimulus, e.g. a pin-prick is an environmen-
tal influence that normally becomes a stimulus when it acti-
vates pain-receptors, but if the person is locally anesthe-
tized and is prevented from otherwise observing whether or
not he is receiving a pin-prick, then the environmental influ-
ence does not become a stimulus.
(ii) The presence of a stimulus which may or may not be
scanned by the brain, e.g. a tactile stimulus caused by a fly
walking on a sleeping person's hand, which may be transmitted
to the brain where it is inhibited rather than scanned.

(iii) The presence of a scanned stimulus which constitutes the necessary and sufficient conditions of a person having a sensation, e.g. a pin-prick becoming a stimulus if scanned by the brain.

It is only when the third stage is reached that a person can be correctly described as having a sensation. Thus the processes identified with the having of a sensation are not physiological processes occurring partly in the peripheral and partly in the central nervous system. Rather, the processes identified with the having of a sensation are processes going on in the central nervous system after the stimulus has reached it. In other words, the processes so identified are BP.

A consideration that further reinforces the view that identity holds between sensations and BP, and not between sensations, on the one hand, and central or peripheral neurological processes, on the other hand, is this: the stimulus which is scanned need not reach the brain through the peripheral nervous system. It has been established experimentally that direct stimulation of certain parts of the brain, in the absence of other stimuli, gives rise to the subject's reporting that he is having certain kinds of sensation. If these parts of the brain are damaged or destroyed, then the person's capacity for having a particular kind of
sensation is impaired. The 'scanning activity' refers to that function of the brain whereby it discriminates between those stimuli that require some sort of response and those that do not. The discrimination is based on such factors as past experience, the intensity of the stimulus and the presence of other stimuli.

It is unnecessary and presumptuous for a philosophical supporter of the identity theory to speculate about the nature of the scanning activity of the brain. The justification for so doing is to give some empirical content to the abstract treatment of the problem. It is very tempting to say that it is the reticular activating system or the reticular formation (RAS) that performs the scanning activity.

Two extremely interesting functions of the RAS are the determination of whether or not a person is conscious, at least in one sense of 'conscious,' and the selection of stimuli for the purpose of responding to them. The RAS, being the control centre of wakefulness determines whether a person is conscious. If the RAS is damaged in certain ways, the person falls into a coma from which he never recovers.


If the RAS of sleeping animals is stimulated, they immediately wake up. The correlation between awakening and direct stimulation of the RAS, in the absence of other stimuli, is, according to Wooldridge, fairly well established. The other interesting function of the RAS is the selection of 'important' stimuli. A stimulus selected by the RAS is one of which a person is aware, while a stimulus not selected by the RAS is never a conscious one. The 'selection' of a stimulus by the RAS amounts to the transmission of the stimulus to certain loci in the cerebral cortex, while the failure to 'select' a stimulus is the inhibition of, i.e. failure to transmit, a stimulus.

Thus the contention of the identity theory, or, in other words, the cash value of T, is that the process of having a sensation is identical with BP. The occurrence of a particular sensation involves awareness of a stimulus of a certain sort, while the occurrence of BP involves a scanned stimulus. The contention is that these occurrences are one and the same.

(6) The particular kind of identity that is alleged to hold between sensations and BP is theoretical identity. Theoretical identity is a kind of property-identity that is empirical and not logical, extensional and not intensional, heterogeneous and not homogeneous.

---

12Chapter Seven is devoted to a discussion of theoretical identity. Here I merely give a sketch.
That theoretical identity is empirical means that the denial of the identity, if the identity holds, does not lead to self-contradiction. The identity holds not between expressions, but between denotations of expressions. The contention is not that the expressions 'sensation' and 'BP' are always used, or should be used, in the same way, but rather that the denotations of these expressions are identical.

That theoretical identity is extensional means that the identity holds only in respect of extensional properties of the denotations of 'sensation' and 'BP.' That is, the identity is not meant to include intensional and modal predicates. The contention is that the denotation of 'sensation' and the denotation of 'BP' are so related that all extensional properties of sensations are reducible to some properties of BP.

That theoretical identity is heterogeneous means three things. First, it means that the expressions 'sensation' and 'BP' occur in the context of different languages. That is, 'sensation' occurs in ordinary language, while 'BP,' appropriately qualified, occurs in the language of physiological theory. Second, it means that the reduction is accomplished if (i) ordinary language contains terms that neither occur in the language of the physiological theory, nor are definable in terms of that language; (ii) any observational data explainable by the common sense view of sensations are explainable by the physiological theory; and (iii) the balance between the strength and simplicity of the two favours
the physiological theory. Third, it means that the reductive relation between the common sense view of sensations and the physiological theory is transitive, irreflexive, and asymmetric.
(1) Let us again consider T:

\[ T: \text{Each occurrence of the process of having a particular type of sensation is identical with the occurrence of BP.} \]

The correctness of the identity theory depends on the truth of T. It is thus of fundamental importance to clarify what sort of considerations are relevant to the truth or falsity of T. There are, I shall argue, both empirical and philosophical considerations involved in determining the truth or falsity of T.

If the occurrence of a particular sensation is identical with the occurrence of BP, then it must be discoverable that there is a perfect correlation between these apparently distinct occurrences. Thus to any particular statement of identity, e.g. 'The headache I had at \( t_1 \) is identical with the occurrence of a BP at \( t_1 \),' there corresponds a statement asserting a perfect correlation between the occurrence of the particular sensation and the occurrence of the BP. Adopting Kim's\(^1\) usage, I shall refer to these as 'psycho-physical

\(^1\)J. Kim, 'On the Psycho-Physical Identity Theory,' \textit{Am. Phil. Quart.}, 3 (July 1966), pp. 227-235. Hereafter abbreviated as 'Kim.'
correlation statements,' or more briefly as 'correlation statements.'

Each identity statement entails a corresponding correlation statement, but no correlation statement entails an identity statement. In other words, the truth of the correlation statement is the necessary, but not the sufficient, condition of the truth of the corresponding identity statement. The existence of a perfect correlation between the two processes is compatible with the distinctness of perfectly correlated processes. For example, the process of the melting of a substance and the process of the temperature of that substance increasing past a certain point are perfectly correlatable, yet distinct processes. Thus the assertion of identity requires justification beyond the truth of the corresponding correlation statement.

The question of whether or not there is a perfect correlation between the occurrence of particular sensations and the occurrence of BP is an empirical or factual one. It is, however, not a purely empirical or factual question.

The empirical or factual claim of the correlation statement will be confirmed if the correlatable processes fulfill the following requirements:

(i) The two processes are spatio-temporally coincidental, i.e. they always occur at the same time and at the same place,\(^2\) if

\(^2\)The problem of the sense in which processes can be said to occur at a place will be discussed later.
they both occur.

(ii) The two processes are non-separable, i.e. when one occurs the other always occurs.

(iii) The two processes cannot be manipulated independently, i.e. changes effected in one process are always accompanied by corresponding changes in the other process.

The correlation statement, however, is not purely empirical. The truth of the empirical or factual claim is a necessary, but not sufficient, condition of the truth of the correlation statement. One reason for this is that it is far from obvious that the having of a sensation occurs at a place. A toothache may be located in the tooth, but where is the having of a toothache located? Whether or not the having of a sensation is spatially localizable is mainly a philosophical or logical problem. What needs to be determined is whether it makes sense to say that the having of sensations is spatially localizable. Empirical testing of the actual location can begin only after it has been settled that the having of sensations could have spatial location.

The conclusion seems to be that while both factual and philosophical considerations are relevant for the truth of T, it is not easy to see how these could be discussed without reference to each other. The factual component of the identity theory, viz., the correlation statement, seems to involve philosophical assumptions, while the philosophical component, viz., the identity statement, involves factual assumptions, i.e. that the correlation statement is true.
There are, in my view, three considerations, requiring three different kinds of argument, involved in trying to prove the truth of T. First, there is the truth of the correlation statement. Second, there is the need for justifying the step from the assumption that the correlation statement is true to the assumption that the corresponding identity statement is true. Third, there is the need to show why the identity theory should be accepted rather than rival philosophical theories that also accept the truth of the correlation statement, e.g. parallelism, epiphenomenalism, the double aspect theory, and some types of interactionism.

Insofar as the first consideration, viz., the truth of the correlation statement, is an empirical or factual problem it should be left for the scientists to consider. Insofar as it is a philosophical problem it is my immediate concern here. In this chapter, and also in later chapters, I shall argue that there are no justifiable a priori objections against the attempt to establish the empirical or factual part of the correlation.

The second consideration, the justification for concluding that the having of sensations is identical with the BP, is the main concern of the present work.

The third consideration, a discussion of other philosophical theories that accept the truth of the correlation statement but deny the truth of the identity statement, will be marginal to my main purpose. I shall indicate in the last chapter how competing theories would be argued against,
but I shall argue against them only insofar as it is a necessary part of my constructive purpose.

It seems to me that the correctness of the identity theory vis-a-vis its philosophical rivals cannot be settled with finality. The reason for this is that the disagreement between them, as I shall try to show, is based, in the last analysis, on different conceptions of what philosophy is and what philosophers should do. The root of the disagreement is that the identity theory attributes far greater importance to science than does any of its competitors. I do not myself know how this could be justified on philosophical or logical grounds, nor, on the other hand, do I think that the refusal to give science pride of place has philosophical or logical justification. The justification, if there is one, lies partly in the remarkable success of science in helping us to understand reality, and partly in predicting that if we continue to act on the assumption that reality is as science depicts it, we shall be successful in coping with it.

I

(2) The truth of the correlation statement is a necessary condition of the truth of the corresponding identity statement. It is, therefore, a matter of crucial importance for the identity theory to show how the correlation between the having of sensations and BP could be established.

A basic requirement for the truth of the correlation statement is that the correlated items be identifiable
independently of each other. If the correlated items are not independently identifiable, then the correlation statement can not even be formulated, since it will not be known what are the items whose perfect correlation is asserted. While if the correlated items were not identifiable independently of each other, the identity statement corresponding to the correlation statement would beg the question.

The requirement, therefore, is to find such different indicators that can be used for the identification of the having of sensations, on the one hand, and for the occurrence of BP, on the other hand. These indicators must be such that they could be used for the identification of the having of sensations without involving any reference to the occurrence of BP; and also such that they could be used for the identification of the occurrence of BP without involving any reference to the having of sensations.

In my view, the required indicators for the occurrence of BP are 'brain-readings,' while the indicators for the having of sensations are instances of verbal and non-verbal behaviour. The way in which 'brain-readings' indicate the occurrence of BP will be discussed below in (3), while the way in which behaviour indicates the having of sensations will be discussed partly in (4)-(6), and partly in subsequent chapters.

(3) How do 'brain-readings' indicate the occurrence of BP?
Let us assume that there exists the framework of a
theory whose field of investigation includes the functioning of the brain. Since BP are 'micro-processes,' i.e. unobservable by the unaided senses, it is necessary to construct instruments for their observation. These instruments will be so constructed that they give a certain signal if and only if an instance of the phenomenon is present for the observation of which the instrument has been constructed. If a given instrument actually makes the signal, then the interpreted signal, i.e. what I have called the 'brain-reading,' can be taken as indicating the occurrence of the BP. The question is: do brain-readings indicate the occurrence of BP without involving reference to the having of sensations? The answer, I think, is in the affirmative.

The argument for brain-readings not doing this may be stated in the following way. In order to design an instrument through which BP can be observed it is necessary to know what BP is going to be observed. In the last analysis, BP are either identified, at the pre-scientific stage, through unaided sense perception, or BP are not identified at all. If they are identified through the unaided senses, then sensation, in some form, is a necessary condition of the construction of the instrument. And if this is the case, then instrumental observation--brain-reading--cannot be taken to indicate the occurrence of BP independently of the having of sensations.

The objection makes the perfectly valid point that perhaps in all cases of instrumental observation there is
involved sense experience. Unaided sense experience may be the necessary condition of instrumental observation. For the construction of instruments, and the interpretation of instrumental observation, may all involve the simple use of the unaided senses. However, from the facts that the occurrence of BP is indicated through instruments, and that instrumental observation involves, perhaps necessarily, unaided sense experience, it does not follow that brain-reading cannot be used to indicate the occurrence of BP without involving reference to the having of sensations. It may be that the occurrence of BP cannot be indicated independently of some sensation or other, but my contention is that the sensation involved in the independent identification of BP need not be the kind of sensation I am discussing.

The sensations whose identity with BP I am trying to establish is one class of sensations; there are other classes. The objection against brain-readings as indicators of BP overlooks the point that the sensation involved in the instrumental observation of BP may not be a member of the class of sensations with which I am concerned. The kind of sensation involved in the instrumental observation of BP may be, and usually is, visual sensation. I am dealing with bodily sensations.

(4) How does behaviour indicate the having of sensations?

This question will be discussed in two main steps.
The first step is to clarify what does and what does not
count as behaviour. I shall do this now largely by comparing and distinguishing behaviour and physiological processes.

The second step, a discussion of how behaviour indicates the having of sensations, will be the subject-matter of the next two sections.

It will be argued that while there is no sharp distinction between physiological processes in general and behaviour in general, there is a fairly clear distinction between certain types of physiological processes and certain types of behaviour. In discussing this question Hebb\(^3\) writes:

> Psychologists and neuro-physiologists thus chart the same bay—working perhaps from opposite shores, sometimes overlapping and duplicating one another, but using some of the same fixed points and continually with the opportunity of contributing to each other's results. The problem of understanding behavior is the problem of understanding the total action of the nervous system, and vice versa.

In a later book Hebb\(^4\) argues again that

> Behavior is the publicly observable activity of muscles and glands of external secretion as manifested in movements of parts of the body or in the appearance of tears, sweat, saliva, and so forth. Talking is behavior; so is a smile, a grimace, trembling, blushing . . . changing one's posture, or following the words of a printed page with the eyes.

Not all psychologists agree with Hebb that the explanation of physiological processes and behaviour is, in the last

---


analysis, one and the same problem. Those who differ, e.g. Skinner, Koffka, Tolman, etc., usually put their objections in terms of the molar-molecular distinction introduced by Tolman. Littman and Rosen suggest that Tolman’s distinction was put to the following use:

Tolman’s approach, then, was to find out what questions could be made systematically meaningful within a purely psychological framework. He argued that a scientific theory has meaning in so far as it is based upon a set of observable events, and that such meanings are a function of the way these events are described. One way of looking at events is in terms of whether they can be described as part or total functions. These functions were termed molecular and molar, respectively. The distinction provided a rationale for separating psychology from physiology by labelling physiology a part process.

According to this distinction, then, physiological explanation is about some part of the organism, while psychological explanation embraces the organism as a whole. Consequently, the distinction between physiological processes and behaviour is analogous to the distinction between the parts constituting the whole and the whole itself. The distinction has the further advantage of allowing for the whole, i.e. behaviour, to have characteristics which the parts, i.e. physiological processes, do not necessarily share.

There is, however, no reason why physiologists should refrain from giving 'molar' explanations. There is no

---

built-in limitation in their subject-matter or method which forces them to deal only with minutiae. The molar-molecular distinction is a conceptual one. Its acceptance or rejection is a methodological question; it says nothing about the phenomena which are to be explained. As Littman and Rosen 7 point out:

The criteria of molarity seem to imply that the subject-matter of psychology is of a special kind, and therefore requires a special manner of description. Behavioral events are molar, so the description too must be molar... [But this is not the case]... It is molar to describe phenomena in a certain way; it is not the content described which is molar or molecular.

There is, of course, a need to draw a distinction between psychological and physiological descriptions. It does not follow, however, that the phenomena described must also be so distinguished. In fact, as Hebb, Littman and Rosen, Pratt, 8 and Spence 9 argue, physiological processes and behaviour form one uninterrupted continuum, where the distinctions made are methodological-conceptual devices required by the logic of the enquiry, and not by the nature of the studied phenomena.

Let us look then at some of the conceptual distinctions required by this enquiry. 'Bodily activity' will be

7Littman and Rosen, pp. 151-152.


used as a neutral term to refer to the three classes of predicates. The first distinction is between two types of predicates that ascribe bodily activities to a person. This distinction is based on the difference between the sorts of explanations which are appropriate to the question 'Why did the bodily activity occur?'

Consider a person having a muscular spasm. The sort of explanation we commonly expect when we ask 'Why does he have the spasm?' is physico-chemical or physiological. 'Physiological processes' will be used to refer to those bodily activities whose occurrence is appropriately explicable in terms of physico-chemical or physiological regularities.

There is another type of bodily activity as well. If a person says: 'I ran a race,' and we ask 'Why did he run the race?', then the explanation we normally find satisfactory is not physico-chemical or physiological, but an explanation based on the motives, intentions, habits, etc., i.e. what I am going to call the 'psychological antecedents' of the person. 'Behavioural act' will be used to refer to those bodily activities whose occurrence is appropriately explicable in terms of the psychological antecedents of the person of whom the behavioural act is predicated.

The problem is: how is it determined what constitutes an appropriate explanation? I shall shirk the problem of what constitutes an explanation in general, and concentrate on how we determine whether or not an explanation, once given, is appropriate, while taking for granted that we
know what constitutes an explanation.

If the person is asked: 'Why did you run the race?', and he replies: 'Because I wanted to win,' or 'Because I always enter the running races of my club,' or 'Because I wanted to test my fitness,' etc., then we are told the sort of thing we expected to hear. Of course, even if we receive the expected type of reply we may be surprised. If, for instance, the person says that he ran the race because he wanted to compensate for his strict toilet training, we might want to have more information before we make up our minds. But, and this is the point, we do not think that his explanation is inappropriate. We might think that it is mistaken, but we are satisfied that he understood the original question.

The appropriateness of the explanation depends on the context in which the request for the explanation is made. Requests for explanation of behavioural acts are usually made in contexts which call for explanations based on psychological antecedents. This, of course, is not necessarily the case. 'Why did he run? (when ordinarily he would have walked)' may be the question one doctor asks another after the would-be athlete has collapsed. The appropriate explanation here might be: 'Because the injection of a hormone overstimulated the motor centre of his cortex,' or some similar story. This, however, merely reinforces the point that the appropriateness of the explanation depends on the context in which the explanation is requested. Thus, whether a bodily activity is
to be called 'physiological process' or 'behavioural act' depends on what constitutes an appropriate explanation of it in a given context.

There is a third class of predicates used to ascribe bodily activities to a person. I shall call these 'bodily happenings,' and use this expression to refer to stomach-rumblings, hiccupping, blushing, coughing, sneezing, yawning, and the like. The members of the class of bodily happenings are border-line cases between behavioural acts and physiological processes. In many contexts a decision has to be made concerning whether the bodily happening should be regarded as a behavioural act or as a physiological process. I shall indicate, on the one hand, what are some of the characteristics of bodily happenings, and, on the other hand, how it could be decided, given that such a decision is required, whether a particular bodily happening should be regarded as a behavioural act or as a physiological process.

It appears that these bodily activities can be thought of as forming a series determined by the extent to which it is possible to control them. Reflexes like knee-jerks and stomach-rumblings seem to be quite uncontrollable. It is conceivable, however, that someone could be taught to control his tendency to blush, or to take preventive measures, e.g. deep breaths, when he recognizes the danger signals of an incoming attack of hiccups. It is not futile to admonish oneself: 'You are old enough not to blush every time a young woman speaks to you,' for improvement in this respect is
possible. Sneezing and coughing, if genuine, are somewhere between hiccups and yawns. But, of course, one can sneeze and cough deliberately, which is hardly possible to do with stomach-rumblings. On the other hand, it is not always possible to control coughs and sneezes, though sometimes we can do it. Yawns, I suppose, are relatively the easiest to control. We can swallow a yawn, and sometimes suppress it altogether. Also, we can succumb to temptation and help a yawn along toward satisfactory completion even in situations where it would be better not to do so. Again, a yawn may be repressed for a long time and then may suddenly overwhelm us. The class of bodily activities referred to as 'bodily happenings' are such that sometimes, and to varying degrees, we can control them, and to a lesser extent, occasionally produce them.

If, for example, I yawn and am asked 'Why did you yawn?', the appropriateness of the answer depends on whether or not I could have helped yawning. If I could have refrained from yawning, but did not, then an explanation based on psychological antecedents seems to be appropriate, e.g. 'I yawned because I wanted everyone to know that sermons bore me.' On the other hand, if the yawn suddenly overcame me and I had no choice, then the appropriate explanation would be a statement of general physiological laws coupled with the initial conditions of my physiology.

Three types of expressions referring to bodily activities have been distinguished: (i) physiological processes,
(ii) behavioural acts, and (iii) bodily happenings. The correct use of 'physiological process' is dependent upon the appropriateness of a physico-chemical or physiological explanation. The correct use of 'behavioural act' depends upon the appropriateness of an explanation based on psychological antecedents. In the correct use of 'bodily happenings,' however, both types of explanation seem to be appropriate at different times. If correct use is dependent upon the nature of the explanation appropriate to the case, then a way must be found to determine which explanation is appropriate.

In addition to the examination of the context in which it is asked: 'Why did the bodily happening occur?,' there is another question which will help decide which type of explanation is relevant, viz., 'Could the agent have been reasonably expected to prevent the occurrence of the bodily happening?' If he could have, but did not do so, then the appropriate explanation should be based on psychological antecedents, while if the agent could not have prevented the occurrence of the bodily happening, then the appropriate explanation is usually physico-chemical or physiological.

'Bodily happening,' therefore is an ambiguous expression. Its ambiguity requires the making of a decision in the case of each bodily activity to which it is used to refer.

---

10I mean to rule out with this qualification such absurdities as committing suicide in order to prevent oneself from yawning.
The decision will have to legislate whether the expression 'bodily happening' is used in the sense in which it is roughy synonymous with the expression 'physiological process,' or in the other sense which makes it almost synonymous with the expression 'behavioural act.' The decision will be based on an examination of the context, and the presence or absence of choice. The outcome of the decision will determine whether the appropriate explanation is based on physiological regularities or on psychological antecedents.

In conclusion, therefore, we can say that the referents of the expressions 'physiological process' and 'behavioural act' are distinguishable according to the explanation which is appropriate in the context in which it is asked: 'Why did it occur?'. 'Physiological process' is used in such a way that an appropriate explanation of the occurrence of its referent is normally in terms of physico-chemical processes, while the occurrence of the referent of 'behavioural act' is appropriately explicable in terms of psychological antecedents. Physico-chemical explanations are usually inappropriate to questions about behavioural acts, e.g. an explanation of what happened in his muscles and nervous system when he ran the race, or before it, does not answer the question 'What reason did he have for running the race?'. Similarly, explanations based on motives and intentions are normally inapplicable if we ask: 'Why did he have a muscular spasm?', for psychological antecedents do not normally explain the functioning of his physiology.
Of course a fully developed identity theory of mind cannot leave the concept of 'psychological antecedent' unanalyzed, as I have done. Some sort of account of all mental events must be given. While recognizing the eventual necessity of giving such an account, I shall not undertake this task here. My problem is only to show that the having of sensations is identical with the occurrence of BP.

(5) The question I am engaged in answering is 'How does behaviour indicate the having of sensations?'. The first step in reaching a satisfactory answer was to show how 'behavioural act,' 'bodily happening,' and 'physiological process' should be understood. From now on I shall use 'behaviour' to refer to behavioural acts and to those bodily happenings that are regarded as behavioural acts. I shall now turn to the second step which is an examination of the relation between sensation and behaviour.

Let us start with the initial, but as we shall see mistaken, assumption that sensation and behaviour are related only contingently. The particular contingent relation alleged to hold between them is causal connection. The having of a particular type of sensation, e.g. pain, and the occurrence of certain types of behaviour, i.e. pain behaviour, are so related, it may be argued, that when a person has a pain this normally causes him to engage in pain behaviour. By 'pain behaviour' is meant such varied behaviour as saying 'I have a pain,' exclaiming 'Ouch,' crying, moaning, flinching, wincing,
grimacing, and the like. The causal connection is established inductively, on the basis of observation, generalization, and prediction.

The statement asserting the existence of a causal connection between pain and pain behaviour is only probably true. It is, however, in principle possible to strengthen the probability of the statement to such a high degree that its truth could be reasonably doubted only if the inductive process in general were doubted. There could in principle be as good a reason for accepting a causal connection between pains and pain behaviour, it is argued, as there is for the acceptance of any other inductively established causal connection.

There is, however, a well-known sceptical objection against the possibility of establishing the causal connection inductively. It seems as certain as anything that people occasionally have pains. It normally presents no difficulty for me to tell whether or not you are in pain, and you, of course, can do the same about me. When in normal circumstances I assert that you are in pain, I do so on the basis of your verbal and non-verbal behaviour. It is possible, however, so the sceptic argues, for you to behave as though you had a pain when in fact you did not: it is possible for you to pretend. Not only do some people engage in pretence, they also do so successfully. Cases of successful pretence are those in which the pretender's behaviour is indistinguishable, by observers, from what his behaviour would be if he were not
pretending. In cases of successful pretence the appropriate pain behaviour and the pretence-behaviour must be indistinguishable. For if this were not the case, then, presumably, the pretence would not be successful.

It is logically possible, so the sceptical objection states, that all cases that are taken to support the existence of a causal connection between the having of a sensation and pain behaviour are cases of completely successful pretence-behaviour; it is logically possible that all instances of apparent pain-behaviour are really instances of pretence-behaviour. If the logical possibility of constant error is not removed, so the sceptic argues, the allegedly causal connection between pain and pain behaviour cannot be established. For the supposition that there is a causal connection presupposes that we can tell whether or not a behavioural act is an instance of pain-behaviour or an instance of pretence-behaviour, and it presupposes also that we can tell that the sensation one is having is a sensation of pain.

In order to meet the sceptical objection one or both of two things must be possible: one must either be able to be directly aware of someone else having a sensation of pain, or there must be some reliable method of inference enabling one to conclude that pain and pain behaviour are causally related. The point of the sceptical objection is that given the assumption that pain and pain behaviour are related only contingently neither of these two things is possible.

Direct awareness of someone else's sensation of pain
is impossible. This is due to the fact there is one and only one way of being directly aware of the having a sensation of pain, viz., feeling it. It is, however, impossible that a person should feel anyone else's sensation of pain. If a person were to feel someone else's sensation of pain he would have to receive the same stimulus, perform the same responses, have the same history to enable him to interpret the stimulus and select the response as the other person does. In other words, if a person were to feel someone else's sensation of pain he would have to be that someone else; and this, of course, is impossible. Nor could he be said to have the other's sensation of pain if both were experiencing identical sensations. In this case each would have a sensation of pain; the fact that the two sensations were tokens of the same type would have no bearing upon their ownership. Both persons could have sensations of pain, and the two sensations would happen to be tokens of the same type.

The other alternative, the sceptic argues, a method of inference yielding certainty, must also be ruled out. Any inference claiming certainty for the step from the occurrence of one event (A) to the occurrence of another event (B), must provide reasons for asserting that from the occurrence of A the occurrence of B can be inferred. In order for the inference to attain certainty the reasons given for the connection between A and B must be conclusive. A conclusive reason for the assertion that B occurred would be knowledge of the occurrence of B obtained in some way other than
through the inference in question. This other way must be either direct observation of the occurrence of B, or else some method of inference yielding certainty. The new method of inference, however, will also have to be vouched for, if it is to have claim of certainty, either by direct observation or by another method of inference. Any method of inference will have to rely, in the last analysis, upon direct observation. Direct observation of someone else having a sensation of pain is impossible. Therefore there is no method of inference enabling one to infer from someone else's pain behaviour that he is having a sensation of pain.

There is a general principle underlying the assumption that there is a purely contingent connection between the having of sensations and behaviour. It is this principle that the sceptical objection calls into question. The principle is this: either sensations and behaviour are related necessarily, or they are related contingently. They cannot be related necessarily, because the logical possibility of total pretence and dissimulation exists. So they must be related contingently. In order, however, to establish the correlation between two purely contingently related events, the presence of each event must be ascertained through direct experience, i.e. non-inferentially. Direct experience of events which are to be correlated normally presents no difficulty. There is, however, a special difficulty in the direct experiencing of the having of sensations: it seems that only the person who has the sensation can experience it
directly, all others can ascertain that someone is having a sensation only through inference. Inference, however, is reliable only if it is ultimately based on the possibility of direct experience. But since direct experience of other people's sensations seems to be impossible, so is the inference from behaviour to the having of a sensation. Consequently the causal connection between sensation and behaviour cannot be established, and as a result of this, behaviour cannot be regarded as an indicator of the having of sensations.

It is patently obvious, however, that we do describe people as having sensations on the basis of their behaviour. If there were only a contingent connection between sensation and behaviour, then given the correctness of the sceptical objection, we could not do what we do all the time. We must conclude, therefore, that the relation between sensation and behaviour cannot be only contingent.

A possible way of trying to meet the sceptical objection is to argue from one's own case. It might be said that when I am directly aware of having a particular sensation, then I behave in particular ways. It is reasonable to expect that if other people have the same type of sensation, they too will manifest the same type of behaviour. The reasonableness of this explanation depends on the assumption that we are all members of the same species, our bodies function in the same way, and that we respond similarly in similar situations.
The difficulty with this argument is that it does not help in escaping the absurdity implied by the sceptical position. For the argument presupposes that, given the sceptical premises, if I am directly aware of having a sensation, then I can and do have justification for saying that it is a sensation I am having, as well as for saying what sort of sensation it is. In other words, the argument presupposes that I can identify and re-identify my own sensations on the basis of direct awareness; and by analogy, I can also identify and re-identify the sensations of others. Neither of these, however, can be done.

In order for me to argue by analogy from my own case I have to be able to tell that what is going on in me is a sensation, and further, that it is a sensation of a certain sort. But how could I be able to do this? I could not rely on some contingent indicator of my sensation, so that I could say that when a certain kind of behaviour is present, I have a certain kind of sensation. For the establishment of the contingent connection between the sensation and the behavioural indicator already presupposes that I know what sort of sensation I am having. If I did not know this, how could the contingent connection be established?

It must be concluded, therefore, that if the relation between sensation and behaviour were purely contingent, then one could not know that what one is having is a sensation, that it is a sensation of a certain sort, and that it is related to a certain kind of behaviour. But it is quite
obvious that we know all this. If we did not, we could not understand the sceptic's words, nor could he. If it follows from a philosophical theory that we do not know that we have a sensation and what sort of sensation it is, then there is something wrong with the theory. Let us, therefore, abandon the assumption that sensation and behaviour are related purely contingently and suppose that sensation and behaviour are related both contingently and non-contingently. We shall see that on this assumption the view that behaviour can function as an indicator of the having of sensations can be justified.

(6) The non-contingent connection that holds between sensation and behaviour is a criterial connection. What I shall do here is to attempt to elucidate certain aspects of the nature of the criterial connection. The concept of criterion is extremely difficult; my discussion of it does not attempt to do anything like full justice to it. My sole purpose here is to explain, as briefly as I can, how the concept of criterion can be used to support the claim that behaviour serves as an indicator for the having of sensations.

We have seen that if one proceeds on the assumption that the connection between sensation and behaviour is only contingent, then behaviour could not be regarded as an indicator of sensation. The reason for this is, to put it slightly differently, that knowing one contingent connection, we can go on to discover other contingent connections. If
behaviour type A is connected with sensation S, then we can
discover another type of behaviour B, and correlate it with
type A, and therefore with sensation S. The question, how­
ever, is how was the first correlation discovered? How did
we discover that behaviour type A is correlated with sensa­
tion S?

Clearly the first such correlation could not have
been discovered on the basis of other correlations, since
ex hypothesi none existed. Thus unless there are some rela­
tions between sensation and behaviour that can be known prior
to the establishment of the first correlation we can never be
justified in saying, on the basis of a person's behaviour,
that he is having a certain sensation. But if there is such
a relation, as there must be, it cannot be a contingent rela­
tion.

It cannot be a contingent relation because before
contingent relations are established a decision has to be
made as to what items are to be correlated. Implicit in
this is the assumption that we have decided what are the items
that could be correlated. But this decision is not about
facts, it is not an empirical matter. It is, rather, the
formulation of a rule. This rule then can be used as a guide
in finding suitable candidates for the correlation. And
finding the candidates is an empirical matter.

So the contingent component in the relation between
sensation and behaviour is the finding of specific correla­
tions between the having of particular sensations and the
occurrence of particular behavioural acts. The non-contingent component is the rule that behaviour is to count as evidence for the having of sensations. Insofar as sensation and behaviour are related non-contingently, behaviour is a criterion of the having of sensations. Insofar as sensation and behaviour are related contingently, behaviour is a symptom of the having of sensations. The question then is: what is to count as criterion, what is to count as symptom, and how are criteria and symptoms related?

What makes something a symptom of X is that normally the presence of the symptom is accompanied by the presence of X. The existence of the symptomatic relation is thus an empirical matter. What makes something a criterion of X is that it is part of the meaning of 'X' that a situation is describable as 'X' if the criterion of X is present. The existence of criterial relation is thus a linguistic matter. It is a matter of fact whether or not one is justified in saying on the basis of the presence of the apparent symptom of X that X is also present. One is justified in saying this if the correlation between the symptom and X is well established. It is a linguistic matter, however, whether or not one is justified by the presence of the criterion in saying that X is also present. The justification for saying this is that one knows the conditions in which a situation is describable as 'X.' It is knowledge of facts that is required for the decision whether something is the symptom of X. It is knowledge of the language that is required for
the decision whether something is the criterion of $X$. What is to count as a symptom of $X$ is learned through observation and experience. What is to count as a criterion of $X$ is learned when one learns the language in which '$X$' is used.

Let us now take an example. What are the conditions in which the presence of a certain percentage of alcohol in the blood is a symptom of intoxication and those in which it is a criterion of intoxication? It would be a symptom if it had been observed that normally when a person is intoxicated there is a certain percentage of alcohol in his blood. This requires that there be independent grounds for describing a person as being intoxicated, as well as there being independent grounds for testing the alcohol content of his blood. The grounds for describing him as intoxicated may be that he has been seen consuming a large quantity of liquor a short time ago, his speech is incoherent, his balance is uncertain, he is claiming that he is dizzy and feels nauseated. The method for testing the alcohol content of his blood may be a chemical analysis of a sample of his blood.

The presence of a certain percentage of alcohol in the blood would be a criterion of intoxication if it had been decided to define 'intoxication,' in part, as 'having a certain percentage of alcohol in the blood.' Other criteria of intoxication may be different types of behaviour.

If alcohol content in the blood is a symptom of intoxication a person could be described as 'discovering' a correlation between them. But if it is a criterion, then there is
no sense in talking about 'discovery.' The connection between the alcohol content of the blood and intoxication would then be a matter of understanding the meaning of 'intoxication,' of knowing the circumstances in which a person can be described as being intoxicated. The existence of the symptomatic relation would be refuted if it were shown that there is no, or only very poor, correlation between alcohol content in the blood and intoxication. The existence of the criterial relation would be refuted if it were shown that 'intoxication' is not used to describe a person when there is more than a certain percentage of alcohol in his blood.

The symptomatic relation is an empirical, contingent relation; its denial, if the relation holds, leads to falsehood. The criterial relation is a linguistic, non-contingent relation; but its denial, if the relation holds, does not lead to self-contradiction. The relation between a criterion of X and X is not entailment. The presence of the criterion of X is neither a logically necessary, nor a logically sufficient, nor a logically necessary and sufficient condition for the truth of a statement asserting the presence of X. The criterial relation allows for the possibility of unusual, abnormal, deviant cases. In these cases either the criterion is present and it would still be false to claim that X is also present, or for cases in which even though the criterion is absent, the claim that X is present would be true.

Let us suppose that the presence of a certain
percentage of alcohol in the blood is the criterion for saying that a person is intoxicated. It is compatible with this that the person is intoxicated even though there is no alcohol in his blood, e.g. if alcohol is introduced directly into his brain so that it produces exactly the same effect as it normally does when it gets there through the blood supply. It is also compatible with there being a criterial connection that a person is not intoxicated even though there is the right percentage of alcohol in his blood, e.g. if he is given some drug that neutralizes the effect of the alcohol.

My contention is that behaviour can serve as an indicator of the having of sensations because there is a non-contingent relation between them, as well as there being a contingent relation. One of the reasons why one is justified in concluding on the basis of the observation of someone's pain behaviour that he is in pain is that it is part of the meaning of 'having a pain' that a person can be so described if he engages in behaviour like flinching, wincing, moaning, crying, exclaiming 'Ouch,' saying 'I have a pain,' and so on. The other reason for being justified in concluding that a person has a pain is that there is a strong empirical correlation between the presence of the sensation of pain and pain behaviour.

A qualification that needs to be added is that it is not pain behaviour per se that is criterial for the presence of pain, but pain behaviour occurring in certain circumstances. An actor giving an excellent performance of a person suffering
from pain would not be properly described as having a pain even though pain behaviour was present. It is important to notice, however, that the circumstances in which pain behaviour occurs and yet does not count as criterion, e.g. acting, imitating, making fun of, are circumstances that presuppose that there are occasions in which pain behaviour is criterial of pain. For unless there were cases in which the actor imitating or making fun of someone else actually did experience the sensation that he is now pretending to have, he would not know how to pretend. What is involved in pretence is acting as though one had the sensation when one does not have it. But in order to do this one must know what it is like to have the sensation on other occasions. And on these genuine occasions pain behaviour is criterial for the presence of pain.

(7) It was argued in (1) and (2) that the truth of the correlation statement is a necessary condition of the truth of the corresponding identity statement. In order to establish the truth of the statement asserting the existence of a perfect correlation between the process of having a particular sensation and the occurrence of BP it is necessary to be able to identify both items of the correlation independently of each other. It was argued in (3) that BP are independently identifiable through brain-readings; that brain-readings are indicators of the occurrence of BP. In (4), (5), and (6) it was argued that the process of having a
particular sensation is independently identifiable through behaviour; different types of behaviour are indicators of the process of having a particular sensation.

In this section I shall consider briefly whether if one is inclined to assert the identity of sensations and some sort of physical processes one should think that these are BP rather than, say, behaviour. After all, the fact required by the identity theory, viz., that behaviour be an indicator of sensation, already establishes a very intimate connection between sensation and behaviour. Why should we then accept the intimate connection, fail to make use of it in asserting the identity statement, and hold instead that it is sensation and BP that are identical?

The reason for this is to be found in the strict requirements of identity. It is a necessary condition of the identity of processes that the occurrence of the two processes be perfectly correlatable, i.e. that either of the two processes should occur if and only if the other process occurs.

Let us take a particular example, e.g. having a pain in the leg, and ask: what are the bodily activities that occur if and only if this sensation occurs? It seems to me that there are only two possible candidates: behaviour and physiological processes. In a trivial sense it is, of course, true that both behaviour and physiological processes occur when a sensation occurs. Indeed, if they did not, the person would be dead. The question is: is there any
particular type of behaviour or physiological process that occurs if and only if a particular type of sensation occurs?

There is, I think, no particular type of behaviour that occurs if and only if a person has a pain in his leg. The person may complain or may refrain from complaining, he may tell others how he feels, or he may keep it to himself, he may be normal and sincere, or he may be a hypochondriac and a liar. And even if there were some one type of behaviour that not only just some persons, but any person, performed when feeling pain in the leg, it would still be the case that the condition of performing it if and only if there is a pain in the leg would be left unsatisfied; for the behaviour might be performed on other occasions as well.

I am not arguing against the view that pain behaviour is criterial for the ascription of pain to persons. On the contrary, I am holding a view that my interpretation of the concept of criterion supports. My position is that there is no one type of behaviour that is either the necessary or the sufficient condition of the ascription of pain to a person. But perhaps there is something going on in persons that could be regarded as necessary and sufficient condition for the ascription of pain. It is this possibility that I shall consider.

The obvious candidate for such a condition are neurological processes. It is, after all, pain receptors that are normally stimulated when a person has a sensation of pain, and it is to the brain that the stimulus travels. A rough
and ready division of neurological processes yields a distinction between processes going on in the central nervous system (located roughly in the brain) and processes going on in the peripheral nervous system (located all over the body). It is certain types of processes going on in the central nervous system, or in the brain, that I claim fulfill the requirement, viz., that they occur if and only if a certain type of sensation occurs.

This view, however, is open to an objection. It might be argued that while a certain sort of brain process, e.g. the scanning activity of the brain, is a necessary condition of a person having a sensation, there is also another necessary condition, viz., the presence of a stimulus. Only the scanning of a stimulus of a certain sort satisfies the condition that the process occurs if and only if a sensation of a certain type occurs. And stimuli come from the peripheral nervous system located all over the body. So it is not just BP alone, but another type of bodily process as well that is needed for the satisfaction of the requirement.

This objection can be met by agreeing that the stimulus, in addition to the scanning activity, is required, but denying that the stimulus is necessarily a bodily process. There is quite an impressive body of empirical evidence for the view that direct external, i.e. non-bodily, stimulation of the brain gives rise to, or is correlatable with persons
having certain types of sensation.\textsuperscript{11} Thus the process that occurs if and only if the person has a certain type of sensation is, probably, the scanning activity of the brain in conjunction with a stimulus that may or may not, but normally does, originate in some part of the body other than the brain.

Hence according to the identity theory the identity statement holds between the process of having a particular sensation and the occurrence of BP; and the corresponding correlation statement holds between the same items. However, the requirement of independent identification for the establishment of the truth of the correlation statement is fulfilled in the case of sensations by behavioural indicators, and in the case of BP by brain-readings.

\textsuperscript{11}Cf. footnote \#10 in Chapter Two.
(1) Introduction

In this chapter I shall further explore the way in which sensations are identified independently of BP through behavioural indicators. It will be argued that there are two different types of behavioural indicators: corrigible and incorrigible. The distinction between corrigible and incorrigible indicators, it will be argued, overlaps to a great extent, but not completely, with the distinction between verbal behavioural indicators, viz., avowals, and non-verbal behavioural indicators, viz., some bodily happenings and non-verbal behavioural acts. According to my view, the typical non-verbal behavioural act is an incorrigible indicator of the having of sensations, while the typical verbal behavioural act is a corrigible indicator of the having of sensations. This, of course, is compatible with the claim that some non-verbal behavioural indicators are corrigible, and that some verbal behavioural indicators are incorrigible. The distinction between corrigible and
incorrigible indicators should be based, according to my view, on the presence or absence of a certain type of cognitive attitude in the person whose sensation is being indicated.

If my contention that there are these two types of behavioural indicators is correct, several important consequences follow. First, the identity theory will have given an account of the nature of avowals. Giving a satisfactory analysis of avowals must be one of the central concerns of any theory of mind, since the problem of avowal is just another form of the problem of how self-knowledge is possible. Second, if my analysis of avowals is correct, it follows that no self-ascription can safely be classed as incorrigible, and that the possibility of incorrigible self-ascriptions is not a distinctive mark of mental states. Third, if there are two distinct types of behavioural indicators, then the identification of sensations will become much more reliable, because identification through either of the two indicators can be checked against identification through the other. Consequently the identity theory will be further strengthened.

(2) The discussion of avowals will be the subject-matter of this chapter, as well as of the next. In the next chapter the problem of whether or not avowals have grounds will be discussed. In this chapter I shall concentrate on the problem of whether or not avowals are incorrigible. It will be argued that typical avowals are corrigible. By this it is
not meant that the typical avowal is mistaken, but rather, that while the typical avowal is true, it is of a sort that permits mistake. In other words, the typical avowal is not logically immune to being mistaken. There is a theory, viz., the expressive theory of avowal, that holds, either that all avowals are incorrigible, or that the typical avowal is incorrigible. The interpretation placed upon the expressive theory of avowals determines which claim one accepts. I shall consider mainly the second alternative, it being the far more defensible claim. According to this view the ground for the claim that the typical avowal is incorrigible is that the typical avowal is similar in certain important respects to non-verbal behavioural indicators.

In Part II of this chapter I shall state what I take to be the expressive theory of avowals. In Part III I shall advance several arguments to show that the theory is incorrect.

(3) I shall use 'avowal' to mean first-person singular, present tense, sentences in the indicative mood. The sentence is an avowal if it is used at least to express, and it will be argued later, also to report, describe, compare, or refer to a psychological state or process or event. Avowals are typically of the form 'I have a . . . ,' 'I feel a . . . ,' 'I am in . . . ,' 'I am . . . ,' where the sentence is completed by a phrase expressing or reporting a psychological state, process, or event.
There are behavioural indicators of sensation that may qualify as avowals, and yet would not fit into the scheme offered above. For instance, a person may exclaim 'Ouch,' and the exclamation may fulfill all conditions of being an avowal except the one about the form of the sentence uttered. It is not relevant for my purposes to give a very tight definition of avowals, if indeed such can be given. The discussion will be restricted to what I shall call, 'typical avowals'; it is acknowledged that there may be atypical avowals.

The question of whether or not avowals are incorrigible depends on the logical possibility of there being an avowal that is sincere and serious, used with the intention of expressing or describing a psychological state, uttered by a normal, healthy, linguistically skilled person who is in no way incapacitated, and yet for the avowal to be mistaken. The term 'mistaken' will be used only in connection with avowals that are erroneous in the above way. The incorrigibility of avowals thus depends on the logical impossibility of there being mistaken avowals. Avowals, of course, may be erroneous in other ways, e.g. be insincere, linguistically inept, etc., without this affecting their incorrigibility. In other words, is it logically possible, i.e. not self-contradictory or absurd to assert, that an avowal, for instance, 'I feel dizzy,' given the above qualifications, could be used to make a false statement? The answer, I think, must be in the affirmative.
II

The Expressive Theory of Avowals (ETA)

(4) The problem

The problem to which the ETA attempts to provide an answer arises in the following way. When I utter an avowal, e.g. 'I am in pain,' my grounds for saying this seem to be different from my grounds for saying that 'He is in pain.' In the case of the third person utterance my ground is usually the behaviour of the person I am observing. But I do not need to observe my own behaviour in order to know that I am in pain. Indeed, it seems that I do not need to observe, infer, or inform myself of anything in order to say with perfect justification that I am in pain. It is easy to be mistaken about the pains of others, but it is not at all easy to see how one could be mistaken about his own pain. The problem of the verification of these two types of sentences is different. In the case of avowals the problem of verification may not even arise, because avowals may need no verification. But if the problem of verification does arise, it is certainly not through my behaviour that I verify that I am really suffering from pain. On the other hand, third person

1As far as I know the ETA does not have a systematic formulation. However, Malcolm, partly in his interpretation of Wittgenstein's position, and partly in the course of making some points of his own, makes greater use of it than anyone else. I shall, therefore, construct what I take to be the ETA by relying mainly on Malcolm's and Wittgenstein's brief remarks about it.
ascriptions of sensations are normally verified on behavioural grounds.

It might be tempting to conclude that the term 'pain' in 'I have a pain' and in 'He has a pain' is used to refer to different kinds of things. In my own case, 'pain' is used to refer to a psychological state, while in the case of others reference is made to publicly observable behaviour. But this conclusion, however tempting, leads to absurdity. It implies that the following state of affairs is possible: my avowal 'I have a pain,' and the third person statement 'He has a pain' used to refer to me, may have different truth values even though they are uttered with the intention of characterizing the same behavioural pattern of the same person at the same time. You and I would then be trying to assert the same proposition by the use of different sentences, and one of these sentences could be used to make a true assertion while the other made a false one, so that we would fail in our attempt to assert the same proposition.

The role of the ETA is to account for the asymmetry of avowals and third person singular sensation ascriptions, and at the same time, to avoid the absurdities mentioned above.

Malcolm\(^2\) states the problem in the following way:

The philosophy of "from one's own case" and behaviorism, though in a sense opposites, make the common

assumption that the first-person, present-tense psychological statements are verified by self-observation. According to the "one's own case" philosophy the self-observation cannot be checked by others; according to behaviorism the self-observation would be by means of outward criteria that are available to all. The first position becomes unintelligible; the second is false for at least many kinds of psychological statements. We are forced to conclude that the first-person psychological statements are not (or hardly ever) verified by self-observation. It follows that they have no verification at all; for if they had a verification it would have to be self-observation. But if sentences like "My head aches" or "I wonder where she is" do not express observations then what do they do? What is the relation between my declaration that my head aches and the fact that my head aches, if the former is not the report of an observation? 

(5) The proposed solution

Having stated the problem, Malcolm, at another place, goes on to provide a solution of it by elaborating a suggestion of Wittgenstein. The question we should ask, according to Malcolm's interpretation of Wittgenstein, is not 'How do words refer to sensations?', but rather 'How does a human being learn the meaning of the names of sensations?'. Wittgenstein's answer to this latter question is:

Words are connected with the primitive, the natural expressions of the sensation and used in their place. A child has hurt himself and he cries; and then adults talk to him and teach him exclamations

---

3O.M., p. 976.


5W.P.I., p. 539.
and, later, sentences. They teach the child new pain behavior.\textsuperscript{6}

Malcolm interprets Wittgenstein as saying that "My words for sensations are used in place of the behavior that is the natural expression of the sensation; they do not refer to it,"\textsuperscript{7} and quotes Wittgenstein's "... the verbal expression of pain replaces crying and does not describe it,"\textsuperscript{8} and some additional supporting passages.\textsuperscript{9}

An important consequence of Malcolm's statement of the EIA is that avowals are incorrigible. Malcolm writes:\textsuperscript{10}

A man cannot be in error as to whether he is in pain; he can not say, 'My leg hurts,' by mistake, any more than he can groan by mistake. It is senseless to suppose that he has wrongly identified a tickle as pain or that he falsely believes that it is in his leg when in fact it is in his shoulder. True, he may be undecided as to whether it is best described as an 'ache' or a 'pain'... but his very indecision shows us what his sensation is, i.e., something between a pain and an ache. His hesitant observation, 'I am not sure whether it is a pain or an ache,' is itself an expression of sensation.

Malcolm supports the view that avowals are incorrigible by three different arguments.


\textsuperscript{7}P.I., p. 539.

\textsuperscript{8}P.I., 244.

\textsuperscript{9}P.I., 256, 271, 288.

\textsuperscript{10}W.P.I., p. 541.
(6) The first argument

The first argument consists of an appeal to and a development of some points Wittgenstein has made. Malcolm writes: 11

Wittgenstein presents us with the suggestion . . . that the first-person sentences are to be thought of as similar to the natural non-verbal, behavioral expressions of psychological states. "My leg hurts," for example, is to be assimilated to crying, limping, holding one's leg. This is a bewildering comparison and one's first thought is that two sorts of things could not be more unlike.

But avowals and non-verbal behaviour are alike in one crucial respect: they are both expressions of psychological states. Non-verbal behaviour is a 'natural' or 'primitive' expression, while avowals are 'sophisticated' expressions. 12 The connection between the primitive and the sophisticated expressions of psychological states is established by the learning of a language. Avowals are linguistic devices used to replace non-verbal behaviour. This view, Malcolm writes, 13 has the merit, inter alia, that:

it breaks the hold on us of the question "How does one know when to say 'My leg hurts?'" for in the light of the analogy this will be as nonsensical as the question "How does one know when to cry, limp, or hold one's leg?".

One of the significant features of the assimilation of avowals to non-verbal behaviour is that, according to Malcolm, it

11O.M., p. 978.
12P.I., 244.
13O.M., p. 978.
removes the temptation to ask the type of question that may lead to an answer which presupposes the possibility of a private language. For if I am justified in asking 'How do I know that I am in a particular psychological state?', then the most tempting answer is to postulate some sort of 'inner sense' or 'introspection.' This would explain how I came to know my psychological state. The explanation, however, presupposes that psychological states are private, that they can be named, and that the name can be used consistently to refer to tokens of a given type of psychological state. Wittgenstein's arguments against the possibility of a private language make this explanation untenable.

Malcolm's point is that if it is shown that 'How do I know . . . ?' type questions about psychological states are inappropriate, and need indeed can never be asked, then the problems of privacy and private language never arise. If avowals merely express and are not used to refer or to describe or to state, then we need not bother about giving an account of how avowals can be used to refer or to describe or to state. And if avowals are merely sophisticated cries, limps, and similar non-verbal behavioural expressions, then they indeed do no more than express; just as their primitive counterparts do. The incorrigibility of avowals follows from this position. For if avowals are only expressions and nothing else, then whenever an avowal occurs only an expression could have occurred. And if avowals are expressions of psychological states, then whenever they occur, they do
express psychological states; indeed, they are logically incapable of doing anything else, since this is the only thing they can do.

(7) The second argument

Malcolm's second argument for the incorrigibility of avowals is based on some observations about the use of certain linguistic devices. Malcolm argues that if a person understands words like 'pain,' then he cannot be in doubt as to whether he has pain. Aune brings out very clearly what I take to be the import of Malcolm's rather striking remarks. Aune writes:

The point can be put this way, although this is not the way that Malcolm puts it: the conditions under which one makes a true statement in uttering the words 'I am in pain' are precisely those conditions under which one exhibits one's understanding of the words one utters. Thus, if a person who is not in pain utters the words 'I am in pain' with no intent to deceive, make a joke, or anything of the sort, he shows, by the mere fact that he utters these words in these circumstances, that he does not possess a complete mastery of the words he is using, and that, although he utters these words, he is not actually making the assertion he seems to be making--and is not, a fortiori, making the false assertion he seems to be making. Hence any genuine, honest assertion that one is in pain is 'incorrigible': for its truth conditions are identical with its being the bona fide assertion that it seems to be.

14 W.P.I., p. 555.
16 Aune, p. 43.
Wittgenstein's remark, Malcolm claims at the same place, points to the same conclusion:

If anyone said 'I do not know if what I have got is a pain or something else,' we should think something like, he does not know what the English word 'pain' means; and we should explain it to him. . . . This explanation, like any other, he might understand right, wrong, or not at all. And he will show which he does by his use of the word, in this as in other cases. If he now said, for example; '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 what to do with.

In other words, avowals are incorrigible because they cannot be mistaken--although they may be intentionally misleading or linguistically inept. Avowals cannot be mistaken, i.e. a linguistically skilled avower cannot use an avowal to give expression to a psychological state and make an unwitting mistake, because if the avower is linguistically skilled and intends to give expression to a psychological state, he does give expression to it. There simply is no such linguistic device as a mistaken avowal. As Malcolm says in a passage already quoted: "A man cannot be in error as to whether he is in pain; he cannot say 'My leg hurts' by mistake."

(8) The third argument

Malcolm's third argument for the incorrigibility of avowals is based on the contention that avowals could be mistaken only if there was some sort of knowledge-claim involved in making them. It makes no sense, Malcolm argues, to

17P.I., 288.
suppose that any knowledge-claim could be involved in the making of avowals. The reason for this is that it makes sense to talk about knowing something only if it makes sense to doubt that that is the case. 'Knowing' derives its meaning partly from a contrast with 'doubting' or 'holding a false belief,' or some synonym. This, Malcolm holds, is the point of Wittgenstein's remarks.

The point about the incorrigibility of the utterance 'I am in pain' lies behind Wittgenstein's reiterated remarks that 'I know I'm in pain' and 'I don't know whether I'm in pain' are both senseless (e.g. 246, 408). Whenever it is meaningless to speak of 'false belief' it is also meaningless to speak of 'knowledge'; and wherever you cannot say 'I don't know ...' you also cannot say 'I know ...' Of course, a philosopher can say of me that I know I am in pain. But "What is it supposed to mean—except perhaps that I am in pain?" (246).

Malcolm's position is based, I think, on the following line of thought. Meaningful use of 'knowing' requires a contrast with 'doubting.' 'I know that I am in pain' would make sense only if 'I doubt that I am in pain' made sense. But one cannot doubt that one is in pain, because doing this presupposes the mastery of the use of 'pain.' One has mastered the use of 'pain,' however, only if one uses 'pain' on appropriate occasions without hesitation. If a person hesitates to say 'I am in pain,' this shows that he is not clear in the use of 'pain,' that he has not mastered its use. The mastery of the use of 'pain' amounts to using 'pain' with the intention of giving expression to a psychological state, only if the

18 W.P.I., p. 542.
avower is in such a psychological state.

Avowals, therefore, are incorrigible. Their incorrigibility, according to Malcolm, does not derive from the possibility that avowals may be mistaken, but as a matter of fact never are. The incorrigibility of avowals is a linguistic matter: avowals cannot be mistaken, because, provided one knows how to use avowals, and uses them only to give expression to a psychological state he actually has, then due to the way in which rules guide the use of avowals, mistaken avowals become logically impossible.

In the next part of the chapter I shall offer several criticisms of the ETA.

III

Criticism of the Expressive Theory of Avowals

(9) The dual role of avowals

One of the important deficiencies of the ETA, as outlined above, is the failure to take into account the implications the dual role of avowals has for the ETA. Many writers argue against the ETA on just this basis, e.g. Mackie.¹⁹

Mackie's statement of the dual role of avowals is perhaps the most succinct; he writes:

The very same phrase, 'I feel cold,' can be a mere expression of the experience of feeling cold, and it can also express a judgement that this experience is of the feeling-cold sort. What is called an avowal is, I suggest, a remark which combines these functions or hesitates between them. If I feel cold, I may shiver, I may say 'Brrr,' or I may say 'I feel cold'; and each of these may perform the same function, each of them can serve as a mere expression of the feeling, a mere reaction to the feeling itself. Saying 'I feel cold' is no doubt a sophisticated, a learned reaction, whereas shivering is unsophisticated and unlearned reaction and nothing more. On the other hand, the remark 'I feel cold' can perform the very different task of stating an explicit judgement that my experience is of a certain sort, and it is only when it performs this function that it is an empirical statement, that is capable of being true or false, or of conveying an item of knowledge.

This forces a dilemma on Malcolm. The attempted assimilation of avowals to non-verbal behaviour can be interpreted either in a weak, or in a strong sense. Assimilation in the weak sense does no more than call attention to the fact that one of the uses of avowals is the performance of the same function as non-verbal behaviour performs. If the assimilation

---


22 Aune, pp. 41-44.

23 Mackie, p. 23.
is to be interpreted in the weak sense, then we may go on to say that avowals have other uses, e.g. reporting, referring, stating, describing, etc. Malcolm sometimes writes as if he preferred the weak assimilation thesis; for instance:

Wittgenstein's examples of the use of 'I am afraid' (pp. 187-188) show how the utterance of that sentence can be a cry of fear, a comparison, an attempt to tell someone how I feel, a confession, a reflection on my state of mind, or something in between.

Assimilation in the strong sense is the thesis that avowals have only and all those functions that non-verbal behavioural expressions of psychological states have. The difference between them is that avowals are sophisticated and learned expressions, while some types of non-verbal behaviour are primitive and unlearned expressions. If the assimilation is interpreted in the strong sense, then the reporting, referring, etc., uses of avowals must be denied. Occasionally Malcolm implies that it is the strong sense of assimilation that he accepts:

When Wittgenstein says that learning the words for sensations is learning "new pain-behavior" and that the words "replace" the natural expressions, he is bringing to light the arresting fact that my sentences about my present sensations have the same logical status as my outcries and facial expressions. [My emphasis]

The dilemma for Malcolm's version of the Eta arises in the following way: if avowals are assimilated to non-verbal

---

24 W.P.I., p. 542.

behaviour in the weak sense, then avowals may be used to refer, to report, or to state. And if avowals can be used in these ways, then, as it will be argued, they are not incorrigible, because it is always at least logically possible that a fully intended reference fails to refer, that a report rests on a mistake, that a description mis-describes. On the other hand, if avowals are assimilated to non-verbal behaviour in the strong sense, then avowals could under no circumstances be other than expressions of psychological states. This would guarantee that avowals must be free of mistakes, but only at the price of absurdity. For avowals occurring in the following contexts would have to be regarded as not describing, or referring, or reporting, or stating, but only as expressing. Consider some examples: (i) "Are you all right?", "No, I feel dizzy"; (ii) "What do you feel if I press here?", asks the doctor, "I feel pain" replies the patient; (iii) "Why do you want to open the window?", "Because I feel hot." I think only a philosopher in the grip of a theory could deny that these avowals are used to state facts.

A possible—albeit inadequate—way of meeting this predicament is to argue that avowals are by definition incorrigible. It may be argued that 'avowal' should be reserved for only those tokens of first-person, present tense, indicative sentences about psychological states that are used to express psychological states. This expedient would make it self-contradictory to deny that avowals are incorrigible, provided the assimilation of avowals to non-verbal behaviour
holds. But this would be begging the question; for the problem is whether avowals can be assimilated to non-verbal behaviour. This is not a problem that can be solved by stipulating interest in only one of the many possible uses of avowals. If avowals are denied to have reporting, or stating, or describing, or referring use, the ETA leads to absurdity. If it is allowed that avowals sometimes are used in these ways, then additional arguments are needed in favour of the view "that the verbal utterance and the natural pain-behavior are each (as I shall express it) 'incorrigible.'"

It seems to me that the strong version of the ETA is mistaken. Henceforth, I shall mean by the 'ETA,' the 'weak version of the ETA.' There will be two main questions that I shall be asking in the remaining part of the chapter: first, are avowals incorrigible in their referring, reporting, stating, comparing, etc., uses? and second, is the referring, reporting, stating, comparing, etc., uses of avowals their typical use? I shall answer negatively to the first, and affirmatively to the second question.

(10) Example of a mistaken avowal

In order to support the claim that avowals are not incorrigible, i.e. that there are some avowals that could be false, it is necessary to provide an example of a mistaken avowal. In order to do this I shall postulate a futuristic

---

26 W.P.I., p. 541.
situation. My case is implausible, but logically possible. All that is needed, however, is the recognition that there may be circumstances in which an avowal would be regarded as false. If this is admitted, then incorrigibility, as a logical thesis, is refuted.

Let us assume that neurophysiology has advanced to such an extent that all the physiological correlates of psychological states are known. It is known, for instance, exactly what goes on in the brain when a person feels pain, has an after-image, feels dizzy, introspects, and so on. The science is so well developed that neurophysiologists can infer what experience a person is having by observing his brain; they can also infer what goes on in his brain by listening to his avowal. In both cases, subsequent testing always confirms the neurophysiologist's inference.

Let us further assume that this advanced science of neurophysiology has been present for generations in its fully developed state. In all cases when it has been tested, and there have been thousands, it was confirmed. Not once, in all the years, has a disconfirming instance occurred. The supposition I am making is that a perfect correlation has been found between psychological states and physiological processes. Or, in other words, that the correlation statement postulated previously is true.27

It is important to notice that this assumption does

27See Chapter Three, Section (1).
not presuppose that physiological processes are identical with psychological states, i.e. I am not begging the question. The presupposition is only that physiological processes are perfectly correlatable with psychological states.

Let us assume that there is a person who, under these circumstances, utters the avowal 'I feel dizzy,' and uses it with the serious and sincere intention of describing his present psychological state. How could he be mistaken? I suggest that he would be mistaken if the following conditions held:

(i) When subjected to non-verbal behavioural tests he shows no signs of being dizzy, e.g. he can stand straight without falling or leaning while his eyes are tightly shut, he can walk on a straight line without losing his balance or missing a step, etc.

(ii) Upon examination, experienced judges of human behaviour, who know the person in question well, find no signs of insincerity, confusion, attempts at frivolity, or linguistic ineptitude. They conclude that the person meant what he said.

(iii) The neuro-physiologist detects signs that the person is in another psychological state, viz., he finds the neuro-physiological correlate of a state of nausea.

(iv) When the person is questioned, after he has made the avowal, he freely states that he feels nauseated, that usually when he feels nauseated he also feels dizzy, and that he has come to expect dizziness to accompany nausea.

(v) The neurophysiologist detects an abnormality in the
neuro-physiological correlate of introspection, and the abnormality is such that it could account for the person's belief that he is feeling dizzy.

(vi) Very shortly after this experience the person is hypnotized and questioned about his experiences during the time in question. Under hypnosis he claims: first, that he remembers having said that 'I feel dizzy'; second, that he remembers having meant what he said; third, that he does not remember having felt dizzy.

(vii) When the person is confronted with the evidence against his avowal he changes his mind and admits the possibility that he was mistaken. I think that on the basis of these considerations it ought to be concluded that the person's avowal was mistaken, and that therefore, avowals are not logically incorrigible.

(11) Refutation of objections

This example may be attacked on several different grounds. It may be argued that if the neurophysiologist told one story and the avower told another story, then one ought to accept the avower's story because the neurophysiologist's theory, or perhaps his instruments, may turn out to have been wrong. Armstrong\(^{28}\) meets the objection in this way:

There is no doubt that this is a possible rejoinder, and that if brain theory were not well founded it would be a rational rejoinder; but why is it the

\(^{28}\) Armstrong, p. 425.
rejoinder that we must accept? Any hypothesis whatever may be "protected" if we are prepared to make a sufficient number of ad hoc assumptions, but to protect a hypothesis indefinitely is not a rational attitude. The fact that we could cling to every deliverance of introspection even against the best-attested brain theory does nothing to show that it would be incorrect to side with the brain technician. In fact, I think it would be rational to side with him against the deliverance of introspection provided that brain theory was well founded.

It might be objected, however, that the neurophysiological theory that we accept rather than accepting the avowal, is itself developed by correlating neurophysiological processes and avowals. Since the theory tacitly presupposes the incorrigibility of avowals, it cannot be used to call into question part of its very foundation. But this objection is not very forceful. In the first place, the scientific theory need not be founded upon the assumption that avowals are incorrigible statements. It is quite sufficient for the theory to be founded upon the assumption that avowals are empirically true statements.

In the second place, even if we accept that the neurophysiological theory discussed has to rely upon incorrigible avowals, we may still hold that not all avowals are incorrigible. That is, just because avowals are needed for the establishment of a neurophysiological theory it does not follow that the very same theory could not call some avowals into question. As Armstrong observes: "our knowledge of the physical world is got by perception but this does not prevent

\[29\] Armstrong, p. 425.
us casting doubt on some perceptions."

In the third place, I find it doubtful that avowals are necessary requirements for the establishment of a neurophysiological theory. For if this assumption were true, then we would be deprived forever from having a neurophysiological theory of animal behaviour. I think that a neurophysiological theory could be developed by relying upon non-verbal behaviour only.

It is quite true, on the other hand, that if such mistaken avowals as I described occurred frequently, then the neurophysiological theory would have to be called into question. The situation discussed, however, was ex hypothesi of the sort that one and only one avowal turned out to be mistaken. My point was that in the hypothetical situation it would be more reasonable to regard the avowal as being mistaken, rather than to abandon the extremely well established theory.

Another objection might be that to proceed on the assumption that some avowals may be mistaken would require a radical reorganization of our conceptual framework. But I do not think that this is an argument; it is rather a complaint. If it were discovered that contrary to my firm belief I have no head, and nor does anyone else, this would be even a greater shock to our conceptual framework. But no one claims that because of this the statement 'I have a head' is incorrigible. Conceptual frameworks are, and have to be, reorganized from time to time. I do not see why we should
exclude the possibility that the time is now.

(12) Temporal objections against incorrigibility

The alleged incorrigibility of avowals, however, may be attacked on other grounds as well. Armstrong\textsuperscript{30} advances two closely connected arguments against the incorrigibility thesis on the basis of temporal considerations. Avowals are, by definition, in the present tense. One reason for this is that if avowals were about past psychological states as well, then memory-claims would be directly involved in the making of an avowal. But no one would want to hold that particular memory-claims are incorrigible. So, if memory-claims were involved in the making of some avowals, these avowal could not be regarded as incorrigible. This commits the champions of the incorrigibility of avowals to the view that the logical character of sentences used to express or to report psychological states changes from one fraction of a second to another. Suppose I say: 'I was in pain half a second ago,' and 'I am in pain now.' Since memory-claims are corrigible the first utterance may be mistaken while, if the ETA holds, the second utterance is logically immune from mistakes. So something must have happened during the fraction of a second that passed while the two utterances were made. Whatever happened must be such that it be capable of changing the status of the utterance from empirical probability to

\textsuperscript{30}Armstrong, pp. 419-421.
to logical certainty. I cannot imagine what would accomplish this.

Furthermore, if we consider 'I am in pain,' we must take it to mean: 'I am in pain now.' But what is the period of time to which 'now' refers? If the pain lasts for a long time, then there is no problem. But what if the pain is just a sudden shot of pain that lasts for an instant? What if the instant is shorter in duration than the time required to utter the avowal? What if the 'now' in the avowal changes into 'then,' since the pain is no longer there? Presumably, what started out to be an avowal turns into a report with a memory-claim built into it. And since particular memory-claims are corrigible, so will be the would-be avowals. Hence in order to hold that avowals are incorrigible, it will have to be denied that one can avow the having of sensations that last for a shorter time than it would take to express them. But surely this is an absurd conclusion: we frequently express the having of sensations that last for a very brief period. Why should expressions of brief psychological states be corrigible, while expressions of longer psychological states be incorrigible? There is an answer to this, but it is not one that the upholders of the ETA would want to take. It may be argued that what enables us to utter incorrigible avowals is introspective knowledge, and to acquire this takes time. If the possibility of introspective knowledge is allowed, then avowals could no longer be regarded as being only sophisticated expressions replacing primitive expressions. Some
avowals then would make knowledge-claims. I shall examine now the implications of this possibility.

(13) Two kinds of awareness

Mackie writes: 31

The strongest argument against incorrigibility, it seems to me, is this. It is one thing to have an experience and it is another to reflect on it, to notice what sort of experience you are having . . . to realize what sort your experience is you . . . must contemplate it, not merely have it. Once this is admitted, it follows that the result of this contemplation, the judgement that the experience is, say, of cold and not of warmth or pain, is open to the possibility of error and the corresponding statement is in principle corrigible. This is the crux of the matter.

It seems to me that the underlying error of the incorrigibility argument is the running together of two different kinds of awareness: the awareness involved in the having or feeling a sensation and the awareness involved in thinking about the sensation one has or feels. I shall call these two kinds of awareness 'non-cognitive' and 'cognitive' awareness, respectively.

It is perhaps misleading to label both the having and the feeling of sensations, on the one hand, and the thinking about the sensation felt, on the other hand, as 'awareness,' because there is a fundamental difference between them. Non-cognitive awareness is a mode of awareness, it is a way of being aware. There is no distinction in its case between the psychological state and the awareness. The psychological

31Mackie, p. 22.
state is the awareness. Cognitive awareness, however, is different. There is need for distinguishing the object of awareness, i.e. the psychological state, from the awareness itself, i.e. some sort of cognitive attitude, e.g. describing, comparing, thinking, evaluating, etc. I shall present arguments now in favour of there being these two kinds of awareness.

Let us recall the argument for the dual role of avowals. Avowals can be used either as expressions, in which case they may be assimilated to non-verbal behaviour, or they can be used as reports, descriptions, referring phrases, and so on, in which case they cannot be assimilated to non-verbal behaviour. My present suggestion is that when avowals are used only as expressions, the attitude of the avower toward the psychological state being expressed is that of non-cognitive awareness. His awareness and his psychological state are one and the same thing. He just feels dizzy, or he just feels pain, nausea, fatigue, or what not. There is no reflection, no self-scrutiny involved; all the owner is doing is feeling the psychological state. But he may do more, and when the avower issues avowals that are not only expressions, but also reports, or descriptions, or referring phrases, he does do more. What he does is to develop a cognitive attitude toward the psychological state that he previously had or felt without any thought of it.
Aune, in trying to make the same distinction, writes:

This kind of awareness is a cognitive one, which resists being lumped together with brute feelings. To be aware of an ache or pain in this sense—or to be aware of it only as language-user can be aware of it—is not, that is to say, just a matter of having or feeling it; it is rather a matter of being able to recognize it for what it is: of being able to describe it, to think about it, to worry about it, and perhaps to appreciate its significance in at least the ethical life of man. All this plainly cannot be done naturally, preconsciously, without any learning whatever.

What needs to be shown now is that the two kinds of awareness are indeed distinct. I shall argue that a consideration of higher mammals and very young children will bear out the distinction. In both cases there is non-cognitive awareness without there also being cognitive awareness.

Sensations like pain, hunger, thirst, fatigue, dizziness can be ascribed to animals on behavioural grounds. Animal trainers, pet owners, experimental scientists have no great difficulty in identifying, recognizing, and even predicting the sensations of animals. There is no doubt that animals have sensations, at least in the trivial sense that the physiological processes that occur when humans have pain, are hungry, thirsty, tired, or dizzy, also occur in animals. But do animals feel pain, hunger, thirst, etc.? What reason do we have for saying that they do? I think the reason is observation of their behaviour. It is not just that animals

---

32 Aune, pp. 50-51.
behave in certain ways in certain situations, but also, that they seem to prefer some situations to others: animals manifest both 'approach' and 'avoidance' behaviour. If they felt nothing, but were merely vessels for physiological processes, they would treat all situations alike. But do they have cognitive awareness in addition to having non-cognitive awareness? Do animals recognize, compare, think about, describe the sensations they have? I do not think this allows of a simple answer. They certainly do not have all the cognitive attitudes human beings have, e.g. they do not describe their sensations. It seems, however, that they may have some, e.g. animals may be able to recognize and compare sensations. But if this is the case, could it not be argued that cognitive and non-cognitive awarenesses always occur together, and are indistinguishable? I think not. For even if we allow that animals have some cognitive attitudes toward their sensations, it must be admitted that there was a first time for the having of the first sensation. At that time questions about comparison or recognition could not have arisen: there was nothing to which the sensation could have been compared. So, there must be at least one occasion, viz., the first one, when animals only feel a sensation, but do not have cognitive attitudes toward it.

What is true of higher mammals, in this respect, is also true of young children. There must have been a first experience of the first sensation for them, when they only felt the sensation without having any cognitive attitude
toward it. There is, however, a difference between the claims about animals and babies. The difference is that my claim about animals cannot be verified, while my claim about babies can be. Many of us have very early memories, and if someone does not, psychoanalysts may be prepared to elicit them. Suppose that a person remembers the first occasion on which he had a sensation. He may well remember how totally overwhelmed he was by the first sensation which happened to be, say, hunger. He just felt it, he was hungry through and through, he identified or recognized nothing, he was just hungry. The argument from first experience becomes even more plausible if we remember that a sensation need not be unique for us to have only a non-cognitive awareness of it: erotic sensations are typically of this sort. I think, therefore, that it is possible to have non-cognitive awareness without cognitive awareness, and consequently that the two types of awareness should be distinguished. Ayer33 argues in the same way:

Knowing what one's thoughts and feelings are, as distinct from merely having them, may be taken to consist in being able to give a true report of them. The two normally go together, but they are not logically connected. Animals and young children, who have not acquired the use of language have feelings and images of which they may be said to be conscious, but, on this view, they do not know that they have them; they may think in images, but they do not know what it is that they are thinking.

Logical arguments against incorrigibility

If it is recognized that there are these two types of awareness, then Mackie's\textsuperscript{34} argument against the incorrigibility of avowals holds. The linguistic equivocation of the two roles of avowals and the epistemological equivocation of the two different kinds of awareness both have at their root the same philosophical error. The incorrigibility thesis, Mackie\textsuperscript{35} argues, ultimately rests on the following argument:

When I am feeling cold the coldness is just the content of my experience: it is what I am aware of; and if I am aware of coldness, how could I be ignorant of it or in error about it? To be ignorant would be not to be aware of it, and to be in error would be to be aware of something other than coldness.

This argument can be interpreted in three different ways, none of which supports the incorrigibility thesis.

First, it may be taken, as it was not meant to be taken, as the assertion that when I am feeling cold then it is cold that I am feeling. Or, to put it generally, if I feel a sensation of a certain sort, then the sensation I feel is of that sort. This is trivial and true, but does nothing to support the claim that when I am cognitively aware of a sensation, then the avowal used to report the sensation is logically incorrigible. For it may be true that some of my avowals are actually mistaken, and the trivial truth of the above interpretation would still hold. If my sensation

\textsuperscript{34}Mackie, pp. 22-26.

\textsuperscript{35}Mackie, p. 24.
is of a certain sort, then it is of that sort. The incorrigibility thesis, however, asserts that I cannot be mistaken about what sort of sensation I am feeling.

The second interpretation is to suppose that whenever one is feeling, say, cold, one knows that it is cold that is felt. In other words, whenever one is non-cognitively aware of cold, one is also cognitively aware of cold. As we have seen, however, it is possible to have non-cognitive awareness without also having cognitive awareness, e.g. the sensations of animals and babies, as well as first sensations, are of this sort. This establishes that non-cognitive and cognitive awareness are different kinds of things. So the second interpretation of the incorrigibility thesis is that two distinct occurrences are related in such a way that if one occurs, the other always occurs. But this interpretation also fails to support the incorrigibility thesis. In the first place, it is not true, as we have seen, that cognitive awareness always accompanies non-cognitive awareness. If this is accepted, then the incorrigibility thesis can amount to no more than the trivial interpretation whose irrelevance is shown above. In the second place, if it is denied that non-cognitive awareness could occur without cognitive awareness, the incorrigibility thesis still does not hold. For we may agree, for the sake of argument, that the two kinds of awareness, as a matter of fact, always occur together. But why should we suppose, as the incorrigibility thesis must suppose, that it is logically necessary that
they occur together? Logical necessity does not follow from even universal empirical coincidence.

Let us turn then to the third interpretation of the above argument for incorrigibility. Mackie\textsuperscript{36} states it in the following way:

Feeling cold is taken to be a cognitive experience, an awareness of cold: it is not quite the same thing as knowing-that-the-experience-I-am-having-is-of-the-feeling-cold-sort, but not quite different either. It is suggested that feeling cold is somehow an awareness of itself. It is clear that if anything were an awareness of itself it would, just by occurring, provide itself with an object, so that whenever this cognition occurred it would be true, but this suggestion contains the seeds of a paradox.

According to this interpretation of the incorrigibility thesis, cognitive awareness of a sensation would be, as it were, contained by the occurrence of the sensation. This could happen in two ways: cognitive awareness of, say, cold would have to be thought of either as being part of the sensation of cold, or the sensation of cold would have to be regarded as being a form or mode of cognitive awareness. However, absurdity results from both alternatives.

If cognitive awareness of cold were part of the sensation of cold, then the following picture would result: a given instance of cognitive awareness, \textit{viz.}, cognitive awareness\textsubscript{1}, would be awareness of a sensation of cold. But since cognitive awareness is part of the sensation, cognitive awareness\textsubscript{1} of a sensation of cold would be awareness of the

\textsuperscript{36}Mackie, p. 24.
sensation of cold and of cognitive awareness$_2$. And then we would have to ask: what is the relation between cognitive awareness$_2$ and the sensation of cold? According to this interpretation, cognitive awareness$_2$ would have to be regarded as part of the sensation of cold. So cognitive awareness$_2$ would be awareness of the sensation of cold and of cognitive awareness$_3$ . . . and so on ad infinitum. Thus the view that cognitive awareness is part of the sensation leads to the necessity of postulating a never-ending series of cognitive awarenesses.

Someone, however, might not be deterred by the infinite regress and still take this view. Those so inclined would have to meet another objection. If the sensation of cold is made up of distinct parts, it is legitimate to enquire what the relation between these parts is. But if this question is asked, it will be seen that the relation is a factual one, i.e. two parts of sensations always occur together. If the relation is factual, then it is conceivable that cognitive awareness may not occur when the sensation occurs. Hence there would be no reason for supposing that an avowal reporting a sensation would be logically incorrigible.

If we take the other alternative and hold that the having of a sensation of cold is a form or mode of cognitive awareness, absurdity results again. There is no doubt that at least sometimes we are cognitively aware of sensations. But if a given sensation were a mode of cognitive awareness, then when we were cognitively aware of sensations we would
be aware of a mode of cognitive awareness, and so infinite regress would follow.

However, there is an additional argument against the view that the having of a sensation is a mode of cognitive awareness. If this were true, then one could not distinguish between different types of experiences. For in order to distinguish between them it would be necessary to compare them. One could compare them, however, only by being aware of both experiences. But if each experience were only a mode or form of cognitive awareness without an object, this could not be done. There would only be two distinct modes of cognitive awareness, occurring perhaps simultaneously, without there being any way, short of postulating some sort of super-awareness, of comparing them.

To sum up: I have argued in support of the view that the incorrigibility thesis of the ETA rests on the equivocation of two kinds of awareness and of the two roles of avowals. When the thesis is examined analytically, it may be interpreted in three different ways. First, incorrigibility may rest on the trivial claim that when I am feeling cold, it is cold that I am feeling. This does not answer the question of how we know in the first place that it is cold we feel. Second, incorrigibility may rest on the empirical coincidence of non-cognitive and cognitive awareness. But from empirical coincidence, even if it is universal, one cannot pass on to logical certainty. Third, if cognitive awareness is regarded either as part of the having of sensations,
or if the having of sensations is regarded as a mode or form of cognitive awareness, absurdity will follow. There is, therefore, no way in which the incorrigibility thesis can be significantly held, provided the dual role of avowals and the distinction between the two kinds of awareness are recognized.

(15) The contrast argument for incorrigibility

The upholders of the incorrigibility thesis may argue, however, that it is the supposition that a person may have cognitive awareness of a sensation that cannot be held significantly. For, as Malcolm\textsuperscript{37} argues, such cognitive claims as 'I know that . . . ' or 'I believe that . . . ' or 'I think that . . . ' cannot be meaningfully prefixed to avowals, because these claims would make sense only if there would be possible contrasting counterclaims. But it makes no sense, Malcolm argues, to say that 'I do not know whether I am in pain' or that 'I falsely believe that I have a pain.' If the claim that I do not know makes no sense, then the claim that I know loses its significance. For there are, according to Malcolm,\textsuperscript{38} certain words in the language that operate in contrasting pairs. These words derive their meaning from the possibility of contrast, and words like 'knowing,' 'believing,' 'thinking' are of this sort. The argument

\textsuperscript{37}See Section (8) in this chapter.

against the incorrigibility thesis results in attempting to talk about knowing, believing, thinking, etc., where contrast is not possible. Consequently, insofar as the arguments against incorrigibility presuppose a cognitive attitude on the part of the avower, the argument is incoherent.

Perhaps the best, though by no means the only, refutation of Malcolm's position is to give actual examples of avowals where it is perfectly legitimate to talk about knowing, believing, or thinking. Suppose that I am asked at one point to list all the facts I know about myself. The list would include statements that I know to be true of myself at a given time. It would be unobjectionable, I think, to include in this list statements about my experiences during the postulated time interval. And if I happened to have a headache, it would be proper to say, in this context, that one of the facts about myself is that 'I know that I have a headache.' Or alternatively, consider a case Malcolm himself gives. 39 There is a person who has a sensation somewhere in between a pain and an ache. He is asked to say what sort of sensation he is having. After some hesitation, during which he introspects, he says somewhat uncertainly 'I think I have a pain.' Or consider cases of people having congenital insensitivity to pain. 40 This rare abnormality occurs when

39 W.P.I., p. 541.

a person is incapable of feeling pain even in circumstances that others would normally consider to be extremely painful. Persons having this disorder are of course able to use the word 'pain' quite well. They learn its use on the basis of third person verbal and non-verbal behaviour. They themselves, however, have never felt pain, have never had occasion to avow the having of a pain, and have never behaved as though they had pain. Suppose they are cured, and under laboratory conditions given such stimuli as are normally considered painful. Might not the person then say, upon first feeling pain, 'I believe I have a pain.' I think all these avowals are quite meaningful and that Malcolm is mistaken in thinking otherwise.\textsuperscript{41}

But Malcolm's case rests on other mistakes as well. It is simply not the case that in all instances when legitimate knowledge claims can be made it must be possible to doubt. In fact, there are some cases in which it is impossible to doubt and remain within rational discourse. What I have in mind is the case of \textit{a priori} knowledge. If there is \textit{a priori} knowledge, then the knowledge will rest entirely on the analysis of the concepts involved. So the conditions for knowing, and the conditions for understanding the use of the concepts would exactly coincide. There would be no room for doubt. Aune\textsuperscript{42} gives the following argument that can be taken

\textsuperscript{41}For additional refuting instances see B. Aune, 'Knowing and Merely Thinking,' \textit{Phil. Studies}, 12 (1961), pp. 53-58.

\textsuperscript{42}Aune, p. 47.
to support the conclusion I want to arrive at:

If a person did not know that one plus one equals two, it would probably be held that he did not know what addition is, or that he did not know what is meant by 'one,' 'two,' or 'equals.' Under these conditions it could be argued that a person could not possibly doubt whether one plus one equals two, since to doubt this would require him to exercise these primitive arithmetical concepts. If it were true, however, that knowing always implies the possibility of doubt, we would have to conclude that 'one plus one equals two' states something that could not be said to be known—which is extremely doubtful considering the normal usage of the word 'know.'

Therefore, if Malcolm wants to hold that the contrast between knowing and doubting is required in certain special cases, e.g. in the case of avowals, in order that knowledge claims could be made, additional arguments are required. The appeal to an allegedly universal feature of language is unsatisfactory, since what Malcolm takes to be a universal feature is not that.

(16) Incorrigibility and language learning

It may be thought that additional arguments are provided by Malcolm's analysis of the learning of the use of avowals. In order to assert the incorrigibility of avowals it is necessary either to argue that avowals have only an expressive use, or to argue that avowals are incorrigible in their reporting, stating, referring, etc., uses as well as in their expressive use. I have argued that if avowals are used to report, state, etc., then insofar as avowals are used in

\footnote{\textsuperscript{43}}\textsuperscript{43}See Section (7) in this chapter.
these ways they are not incorrigible. I shall argue now that the incorrigibility claim cannot be made out successfully even if we accept that avowals have only an expressive use. The reasons for this are: first, if avowals are regarded as part of language, then they cannot be assimilated to non-verbal behavioural expression, so the analysis leading to incorrigibility, and presupposing the possibility of assimilation, cannot be used in the case of avowals; second, in learning the expressive use of avowals there are always cognitive considerations involved, and for this reason avowals cannot be purely expressive.

Malcolm gives a brief indication of how he supposes the learning of avowals takes place. Malcolm writes that as a child

... grows older and begins to talk it will normally come about that often when his behaviour and circumstances are those of a person in pain he will say the words 'It hurts,' or some synonymous ones; and hardly ever will he say them when either his behaviour or circumstances do not satisfy the original criterion of pain. This development fulfills our criterion of his understanding those words.

But surely this cannot be the case; more needs to be said. On the basis of precisely analogous considerations we might say that a well conditioned parrot, i.e. one that emits 'It hurts' only when it is in pain, also "fulfills our criterion of his understanding those words." By emitting the sounds the parrot is not using language. Nor would a person be using

---

language if he did no more than emit the sounds. Now such an utterance of 'It hurts' is indeed incorrigible, and incorrigible in precisely the same way as non-verbal behaviour is. But this kind of utterance is not an avowal, for the simple reason that it does not involve understanding the sounds uttered. As Aune puts it:

To understand the conventional significance of a group of words is at least to understand what one can use these words to say or do . . . merely to exercise a habit of responding with special noises in certain circumstances is not, by itself, to use language at all. One typically uses something for a purpose or end. And so with a sentence: one uses it, because it has a certain conventional significance, to inform, to amuse, to deceive, and so on.

The asymmetry between non-verbal pain behaviour and an utterance of 'It hurts,' said with understanding, is due to the fact that the avower must be prepared to clarify, to use synonyms for, to re-express, to discuss, to argue about, to joke about, to appreciate comments about, his avowal. The performance of a non-verbal natural and primitive behavioural expression of pain needs to fulfill none of these requirements. So if avowals were incorrigible, they could not be incorrigible in the same way non-verbal behaviour is. Malcolm's mistaken contention is that they are incorrigible in the same way, and for the same reason.

The incorrigibility thesis, however, can be attacked on different linguistic grounds as well. If avowals are

---

\[45\] Aune, p. 37.
incorrigible because they are purely expressive, then the absurd conclusion will follow that one understands only those avowals that he himself utters. According to Malcolm, if I say 'I am in pain,' the question of whether I am really in pain cannot arise for me. But if someone else says 'I am in pain,' the question of whether he is really in pain can arise for me. In order to understand avowals I must understand how any and all tokens of avowals are used; this includes understanding the avowals of other people. But in order for me to understand the avowals of others I must know what counts for and against the avowals, I must know under what conditions I should accept the avowals, when I should ask for clarification, and so on. In getting to know all these things, cognitive elements are introduced into my understanding of the use of avowals. However, my use of avowals is not learned from my own case, but rather it is learned on the basis of third person uses of avowals. This, as I understand it, is the assumption underlying Malcolm's proposed solution of the other minds problem. But as we have seen, one's understanding of the avowals of others is permeated with cognitive elements. Consequently my own avowals cannot be any freer of these cognitive elements than can be my understanding of the avowals of others. Thus, it follows from Malcolm's own assumptions, that avowals cannot be purely expressive. And insofar as they are not purely expressive, they are incorrigible.
IV

(17) Conclusion

The disagreement with the ETA centered on two questions: first, are avowals incorrigible in their non-expressive uses?, and second, are avowals incorrigible in their expressive uses?

On the basis of the arguments presented the following conclusions can, I think, be drawn:

First, avowals are not incorrigible in their non-expressive uses. That is, it is logically possible that a non-expressively used avowel could be mistaken. Of course, as a matter of fact, avowals are rarely or never mistaken.

Second, if avowals are incorrigible at all, they must be so in those expressive uses in which they closely resemble non-verbal behavioural acts.

Third, it is very doubtful, however, that a purely expressive avowel could ever occur, since avowals are linguistic devices the use of which seems always to require a certain amount of cognitive awareness.

Fourth, the problem of whether or not there are purely expressive avowals is not a very important one for my purposes. The important point is that, in agreement with the ETA, I hold that the typical instance of non-verbal behavioural act is incorrigible, while the typical instance of an avowel I hold, in disagreement with the ETA, is corrigible. It is not very important to determine into which class purely
expressive avowals belong, if indeed there are any.

Fifth, there are, therefore, two different kinds of behavioural indicators of sensations: corrigible, viz., the typical non-expressively used avowal, and incorrigible, viz., the typical non-verbal behavioural act. This implies that knowledge issuing in incorrigible statements is not a distinctive mark of the mental; and that the identification of sensations through behavioural indications is more reliable than originally supposed since the two methods of identification can be used to confirm each other.
CHAPTER FIVE

THE GROUNDEDNESS OF AVOWALS

I

(1) Introduction

In this chapter I shall examine another argument for the incorrigibility of avowals. The argument is based on the assumption that avowals are, as Gasking \(^1\) calls them, 'infallible' and 'ungrounded.' The argument for the incorrigibility of avowals is based on the supposed linguistic fact about avowals that they simply are not used in a way which would make it possible for mistakes to arise, and on the epistemic feature of avowals that they need have no ground to be justified. It is argued that the linguistic fact about avowals guarantees their infallibility, while the epistemic feature guarantees their ungroundedness.

I shall argue that the supposed infallibility of avowals does not solve the problem of how a person can have knowledge of his sensations. The infallibility of avowals

does suggest a linguistic explanation of the role avowals play in language, but it does not explain how it is that when avowals are used to make a cognitive claim they are based on allegedly indubitable knowledge. It is in answer to this question that Gasking, and also Lean,\(^2\) argue for the ungroundedness of avowals.

I shall argue that avowals are not ungrounded and consequently they are not incorrigible. It is logically possible, I hold, for there to be mistaken avowals. By 'mistaken' avowal I mean that it is logically possible for there to be an avowal that is used for making an unwittingly false statement about one's psychological states. Gasking and Lean do not deny that avowals may be erroneous in the sense that an avowal may be insincere, linguistically inept, uttered in jest, or made by a person who is in some way incapacitated, e.g., hypnotized. My disagreement with Gasking and Lean is about the possibility of there being a particular way in which avowals may be erroneous, viz., by being used for making an unwittingly false statement. The term 'mistaken' will be used only in connection with the possibility of this particular kind of error. So, in the present discussion, the alleged incorrigibility of avowals will depend on their supposed ungroundedness; the ungroundedness of avowals would remove the possibility of their being mistaken,

but not the possibility that they may be erroneous in some other way.

First I shall consider Gasking's statement of the problem. After this I shall discuss Gasking's solution of the problem in terms of the alleged infallibility and ungroundedness of avowals. I shall then consider Lean's critical discussion of Gasking's solution, and state what I take to be Lean's solution. Lastly, in terms of Gasking's and Lean's arguments I shall state the identity theory's account of avowals.

(2) Gasking's statement of the problem

Gasking\textsuperscript{3} argues that the problem of avowals is made up of a major and a minor problem. "The major problem is to understand how, of two statements asserting the same proposition, one could be fallible and the other an infallible statement." The minor problem is "that of understanding how, of two statements asserting the same proposition, one could be grounded and the other ungrounded." In my discussion of avowals Gasking's major problem will become the minor problem, and vice versa.

Let us start with Gasking's distinction between ego and non-ego statements. Gasking\textsuperscript{4} writes:

'I feel hungry' is a present tense statement in the first person singular--I shall call it, for short,

\textsuperscript{3}Gasking, p. 156.
\textsuperscript{4}Gasking, p. 154.
an 'ego-statement.' Corresponding to any ego-statement are various non-ego statements, in the past tense or in other persons, which say the same thing, with reference to the same person and the same time. To the ego-statement 'I feel hungry'; which I now make, there correspond such non-ego statements as your statement to me in the words 'You feel hungry.'

My use of 'avowal' overlaps very largely with Gasing's use of 'ego-statement.' The difference for my purposes is negligible.

Concerning the infallibility of avowals Gasing writes:

Of any statement whatever it makes sense to say that the person making it was untruthful in doing so, or that he was lying or insincere. And it usually also makes sense to say that he is mistaken in saying what he does. But of some ego-statements, according to a widely held philosophical view, it makes no sense to say this; e.g., when Smith says 'I am in terrible pain' it does not, on this view, make sense to say that Smith might be mistaken. I shall call such utterances . . . 'infallible statements.'

Infallible ego-statements, i.e. ego-statements whose corresponding non-ego statements are fallible, Gasing calls 'exceptionally infallible.' Avowals are not only exceptionally infallible, but ungrounded as well. Gasking writes:

You would commonly make such a non-ego statement as 'Smith feels hungry' on such grounds as that he has not eaten anything for some time, that he is now bolting a large meal, and so on. And it is frequently held that unless you have some ground for them you are not entitled to make statements of this sort. But for an ego-statement of the same proposition the speaker need not have any

\[
\begin{align*}
5 & \text{Gasking, p. 154.} \\
6 & \text{Gasking, p. 155.}
\end{align*}
\]
grounds, Smith may well be perfectly entitled to say 'I feel hungry' even though he has no grounds for saying this. Indeed, the very notion of a ground for such non-ego statements seems an absurdity. An ego-statement which needs no grounds, whereas non-ego statements of the same proposition require grounds, may be called 'exceptionally ungrounded.'

All exceptionally infallible avowals are also exceptionally ungrounded. If it makes no sense to suppose that an avowal could be mistaken, then it makes no sense to ask for its ground. The relation, however, is not symmetrical. There may be exceptionally ungrounded avowals which are not exceptionally infallible. Gasking's example is: 'I just remembered the formula for common salt.' The person saying this needs no ground for it; he remembers and that alone justifies him in his avowal. But what he remembers may not be the formula for common salt. Consequently, his statement, while ungrounded, may be fallible.

(3) Gasking's solution of the problem

I shall first discuss Gasking's solution of his major, and my minor, problem, viz., the fallibility of avowals. And then Gasking's solution of his minor, and my major problem, viz., the ungroundedness of avowals, will be examined.

The infallibility of avowals

The requirement is to explain how avowals came to have only an infallible use in language. The question could be asked generally, i.e. how did language come to have this

---

7Gasking, p. 155.
particular device? The answer should be an historical account given by linguists and philologists. There is, however, also a particular question, i.e. "how does an individual human being come to be able to use sentences like 'I'm hungry' or 'I'm afraid'--and how is it that throughout his life he keeps consistently to one use of these locutions, that is, the correct one?"\(^8\)

In reply to this question Gasking makes use of Skinner's\(^9\) distinction between two types of conditioning. There are two different policies one could adopt in teaching a child the use of a certain sentence. The first policy is to "link the sentence with a certain sort of objective stimulation, by encouraging his utterance of it (giving him approval or any other reward) if and only if he utters that sentence in a situation of that sort."\(^10\) "The second policy would be always to behave towards the learner in exactly the same way whenever he utters the sentence, no matter what the objective situation."\(^11\) Gasking\(^12\) then concludes that it is evident that the first policy . . . must necessarily be adopted when we wish to teach someone to make fallible true-or-false-statement. . . . But the second policy works in a completely different way. Here no discriminating between situations is

---

\(^8\) Gasking, p. 162.


\(^10\) Gasking, pp. 162-163.

\(^11\) Gasking, p. 163.

\(^12\) Gasking, p. 163.
required--here there is no right and wrong--for the results of making the utterance depend in no way upon the sort of situation in which it is made.

In teaching the use of avowals it is, of course, the second policy that is followed. If we want to teach a child the use of the avowal 'I am cold,' for instance, then whenever the child utters the avowel we behave in a particular way, viz., take action to increase the temperature. The consistent application of this policy, regardless of the actual temperature, will teach the child that the avowel should be used for one purpose only, i.e. to alleviate cold.

The linguistic and psychological explanations of the infallibility of avowals qua linguistic devices, however, leave unanswered what I take to be the major problem. The linguistic and psychological explanations tell us how avowals became infallible once the sensations avowed were known. The major problem, however, is to explain what it is in a person that enables him to know with apparent certitude the sensations which he later avows.

In order to distinguish the problem of the infallibility of avowals from the problem of the ungroundedness of avowals consider the following analogy: a computer is programmed to type out 'machine error' every time a mechanical defect occurs inside the computer, and not otherwise. On a particular occasion the computer in fact types out 'machine error.' There are two questions I want to ask about the performance of the computer: (i) what is the explanation of the fact
that the computer types out 'machine error' only when there is a mechanical defect inside it? and (ii) what must be true of the computer for the explanation given in (i) to hold? If it is not pressed too far there is an analogy between (i) and the question 'How are infallible avowals possible?', and between (ii) and the question 'How is the indubitable knowledge upon which avowals are based possible?'

The reply to the first question presents no difficulties. The explanation of the fact that the computer types out 'machine error' only if there is one, is that the computer has been programmed to do precisely this; just as children are taught to avow only if they have something to avow. The analogy, however, should not be carried beyond this point, for children, unlike computers, can be insincere. This does not destroy the analogy, it merely shows that children can do things that the computer in question cannot do.

The second question, if asked about the computer, is once again not puzzling. That which has to be true of the computer so that it could be programmed in the required way can be specified by drawing up a list of the mechanical parts which should be built into the computer in order that it be capable of performing certain sorts of tasks. The fact that the computer has mechanical devices built into it which make it possible so to programme it as to type out 'machine error' is ex hypothesi free of mechanical defects.
only if there is one, explains that the computer never makes a mistake.

I want to suggest, by analogy, that the physiological constitution of human beings explains the fact that the knowledge upon which avowals are based is indubitable. That is, to the question: 'How is it possible that whenever X says "I am in pain," he is in pain?' I give the answer in two steps: (i) linguistic training and the nature of language explain the fact that avowals are used in such a way that mistakes are ruled out, and (ii) the physiological constitution of human beings explains the fact that one is never mistaken in avowing psychological states. The indubitability of the knowledge upon which avowals are based is explained by the fact that we feel a sensation if and only if we have a sensation, and that we never feel a sensation if we do not have one. What this amounts to, I shall discuss later.

To complete the analogy: as the computer types out 'machine error' only if it has one, because it has been programmed this way, so a sincere, linguistically skilled, normal person says seriously 'I am in pain' only if he feels a pain, because he was taught to use avowals in this way. Further, as the mechanical constitution of computers explains how they can be programmed to be accurate in typing out 'machine error,' so the physiological constitution of human beings explains how they can be taught to avow the having of a pain, for example, if and only if they have a pain.

The analogy, however, does not help us over the
crucial point. It makes sense to say of a man that he is cognitively or non-cognitively aware of having a pain when he avows it, but the ascription of either kind of awareness to computers is, to say the least, controversial. It is necessary, therefore, to enquire into the problem of what it is that human beings do when they are aware of a psychological state that computers do not do when they perform the tasks for which they have been programmed. This leads us to the problem of the alleged ungroundedness of avowals.

(4) Gasking's account of the ungroundedness of avowals

Gasking argues that avowals as well as some non-ego statements are ungrounded. In his view it is possible for an avowal and for a non-ego statement to be ungrounded and yet be justified. Gasking holds, however, that the explanation of the incorrigibility of avowals is not to be found in the ungroundedness of avowals, but rather in their infallibility. In agreement with Lean, I hold that the crux of the problem is the alleged ungroundedness of avowals. Gasking's linguistic explanation of the infallibility of avowals accounts for our language having, and for people learning, the use of this linguistic device. But it does not explain what it is in human beings that makes them capable of learning the use of avowals. If there were a kind of knowledge that required no grounds, it would explain the incorrigibility

14Gasking, pp. 156-161.
of avowals. Lean argues, if I interpret him correctly, that there is such knowledge. I shall argue that the knowledge involved in the making of avowals, i.e. cognitive awareness, is not ungrounded, and consequently avowals are not incorrigible.

Gasking produces several examples of ungrounded non-ego statements. 'The ceiling looks orange to me' has no ground, for there is nothing that one could cite as evidence over and above the fact that it looks orange to him. Mrs. Smith's ability to tell whether her husband is tired, just by looking at him, without being able to say how she knows it, is another case in point. The hypothetical Mr. Timex who can always tell what time it is without consulting a time-piece also seems to support Gasking's contention that there are ungrounded non-ego statements. I shall discuss the chicken-sexer example.

A trained chicken-sexer can tell by looking at day-old chicks whether or not they are male or female; his judgements are reliable. Chicken-sexers are trained: they are shown photographs of day-old chicks and asked to guess whether a chick is going to be a hen or a cock. The trainer, who knows the sex of the photographed chicks, encourages the trainee if he guesses correctly, and corrects him if the guess is incorrect. Gradually, as the training progresses, he guesses correctly more and more frequently. Eventually, Gasking\textsuperscript{15}

\textsuperscript{15}Gasking, p. 159.
concludes,

if he is a good pupil, he reaches the stage where he can regularly get it right. When fully trained he can justifiably make the statement 'That chick is a male.' But he has no ground for his state-

Of course, one is inclined to say that there must be some-
thing about the physical make-up of the chicks that a trained observer notices, and it is this that is the ground of his judgement. This, however, is an unsupported assumption. For neither trained observers, nor the chicken-sexers themselves have any idea about what it is that enables the chicken-sexer to judge correctly. The chicken-sexer does not know whether his judgement has a ground, so he has no ground for it. Nevertheless, Gasking argues, he is certainly justified in making the judgement, because he is a trained chicken-sexer and his predictions about the sex of day-old chicks are normally correct.

The justification and the ground of a statement, therefore, have to be distinguished. The statements of the chicken-sexer are ungrounded but justified, so is the state-
ment that 'The ceiling looks orange to me,' or Mr. Timex's pronouncements about the exact time. Insofar as these state-
ments are ungrounded, they are like avowals. The difference between ungrounded non-ego statements and ungrounded avowals is that the former are fallible, while avowals are infallible. Gasking's explanation of the infallibility of avowals is, as we have seen, in terms of language-teaching, i.e. certain sentences come to have only a correct use.
In Gasking's opinion, there is no asymmetry, in respect to groundedness, between avowals and non-ego statements; the asymmetry holds between grounded and ungrounded statements. The class of ungrounded statements contains avowals as well as non-ego statements. With respect to fallibility there is indeed an asymmetry between avowals and non-ego statements. The explanation of this, however, is not to be found in the nature of our knowledge of the psychological state avowed, but rather in the way in which language is used to express them.

(5) Lean's criticism of Gasking

Lean accepts, with some reluctance, Gasking's distinction between the ground and justification of statements. He agrees with Gasking that both avowals and non-ego statements stand in need of justification. Lean, however, disagrees with Gasking's inclusion of some non-ego statements into the class of ungrounded statements. Avowals, and only avowals, are exceptionally ungrounded and exceptionally infallible. Non-ego statements are always grounded, and consequently they are always, at least in principle, fallible. 16

Both avowals and non-ego statements have to be justified. They are, however, justified differently. The justification of the avowal 'I have a toothache' is having a toothache. The fact avowed justifies the avowal. Non-ego

16 Lean, pp. 177-182.
statements are not justified by the fact stated. The justification of the chicken-sexer's assertion that a particular chick is going to be a hen is not just that the chick grows up to be a hen. The chicken-sexer would still have been justified in his judgement if the chick turned out to be a cock later on. The chicken-sexer's justification is that he was usually right in the past. This justification holds in the present regardless of whether or not the statement turns out to be correct.

Consider Mrs. Smith as another example. She can tell whether her husband is tired when he comes home from work, but she cannot say how she knows it—she just knows it. Her justification for saying that Mr. Smith is tired cannot just be that he is indeed tired. Smith could be tired without his wife having any justification for asserting that he is; and he may not be tired, just act as though he were, and his wife would still be justified in saying that he is tired. Nothing but feeling tired justifies Smith's avowal that 'I feel tired,' whereas Mrs. Smith is justified on other bases in asserting that Smith is tired.

Lean does not, but we ought to, notice that the sentence 'I feel tired' can be used to make two different types of statements. 'I feel tired' may be used either as an inference or as an avowal. If Mr. Smith's statement 'I feel tired' is only an inference, then he has the very same ground and justification for it as does Mrs. Smith. He may say 'I normally pay attention to my paper,' 'I am not usually
clumsy,' therefore, 'I must be tired.' The logically different situation, i.e. the one which Lean needs for his purposes, is in which 'I feel tired' is not used as an inference. Smith's utterance of it still needs justification, but it has no ground. Smith's justification of the avowal-use of 'I feel tired' does not rest upon inference from his behaviour. The inferential use of 'I feel tired' may be contradicted by behavioural evidence. The avowing-use, Lean needs to argue, could not be contradicted by any evidence extraneous to Smith's feeling tired.

The justification of ungrounded avowals is the fact avowed, and nothing outside it; the justification of non-ego statements, whatever it is, is not the fact asserted. The justification of avowals and non-ego statements, therefore, Lean argues, must be distinguished. The distinction, Lean goes on to say, is based on the distinction between a statement not having a ground and a statement whose ground the person asserting it does not know. Statements of the first sort are genuinely ungrounded; statements of the second sort are grounded, but their ground is not known. Avowals are genuinely ungrounded statements, while the non-ego statements made by Mrs. Smith, Mr. Timex, and the chicken-sexer, and the person saying that the 'ceiling looks orange to me' are ungrounded only in the sense that the person asserting them does not know the ground for the assertion. Avowals have no grounds, while apparently ungrounded non-ego statements do
have grounds.\textsuperscript{17}

Let us now ask: What do Gasking and Lean mean by 'ground'? Gasking does not say; he hopes, I think, that his examples will show what he means. But they do not, and Lean is correct in criticizing Gasking for using 'ground' as though its meaning were obvious. Lean's attempt at telling us what 'ground' is deserves to be quoted in toto. He says\textsuperscript{18} that the 'notion of "being grounded" that is, or at least more closely approximates to, the one in common currency' is the following: that the

statement-situation be one in which a person judges something to be the case, in response to or in consequence of something else—something he believes, or that he has learned, or that he once did or does now perceive or sense—of which something he need not at the moment, nor even ever, be consciously aware, but which is for him associated in such a way with what he judges or asserts to be the facts, that if it were appreciably and significantly different in content or character, or absent, or else not thus associated 'in his mind,' he would judge or state the facts to be otherwise, or would be unable to judge or state at all, except haphazardly.

As far as I can make out Lean is saying that whenever a person makes a judgement, \(y\), he does so because of its relation to a state of affairs, \(x\). This relation is one in which if \(x\) were different, \(y\) would be different. The state of affairs, \(x\), has as one of its associated consequences, the judgement, \(y\).

'Being a ground' is being what the state of affairs, \(x\), is to

\textsuperscript{17}\textit{Lean}, p. 182.
\textsuperscript{18}\textit{Lean}, p. 179.
the judgement, y. It is not necessary for the person making
the judgement, y, to realize that the state of affairs, x,
has led to the judgement, y. The two may be associated with­
out the person, in whose 'mind' they are associated, knowing
it.

If this is employed as the criterion of groundedness,
then, Lean19 holds, all non-ego statements will be found to
be grounded and all avowals ungrounded. The statements of
the chicken-sexer, Mr. Timex, Mrs. Smith and the person mak­
ing the colour-judgement, according to Lean, are all grounded
because

what is important is that in making their state­
ments they do so in consequence of having learned
an association between what they assert and some­
thing else to which they respond, even if they are
not consciously aware of this.

In the case of avowals, however, what they verbalize is not
something else than that to which the speaker is responding.
What is verbalized in an avowal is a psychological state,
and the very same psychological state is also that to which
the speaker is responding.

Avowals, therefore, are ungrounded, while non-ego
statements are grounded. But we still need an explanation
of the ungroundedness of avowals.

(6) Lean's solution of the problem

In the case of avowals, infallibility and

ungroundedness come to the same thing. Avowals are infallible because they are ungrounded.²⁰ They are ungrounded in the sense that it is logically impossible for them to have a ground. Avowals require no ground because they make no claim whatsoever. And there couldn't even be such a thing as a ground for them, because a 'ground' is something in the speaker's awareness, consciousness, or 'mind,' which leads him to make a claim about something else; whereas an avowal refers or pertains only to what is in the speaker's 'mind' itself.²¹

Our skill in talking about sensations, images, etc., is acquired on the model of our skill of talking about physical objects; in this respect Gasking and Lean are in agreement. Physical object language has a claim-use built into it; this, however, is not the case with the derivative avowal-use. In uttering an avowal, Lean argues, a person does not make a claim. The essential element of the claim-use, that of referring to something else over and above the psychological state, is lacking in the case of avowals. In the normal claim-use the speaker, as a result of having a sensation, feeling, thought, etc., makes a statement about something else. The 'something else,' a particular state of affairs, has come to be associated with his psychological state; he expects it on the basis of his past experience. But this expectation, as well as the association with something else, is absent from

²⁰Lean, p. 183.
²¹Lean, p. 184.
avowals. Hence avowals do not make a claim. 22

Avowals are about sensations, feelings, images, etc. Ground and reference are identical, and therefore, indistinguishable in their case. It is not just that avowals are not grounded on anything, it is also that they could not be grounded. Avowals could not be grounded for everything else is grounded upon what avowals are about, i.e., psychological states. All statements are grounded, ultimately, upon sensations, images, thoughts, etc., therefore, those psychological states themselves cannot be grounded. They are like primitive descriptive elements which cannot themselves be described. To demand a ground for psychological states is to say that no arguments are conclusive, that they all lead to infinite regress, that we never know anything with certainty. 23

The infallibility of avowals follows from their ungroundedness. A statement is fallible only if there is a discrepancy between its ground and reference. In the case of avowals there is not and cannot be such a discrepancy. The discrepancy, however, is present even in the simplest of perceptual judgements. Avowals, and only avowals, are exceptionally ungrounded and exceptionally infallible, because their ground and reference are identical. 24

---

22 Lean, p. 185.
23 Lean, pp. 185-186.
24 Lean, p. 186.
II

(7) The identity theory's account

If the ground-justification distinction is employed to state the difference between avowals and non-ego statements, the way in which Gasking's, Lean's, and my solution of the problem differ can be stated with some clarity.

Gasking's view is that avowals and some non-ego statements are alike in being ungrounded but justified. Some non-ego statements may be justified even though their ground is not known, but there are no ungrounded non-ego statements. Gasking and Lean disagree, because Lean thinks that avowals are uniquely ungrounded, and the infallibility of avowals follows from their ungroundedness. Gasking thinks that avowals are not uniquely ungrounded, and accounts for the infallibility of avowals in terms of language teaching. My position is that avowals and some non-ego statements are alike in being justified without their ground being known. I hold, however, that avowals and non-ego statements do have ground, although their ground may not be known. So I am in partial agreement and in partial disagreement with both Gasking and Lean. I agree with Gasking that avowals and non-ego statements are not asymmetrical in respect to groundedness. But I disagree with Gasking that they are alike in being ungrounded; I think rather that avowals and some non-ego statements are alike in being grounded without it being known what their ground is. I agree with Lean that Gasking overlooked the
distinction between a statement having no ground and its ground not being known. And I agree with Lean that Gasking's examples of allegedly ungrounded non-ego statements are in fact grounded non-ego statements whose ground is not known. But I disagree with Lean's view that avowals are uniquely ungrounded. I think that avowals, like Gasking's allegedly ungrounded non-ego statements, are grounded, but it may not be known what their ground is. Avowals, according to my view, are not incorrigible because, on the one hand, they are grounded. On the other hand, their alleged infallibility does not touch the basic question of whether indubitable knowledge of one's psychological states is possible.

Lean advances several arguments for the unique ungroundedness of avowals. Of these I shall consider the two most important ones: (i) that avowals are not used to make a claim, and (ii) that there is no distinction between the ground and reference of avowals, or in other words, avowals have no ground. I shall argue against these in turn.

(8) Do avowals make a claim?

Lean argues that avowals do not make a claim because the essential requirement for a claim use is absent. In the claim use of an expression, the speaker, as a result of some sensation, feeling, image, thought, belief, or the like (of which he may or may not be consciously aware), makes a statement about something else, that he takes to be the case. That the speaker

25Lean, pp. 184-185.
should assert the particular belief or conviction expressed in his statement, is of course the result of his having somehow acquired, together with appropriate language habits, some more or less spontaneous association between the particular visual sensation he has (to take the simplest case), and something else. This 'something else' will be something well correlated with it in his past experience, that he has come to expect, quite automatically, when he has that particular sensation—namely, all the relevant possible experiences (to put it briefly) that are associated with there being a physical object before him that he is visually perceiving.

Avowals are about, say, sensations and not about what is associated with sensations. Hence the 'something else' is lacking, and consequently avowals do not make a claim. One has to enquire, however, what the 'sensations' are that Lean is talking about. He says:

avowals are merely verbalizations, so to speak of the describable content of the speaker's imagery, sensations, feelings, attitudes, inclinations, and the like—in the sense in which, say the particular after-image that one person may be 'seeing' . . . may be said to differ in 'content' from that which another person is experiencing . . . (even though in either case the reference, of course, is to nothing that exists independently of the respective awareness).

It seems then that sensations form a particular way of being aware, and have contents inseparable from awareness. But what is this content? What is the cash-value of 'having a kind of awareness'? What does it mean to say that the 'kind of awareness' is a mental event? Is being aware, together with the appropriate content, public or private? Lean does not say. He does not even give a hint of how he would

26 Lean, p. 134.
analyse 'content of awareness,' 'conscious awareness,' and 'contents of awareness being associated with something else,' and many other expressions.

Let us ask, therefore, whether or not the sensations Lean is talking about have publicly observable manifestations. If we assume that sensations, according to Lean's interpretation, can have no publicly observable manifestations, i.e. they are logically private, then we are taking the first step on the path to absurdity. If, however, we assume that sensations are not logically private, then we have to assume also that the occurrence of sensation of a certain sort is correlatable with the occurrence of some sort of behavioural and/or physiological process. The existence of this correlation is sufficient to refute Lean's argument that avowals do not make a claim.

My point is that when a person utters an avowal he may, though he need not, use it to describe or state or report that there is something going on in him. The justification for the descriptive or stating or reporting use of avowals is that the avower has learned in the past that when such-and-such is going on in him, he should describe, report, or state that it is going on in him by the use of avowals. The case of there being a person, of something going on, of the person having learned to describe what is going on, conforms perfectly, I submit, to the example of claim-making given by Lean, and quoted above.

A person who says that he is observing the sunset may
claim that there is something going on in the world. This claim is based on his having learned to describe the sort of thing that he now observes. Analogously, a person who says that he is having a toothache may claim that there is something going on, not in the world, but in him. This claim is based on his having learned to describe the sort of thing that is now going on. The difference between the visual sensation and the sensation of pain is that in the former case what is going on is outside the skin, while in the latter case it is inside the skin. Apart from this difference, in both cases there is something going on; and the justification for this claim is that in the past, when the same sort of thing was going on, the person was taught that this claim could be made. Avowals may, therefore, be used to make a claim.

(9) The ground and reference of avowals

The important point for Lean is that even if avowals can, in some sense, be said to make a claim this is not the sense in which non-ego statements make a claim. Consider the non-ego statement 'I see a tree.' The ground and reference of it must be distinguished; the statement makes a claim precisely in virtue of this distinction. The reference of a non-ego statement is always a fact about the world; in this case the fact that there is a tree. But what is the ground of the non-ego statement? The ground, according to Lean,27 is

27Lean, p. 185.
this complex expectation that is at the bottom of his
notion of there actually being before him the particu-
lar fact he is asserting; and it is also what his
stating of his statement reflects and can be said, in
this sense, to 'express.' But the statement, itself,
which he utters, does not express this. For the
statement does not refer to his sensory awareness
itself or its specific content, nor to its being
correlated with something else in his experience
that he has come to expect—and he may not even be
aware of any of these things as such. His statement,
therefore, can not in any straightforward sense be
said to be about any of this. These would (on my
vision of 'grounded,' though not on Mr. Gasking's)
even in the simplest case of perceptual reporting,
be the ground of the speaker's statement.

Thus if I understand Lean correctly, the reference of a non-
ego statement is to the state of affairs asserted by the
statement, while the ground of a statement is the expectation,
based on past experience, that when a person is in a given
psychological state, e.g. having a visual sensation of a tree,
then there corresponds to his psychological state something
else, e.g. a real tree.

Non-ego statements make a claim because in their case
the reference of the statement, i.e. a certain state of
affairs, is distinct from the ground, i.e. the expectation
that corresponding to a psychological state there is a certain
state of affairs. But avowals, according to Lean, do not make
a claim in this sense. Lean argues:28

Avowals are ungrounded because they do not involve
this essential twofold relationship that is at the
bottom of a genuine claim. For with avowals, their
'ground' or 'basis' is precisely what they refer to—
is identical with what they say or express. They
are merely verbalizations of the speaker's mental

28 Lean, p. 185.
states, sensations or the like, per se; and they do not refer to any expectation the speaker might have about something else on the basis of them, or even about themselves.

So, according to Lean, avowals are ungrounded because, as he question-beggingly puts it, there is no difference between their ground and reference. They refer to a psychological state, viz., the one being avowed, and they are grounded on that very same psychological state, for there is no additional state of affairs which corresponds to that psychological state. It is in this sense that we should interpret Lean's remarks\textsuperscript{29} that avowals are in this respect comparable to analytic or tautological utterances, the 'truth' of which lies only in the explicit or implicit linguistic principles of which they themselves are the verbal formulations, and is not contingent upon anything outside of themselves, which they might then fail of representing.

I shall argue that Lean's account is mistaken, that avowals, like non-ego statements, do involve the 'twofold relationship' essential to making a claim, that, in a word, the ground and reference of avowals are distinct.

The reference of the avowal 'I have a toothache' is the toothache. Lean, I suppose, would not disagree. But what is the ground of the avowal? Surely, its ground is the expectation, based on past experience, that whenever I am cognitively aware of a toothache there is a toothache. The claim which avowals may be used to make is that the avower is

\textsuperscript{29} Lean, p. 186.
cognitively aware of some fact about himself. Lean, I am sure, would disagree.

It could be argued against my suggestion that the toothache, to which reference is made, and the awareness, the putative ground of the avowal, are one and the same. The attempt to distinguish the having of a toothache and being aware of the having of a toothache is, therefore, doomed to failure. But I do not think that this is the case. Let us remind ourselves of the distinction between non-cognitive and cognitive awareness. It was argued that non-cognitive awareness is a form or mode of being aware; there is no distinction between non-cognitive awareness of a psychological state and the psychological state. But cognitive awareness is an attitude a person may or may not have toward his psychological state. It is the attitude involved in describing, comparing, thinking about, evaluating the psychological state of which one is, indeed, cannot help being, non-cognitively aware.

Now, my point is that Lean is quite right in saying that the ground and reference of avowals are not distinct if only non-cognitive awareness is present. But Lean is mistaken in denying that the ground and reference of avowals are not distinct if cognitive awareness is also present. Lean's mistake is due to his failure to draw the distinction between the two kinds of awareness.

30See Chapter Four, Section (13).
When a person is cognitively aware of the psychological state that he avows, e.g. in saying 'I have a toothache,' the reference of his avowal is to the psychological state, viz., to the having of a toothache. Similarly, the reference of the non-ego statement 'I see a tree' is to the tree. The ground of the non-ego statement is the expectation that to the psychological state of having a visual sensation of a tree there corresponds a tree in the world. In the same way, the ground of the avowal is the expectation that to the psychological state of being cognitively aware of a toothache there corresponds a toothache. The difference between the two cases is that the correlate of the visual sensation is a fact 'in the world,' while the correlate of the cognitive awareness is a fact 'in the mind.' The correlate of visual sensation is not a psychological state, while the correlate of cognitive awareness is a psychological state.

These two differences, however, are not such as to undermine the distinction between the ground and reference of avowals of those psychological states of which the avower is cognitively aware. For the dichotomy of 'in the world' and 'in the mind' need be taken seriously only by a dualist. Identity theorists can and do deny that the distinction expresses anything basic about persons and their relationship to the environment. If 'mind' and mental events, states, and processes, are analyzable in terms of, or are reducible to non-mental occurrences,—occurrences not fundamentally unlike those that go on 'in the world'—then occurrences 'in the
mind' and occurrences 'in the world' will not be fundamentally different. So this difference between non-ego statements and avowals should not present a problem to the identity theory. Nor do I think that the other difference should stop us from noting the similarity, in respect to groundedness, between non-ego statements and avowals. It is true that the ground of the avowal of a psychological state is another psychological state, while, if we accept Lean's account of 'ground,' the ground of a non-ego statement is not the state of affairs to which it refers, but a psychological state. That is, the ground and reference of avowals are the same type of states, while the ground and reference of non-ego statements are different types of states. One may well agree to this and still hold, in disagreement with Lean, that what an avowal may be used to refer to, and that upon which an avowal is grounded, are different. Perhaps not different in kind, but just different particulars of the same type.

(10) What is meant by 'ground?'

It is possible to object to my argument for the groundedness of avowals in the following way. It might be said that I argue against Lean's view that avowals are ungrounded, and in doing so I accept Lean's analysis of 'ground.' If this analysis is mistaken, then it may well be that if the correct analysis of 'ground' is given, avowals will prove to be ungrounded. And as a matter of fact, Lean's analysis is unusual. Its major shortcoming is, I think, that
it confuses logical and psychological considerations. For if grounds were expectations based on the past experience that corresponding to a certain psychological state, e.g. to having a visual sensation, there is a certain state of affairs 'in the world,' then a perfectly ordinary situation would have to be regarded as being very close to impossible. If grounds were what Lean says they are, there could not be statements with unknown grounds. If grounds were expectations based on past experience, then for a statement to be grounded sometimes, someone must have had the expectation. But clearly this need not be the case. Lean himself agrees that the chicken-sexer's utterances have a ground, yet neither he nor anyone else knows what this ground is. So Lean has to suggest that the chicken-sexer has an expectation based on past experience; an expectation that he is not now and never was aware of having, and that no one else is or ever was aware of either. The question of whether the chicken-sexer's statements have a ground would then depend on a psychological analysis leading to the confirmation of Lean's hypothesis.

It seems to me, however, that there is a much less cumbersome way of analysing 'ground.' I shall say that a statement \( p \) has a ground if and only if there is another statement \( q \) such that (i) \( q \) implies \( p \), and (ii) the truth of \( q \) is evidence for the truth of \( p \). The second requirement is meant to exclude the objection that if only (i) holds, then all statements could be said to have grounds, because \( q \) may imply \( p \) in virtue of \( q \) being self-contradictory, or \( q \) being
"either p or r." That is, if only (i) holds, it is always logically possible to find a statement that implies, in a purely formal sense, another statement. So I want the implication holding between q and p, in the definition of 'ground' to be understood as an evidential relation.

It is perfectly compatible with this analysis of 'ground' that a statement has a ground and that no one knows this to be so, or simpliciter, no one knows what the ground of the statement is. Yet a person may be justified in asserting the statement even if he does not know that it has a ground, because his past assertions that the statement holds have turned out to be true most of the time. This, of course, is precisely what happened in the case of the chicken-sexer's assertions. As Gasking\(^{31}\) says:

As determinists we are entitled to claim that there is something about the chick that determines his (i.e. the chicken-sexer's) answer, one way or another. But we don't know what it is, and nor does he.

Thus given my analysis of 'ground,' we can say that the chicken-sexer's statements have a ground, even though neither he nor anyone else knows what it is, and yet the chicken-sexer's statements are justified on the basis of the chicken-sexer's past success.

There is an interesting problem that arises here: can there be a statement (p) such that p has no ground, i.e. there is no q that both implies and is evidence for the truth

\(^{31}\)Gasking, p. 159.
of \( p \), and yet there be a person who is justified in asserting that \( p \) holds? My inclination is to deny this possibility, but I shall not argue the point. The consideration relevant for my purposes is: given the above analysis of 'ground' do some avowals have grounds? It seems to me that some avowals do have a ground.

(11) The groundedness of avowals

The ground of the avowal 'I have a toothache' is a state of affairs expressed by a statement that implies and is evidence for the truth of the avowal, viz., 'I am cognitively aware of having a toothache.' In support of this position it is necessary to recall the distinctions made in Chapter Four between the two roles of avowal, on the one hand, and between the two kinds of awareness, on the other hand.

It was there argued that avowals may be used expressively and non-expressively. If there are avowals used purely expressively, i.e. avowals which are made without any cognitive consideration, (I argued that this suggestion is implausible), then, it might be held, such avowals have no grounds. However, even if it were true that there are some avowals used purely expressively, and these avowals have no grounds, it would still be misleading to conclude that avowals in general are ungrounded. It should rather be concluded that considerations about groundedness are irrelevant. For only statements can have grounds, and avowals used purely expressively are not used as statements. So it is not that we thought that purely
expressively used avowals, if there are any, have a ground and that they turned out to be ungrounded. It is rather that the question of groundedness does not even come up.

If, however, the avowal is used non-expressively, i.e. to report, or to describe, or to make a reference, and so on, then it is used to make a statement, and the question of its ground does arise. It is in this second sense, i.e. as used non-expressively, that the avowal 'I have a toothache' should be interpreted.

My contention is that the avowal 'I have a toothache' interpreted as being used to make a statement, has a ground and its ground is the statement 'I am cognitively aware of having a toothache.' It was argued in Chapter Four that non-cognitive and cognitive awareness must be distinguished. A person may be aware of a psychological state without having any cognitive attitudes toward it. Indeed, it may be that a person is always non-cognitively aware of his psychological states, i.e. one cannot have a pain without being non-cognitively aware of it. But a person may be non-cognitively aware of a psychological state without being cognitively aware of it. It is, in my view, cognitive awareness of a psychological state that is the ground of all avowals which are used to make a statement. That is, if the avowal is used to report, describe, refer to, or compare a psychological state, it is presupposed that the avower is cognitively aware of the psychological state. How could he report, describe, refer to, or compare the psychological state if he were not
cognitively aware of it?

To say only that the ground of avowals is cognitive awareness of the psychological state is not to say very much. The assertion is almost vacuous unless it is specified what cognitive awareness is. To say that cognitive awareness is the ground for saying 'I have a toothache' is like saying that the chicken-sexer knows something about the chicks. It is, however, what he knows that is the ground of his statement concerning the chick. We are accustomed to thinking in deterministic terms. The assumption that statements have grounds comes very naturally. But to argue on the basis of this assumption alone against the view that some statements are ungrounded is to argue very weakly. The argument against the ungroundedness of avowals becomes telling only if it is at least indicated what cognitive awareness is; to say only that there is cognitive awareness is not enough. In the next section of this chapter I shall attempt to say what the identity theory takes cognitive awareness to be.

III

(12) What is cognitive awareness?

The identity theory's position is that awareness is identical with the scanning activity of the brain. It was suggested that the scanning activity of the brain is performed by the RAS in transmitting or inhibiting stimuli. An appropriate stimulus transmitted to the cerebral cortex becomes a sensation,
while an inhibited stimulus is not felt at all.\textsuperscript{32} Supposing this scheme is correct, it is still problematic how one would distinguish non-cognitive and cognitive awareness.

In offering two replies to this I make no claim to reliance on facts: I offer speculations. My aim is to give some empirical flesh and blood to the otherwise abstract scheme. It may well be that I am mistaken in offering this empirical content, but this does not matter. The philosophical task, as I see it, is to show that some empirical content can be given. The specification of what it is, is the proper task of the scientist.

Given that awareness and the scanning activity of the brain are identical, the distinction between cognitive and non-cognitive awareness may be drawn in one of the following two ways. If a stimulus is scanned, then the person is aware of it, but whether he is non-cognitively or cognitively aware of it depends on which part of the cerebral cortex is activated by the RAS. If a stimulus is transmitted to certain locations, the person is non-cognitively aware, while if it is transmitted to certain other locations the person is cognitively aware.

A second possibility is that the difference between non-cognitive and cognitive awareness is that while in both cases the stimulus is transmitted to the same location on the cerebral cortex, in the case of cognitive awareness there are

\textsuperscript{32}See Chapter Two, Section (5).
more stages involved in the transmission than there are in the case of non-cognitive awareness. It is the involvement of these additional stages that enable the person to think about, compare, recognize, evaluate, and so on, the sensation he is having. This would enable us also to explain the fact that our response to sensations of which we are non-cognitively aware is more direct and immediate, e.g. wincing caused by a sudden shock of unexpected pain, than the response is to sensations of which we are cognitively aware, e.g. deciding to consult a physician about persistent dizziness. The explanation is that there are a smaller number of stages, and consequently less time, involved in responding to sensations of the first kind than there are in responding to sensations of the latter kind.

There is, however, an objection against the separation of cognitive awareness and its object. The objection is independent of the particular scientific account given of what goes on in the brain when a person is cognitively aware of the having of a sensation. Ryle,33 among others, argues that the separation of cognitive awareness and the having of the sensation leads to an infinite regress. For, it is argued, if one is cognitively aware of some mental state, then it should be possible to be cognitively aware of being cognitively aware of a mental state. So that "there would

have to be an infinite number of onion-skins of consciousness embedding any mental state or process whatsoever." But I do not think that the identity theory involves infinite regress. Armstrong suggests that

there is no logical objection to the introspective awareness of experiences [my 'cognitive awareness'], to the simultaneous introspective awareness of that awareness, and so on as far as we please. This will always involve an ultimate awareness that is not itself an object of awareness. . . . How far such an awareness goes in fact is an empirical question, to which the answer seems to be "Not very far." We can speak of awareness of awareness of awareness of awareness of . . . I, but no psychological reality seems to correspond to our words.

In other words there is a limit to simultaneous acts of cognitive awareness. The limit is the psychological limit of the performance of simultaneous self-scanning processes by the brain. The supposition of an infinite series of simultaneous self-scanning brain processes is logically permissible, but empirically vacuous. There is, after all, a limit to what any physical system can do. The limit of the brain's performance of an additional self-scanning process should be the limit of our postulation of additional sets of cognitive awareness.

The identity theory, however, still needs to explain how, given the correctness of its position, it is possible for a person to be cognitively aware of a psychological state, issue an avowal, and yet not to know what the ground of the avowal is. It was previously noted that from this assertion

34 Armstrong, p. 427.
alone, viz., that the ground of an avowal is cognitive awareness of the psychological state avowed, one cannot infer that the ground of the avowal is known. The assertion in itself amounts to no more than saying that the avowal has a ground; but the avower does not know what the ground is. Instead of talking about cognitive awareness, as I have done, the usual pre-analytical way of describing the situation is by saying that one just knows that he is in a given psychological state. This assertion is analogous to the chicken-sexer's claim that a particular chick is going to be a hen. We may assume that both these assertions have grounds, but unless we specify what their grounds are, we have to admit that the assumption is largely unfounded.

The identity theory, rather than accepting the ungroundedness of avowals, proceeds initially on the basis of this weakly founded assumption. However, if the theory is correct both in its analysis of cognitive and non-cognitive awareness, and in asserting the identity of cognitive awareness and the scanning activity of the brain, then the assumption becomes well founded. The typical avower, however, does not have at his disposal the correct analysis of the state of affairs that he describes as 'just knowing.' Even less is he likely to be able to assert and to justify the identity of cognitive awareness and the scanning activity of the brain. It is not at all surprising, therefore, if all the typical avower can say when asked how he knows that he is in a certain psychological state is that he just knows it, without being
able to explain how he knows it. But from the fact that a
typical avower cannot explain how he comes by his knowledge,
it does not follow that there is no explanation. There is an
explanation, and the identity theory attempts to provide it,
in two steps. The first step is an analysis of 'just knowing'
in terms of cognitive awareness. This analysis is neutral
between different theories of mind. The second step is the
assertion and the justification of the identity of cognitive
awareness and the scanning activity of the brain. This step
is taken only by the identity theorist. Avowals, therefore,
have a ground, viz., cognitive awareness which is identical
with the scanning activity of the brain, while the typical
avower does not know what the ground of his avowal is.

(13) Why is it that avowals rarely need correction?

The identity theory denies that avowals are incorrigible,
and denies also that avowals are ungrounded. The theory
has to provide an explanation of the fact that while avowals
are correctible there is rarely or never any need to correct
them. In other words, one may agree that it is logically
possible that some avowals may be mistaken, and yet hold that
this logical possibility is never actualized. The identity
theory has to account for this immunity—although not a
logically necessary immunity—to being mistaken.

I shall consider and reject a tempting compromise
advocated by Ayer and somewhat hesitantly endorsed by Mackie. Ayer argues that while avowals are not logically incorrigible, all sincere, linguistically competent, serious avowals are empirically incorrigible. The cash-value of 'empirical incorrigibility' is that it is logically possible that an avowal is mistaken, but only the avower, if indeed even he, can actually correct the avowal. Ayer rejects logical incorrigibility, but accepts privileged access. The avower, according to this view, is the ultimate authority on his own psychological states.

An argument supporting the empirical incorrigibility of avowals is the following. The situation required by the logical possibility of a mistaken avowal is that there be a person who is cognitively aware of, say, feeling dizzy, and who is mistaken in asserting this in an avowal; for actually, he is only feeling nauseated. This is a difficult enough position to maintain, but I think the logical possibility of this happening must be admitted. However, it is empirically impossible to correct this mistake once made, for its correction would require: first, that the avower realize that he is feeling nauseated, and second, that he realize that he is mistaken in thinking that he is feeling dizzy. But if anyone realized this, the realization would remove the mistake. The


argument that the mistake the avower makes now may be discovered in retrospect later, does not help. For it is always possible that the retrospective correction is based not on a mistaken avowal, but on a mistaken recollection of the nature of the psychological state avowed, a state which was avowed correctly. So, an avower may be mistaken, but if he is, he cannot detect his mistake. Hence avowals are empirically incorrigible.

Ayer's compromise, however, does not work. For if the identity theory is correct in arguing that avowals are not logically incorrigible—in part because avowals are not ungrounded—then it becomes possible for someone to be an authority concerning the ground of another person's avowals. If the ground of avowals is cognitive awareness of the psychological state avowed, and if cognitive awareness is identical with the scanning activity of the brain, then a millenial neurophysiologist would be in a better position to observe the ground of an avowal than the avower himself. And then neurophysiologists could do what the avower perhaps cannot do, viz., detect the mistake.

The neurophysiologist, however, is millenial and the millenium is not at hand. The identity theory still needs to show how it is that avowals rarely or never are mistaken—even though they are logically corrigible. The question really amounts to asking why, if avowals are like empirical statements, avowals are not as frequently mistaken as are empirical statements. The answer is that avowals are about psychological
states, empirical statements are about the world. Typical avowals are less likely to be mistaken than are typical empirical statements because one's knowledge of one's psychological states is normally more reliable than one's knowledge of the world. This is not surprising; after all, a person is in much more intimate contact with himself than he is with the world. Some of the things that go on inside oneself are far more important, from the agent's point of view, than what goes on in the world. And while the identity theory rejects a radical inner-outer distinction that roughly corresponds to the mental-physical dichotomy, the distinction between the portion of the physical world that lies within my skin and the portion that lies outside of it must be maintained. The varying degrees of interest attached to these two sections of the world could, I suppose, be given an evolutionary explanation, e.g. that one is more interested in what goes on within the skin, because it is more important for survival. At any rate, one comes very close to reaching rock-bottom here. The most natural answer seems to be that this is the way human beings are.

(14) Conclusion

In Chapter Three it was noted that the truth of the correlation statement is a necessary condition of the truth of the corresponding identity statement. The establishment of the correlation statement, however, was found to be dependent on the independent identifiability of both terms of the correlation, viz., the process of having a sensation and BP. It was
argued that BP can be identified independently of sensations through brain-reading, while sensations can be identified independently of BP through behaviour.

The relation between BP and brain-readings was found to be relatively uncomplicated. The real problem was to give an account of the relation between sensation and behaviour. It was argued that the relation is partly contingent and partly non-contingent.

Behaviour is an indicator of the having of sensations, but there are two different ways in which behaviour can be used as an indicator for sensations: certain types of behaviour are logically incorrigible indicators, while certain other types are corrigible indicators of sensations. The typical non-verbal behavioural act is a logically incorrigible indicator, while the typical avowal is a corrigible indicator.

Chapter Four was devoted to showing that avowals are corrigible, i.e. it makes sense to suppose that avowals may be mistaken. It is, of course, rarely or never the case that an avowal is in fact mistaken. The logical possibility of a mistaken avowal, however, is sufficient for the refutation of the incorrigibility thesis as maintained by the expressive theory of avowals.

As a result of those arguments it is possible to maintain that incorrigibility is not a distinctive mark of the mental, i.e. there is no special kind of knowledge involved in knowing one's own psychological states. Also, it is possible to make the identification of sensations through behaviour more
reliable, since the two ways in which sensations are identifiable can be used as tests of each other.

The subject-matter of Chapter Five, the present chapter, is an argument to the effect that avowals are corrigible because they are not ungrounded statements. It is argued that the ground of avowals is cognitive awareness of a psychological state. An identity account of cognitive awareness is proposed.

I consider the claim that sensations and BP could be correlated as partially made out. The question of whether there actually is a perfect correlation between them is an empirical question. My concern has been to remove some of the philosophical obstacles that stand in the way of formulating the correlation statement.
CHAPTER SIX

THE SPATIAL LOCALIZABILITY OF SENSATIONS

I

(1) One of the main objections against the identity theory is that sensations cannot be identical with BP, because BP are, while sensations are not, spatially localizable. The purpose of this chapter is to examine and to reply to this objection.

The problem of whether or not sensations are spatially localizable arises for the identity theory in two different but connected ways. First, the problem arises in connection with the possibility of testing the correlation statement. It may be argued that the obvious way of testing whether two processes are correlatable at all, is to test whether they occur at the same place at the same time. If they do, they may still not be correlatable; but if they do not, then they cannot possibly be correlatable as perfectly as the identity theory requires. The objection against the identity theory is then, that the truth of the correlation statement, required by the corresponding identity statement, cannot be tested, because it is nonsensical to ask for the spatial location
of a sensation.

This objection may be met by pointing out that the correlation between two processes can be tested without reference to their spatial localizability. Let us suppose, for example, that we wish to test whether the passage of time is perfectly correlatable with what time-pieces indicate. It is obvious that we would not go about testing this correlation statement by looking for the spatial location of the allegedly correlatable processes. It may be argued then, in reply to the objection, that sensations and BP may be correlated without either of them needing to be spatially localized.

The objection, however, can be restated. This time it is not directed against the truth of the correlation statement. It is directed, rather, against the truth of the identity statement. It may be argued that the identity is an extensional identity, so that each and every property that BP have, sensations must also have; but BP have the property of being spatially localizable, while sensations do not. Therefore they cannot be identical. It is this second way of interpreting the objection that will be discussed in this chapter.

It will be argued that sensations are processes, so that the problem of their localizability is partly the general problem of the spatial localizability of processes. Most critics of the identity theory, although there are notable exceptions, are unclear in their treatment of sensations, because they are unclear about whether sensations are entities
or processes. If sensations are processes, and if the argument against the identity theory is that sensations as entities have no spatial location, then the objection is an ignoratio elenchi.

Sensations seem to have spatial location. We do say, for instance, that 'I have a pain in my leg.' The appearance, however, may be misleading. If it is, it has to be shown why it is that some sensations, at least, appear to have spatial location when in fact they have none. Several theories will be discussed. Each of them attempts, but as I shall argue, fails to explain what happens when an avowal is uttered that seems to report that the avower has a sensation located at a certain part of his body.

Next, I shall introduce the analysis supporting the identity theory. The analysis will also have to be capable of meeting those objections that proved fatal for previously discussed theories.

(2) Sentences of the following sort will be called 'PL-sentences' (pain-location sentences): 'I have a toothache,' 'There is a pain in my leg,' 'My stomach hurts,' and so on. The problem about the spatial localizability of sensations arises because of the pre-analytic tendency to interpret in apparently conflicting ways what it is that PL-sentences are used to report.

The avowal 'I have a toothache' is, I think, most naturally interpreted as reporting that I have a pain and the place of the pain is in my tooth. This implies that the pain
has a spatial location. But now suppose that the following happens: I go to the dentist who tells me that there is nothing wrong with my tooth, it is an inflammation of the gum that must hurt. After treatment the inflammation is cured and the pain ceases. What happens to the original interpretation of the avowal? On the one hand, it would be unreasonable not to accept the dentist's diagnosis. On the other hand, I may be sure that whatever was wrong with me, it was a 'tooth-achy' sensation that I felt. To reconcile these, I may interpret the avowal 'I have a toothache' as reporting that 'I have a "toothachy" sensation.' On the first interpretation of the avowal, given the correctness of the dentist's diagnosis, I would have to say that I was wrong, that the place of the pain is not in my tooth. On the second interpretation, however, I need not give up my claim, no matter what the dentist says. The reason for this is that the first interpretation implies that pains have a spatial location, while the second interpretation allows for the possibility that pains may not be in space at all.

If one accepts the view that pains are in space, then 'I have a toothache' would have to be regarded as mistaken if it is an inflammation of the gum that is causing the pain. If one allows for the possibility that pains are not in space, then 'I have a toothache' would not have to be regarded as mistaken no matter what it is that hurts.

Thus there appears to be a conflict between ways in which PL-sentences can be interpreted. PL-sentences may be
taken to refer to the place of the pain. But PL-sentences sometimes are not withdrawn even if it clearly demonstrated that a given PL-sentence failed to refer to the place of a pain. The justification for this is that it may be denied that PL-sentences are used to refer to the place of a pain.

The theories I am going to discuss are attempts at coming to terms with the apparently conflicting interpretations of what PL-sentences are used to do.

II

(3) The naive view

PL-sentences are used to assert that:

(A) The place of a pain is the place where the pain is felt to be.

For example, 'I have a pain in my left leg' is to be interpreted, on this view, as asserting that I have a pain and the place of the pain is in my left leg.

The naive view is best regarded as a statement of the problem, rather than a proposed solution to it, for it gives no account of the way in which a pain is in my leg. A pain is not in my leg in the same way as a marble is in a box. If there is a marble in a box, and if the box is in the cupboard, then the marble is in the cupboard. But if there is a pain in my leg, and my leg is in a cast, then it does not follow that the pain is in the cast.

Another consideration pointing at the differences between the spatial location of pains and marbles is that a
marble may or may not be in a container, but it seems that 
pains must be in bodies. The naive view assumes, but does 
not explain what Coburn \(^1\) calls the 'body-linkedness' of 
pains. Nor does the naive view explain how it is that I may 
feel that a pain is at a certain place in my body, but that I 
could be persuaded that the place of the pain is not the place 
where I feel the pain to be.

The naive view is not necessarily mistaken, it is just 
that much more needs to be said before a satisfactory analysis 
of PL-sentences can be arrived at.

(4) The simple behaviourist theory

PL-sentences are used to assert that

(B) The place of a pain is the place toward which pain behav-

For example, 'I have a pain in my left leg' is to be inter-

preted as asserting that I have a pain and the place of the 
pain is the place that I complain about to the doctor, that I 
try to protect, that I try not to put too much weight on, 
that I try to use as infrequently as I can, that I point at 
when asked where it hurts, and so on. \(^2\)

This theory explains the body-linkedness of pain by 
pointing out that as a matter of fact pain behaviour is

\(^1\)R. C. Coburn, 'Pains and Space,' J. of Phil., 63  
(1966), pp. 381-396. Hereafter abbreviated as 'Coburn, 1966.'

\(^2\)This account of PL-sentences seems to be implicit in 
always directed toward some part of one's body, and is never
directed toward anything that is not part of one's body. As
Wittgenstein put it:

"We can't have (haven't as a rule) pains in another
person's tooth." In this proposition the word
"can't" is used in the same way as in the proposi-
tion "An iron nail can't scratch glass." (We could
write this in the form "experience teaches us that
iron nail doesn't scratch glass" thus doing away
with the "can't").

But the theory fails because the analysis cannot be used to
explain PL-sentences about phantom limb sensations. A person
may be disposed to manifest pain behaviour toward his leg.
When, however, it is pointed out to him that unknown to him
his leg has been amputated he will withdraw his claim that
the place of the pain is in his leg.

The disposition to manifest pain behaviour is directed
not toward the place of the pain, but toward the place the
pain is believed to be. This belief, however, may be erroneous,
as the phantom limb case shows. If the belief is erroneous at
least sometimes, then the place of the pain cannot be the
place toward which pain behaviour is directed; for these may
be two different places.

(5) The causal theory

PL-sentences are used to assert that

---

3L. Wittgenstein, The Blue and Brown Books, (New York:

4K. Baier, 'The Place of a Pain,' Phil. Quart., 14
(1964), pp. 138-150. Hereafter abbreviated as 'Baier.'
(C) The place of a pain is the place where the cause of the pain is.

As it stands this formulation is inadequate because it fails to recognize the body-linkedness of pains. It would be possible to say, according to (C) that the place of my pain is in the stove, if it was the stove that caused my pain. To meet this difficulty (C) can be reformulated as

(D) The place of the pain is the place where the first bodily link is in the causal chain producing the pain.

For example, 'I have a pain in my left leg' is to be interpreted as asserting that I have a pain and the place of the pain is the place where there is an organic damage causing me to feel pain in my left leg. Or as Ryle puts it:

The point here being made is that whether we are attaching a sensation to a physiological condition or attaching a feeling to an emotional condition, we are applying a causal hypothesis. Pains do not come already hall-marked 'rheumatic.'

The theory is capable of handling the phantom limb case by arguing that the place of the pain is not in the amputated leg, but rather it is the place of the cause of the pain, e.g. the stump, or an inflamed nerve-ending.

The theory has the further feature, considered to be an advantage by some, that it leaves it an open question whether or not pains per se have spatial location. For one

---


may hold in accordance with the causal theory that while the cause of the pain has a place, the pain itself has none. The causal theory is thus particularly suited for some dualistic theories.

The theory fails, however, because it cannot account for PL-sentences about referred pains. In some cases a person may feel pain at one place, e.g. in his left leg, while the cause of the pain is at another place, e.g. in his spine. If this is pointed out to him, the person will not conclude that he was wrong in locating the pain in his leg, although he may well accept that the injury or abnormality causing the pain is in his spine. He will not come to complain about his spine rather than about his leg; he will not behave as though it was his spine and not his leg that hurt. So, relying on the person's verbal and non-verbal behaviour, the view that the place of a pain is the place where the first bodily link in the causal chain producing the pain is has to be rejected.

An interesting question that arises in this connection is: why is it that in the phantom limb case a person would change his mind about the place of his pain, while in the referred pain case he would not? The explanation, I think, is to be found in the body-linkedness of pain. It is a necessary condition for determining the place of a pain that the place be somewhere in the body. But it is not a necessary condition for determining the place of pain that it should

7 Baier, pp. 140-141.
coincide with the place of the cause of the pain.

(6) The local sign theory

PL-sentences are used to assert that

(E) The place of a pain is the place where playing the 'wound-locating game' and going by the local sign we judge that the first bodily link in the causal chain producing the pain is located.

Baier\(^8\) elucidates the 'wound-locating game' in the following way:

Poking around to discover which pokes hurt the patient most enables the doctor to say where the injury is, namely, in the place of the poke which hurts most. Correlation between the place of the poke and the intensity of the pain helps locate the internal injury not otherwise easily locatable. We ourselves may get adept at locating our internal injuries in this way. Next, we may learn to locate, from the feel alone (i.e. from what psychologists have called the "local sign"), where our injuries are . . . at this stage, we are still playing a game locating "public" objects . . . the location claims . . . are subject to correction by the five senses.

The role of the local sign, according to Coburn,\(^9\) is that of

(a) indicating more or less straightforwardly that we have a pain of a certain kind, viz., a pain of a kind which we've learned through experience to be associated with something unusual (and generally untoward) happening in our left leg, and which, accordingly, constitutes a sign that something unusual (and probably untoward) is happening in this leg; and a fortiori (b) indicating more or less obliquely the place of those events or conditions which are the cause or causes of the pain we feel.

There are, however, two 'pain-locating games' compatible with (E). The first is pain-locating through wound-locating, i.e.

---

\(^8\)Baier, p. 142.

going by pokes; the second is pain-locating through local sign, i.e. going by the 'feel' of the pain. It is possible that the two pain-locating games produce different results.

Cases where the two distinct types of pain-locating attempts yield the same result are unproblematic. For example, 'I have a pain in my left leg' could be such an unproblematic case if, on the one hand, I have a 'leg-achy' feeling, and on the other hand, going by pokes, I locate the injury in my left leg. Pain-locating through wound-locating, and pain-locating through local signs, in this case yield the same conclusion.

But not all cases are unproblematic. In the case of referred pains, for instance, the two distinct types of pain-locating attempts yield different results. If the pain I have in my leg is a referred pain, then going by the 'leg-achy' feeling I would say that the place of the pain was my left leg. But going by pokes, I would conclude that the place of the pain was in my spine, because the pain that I felt in my leg increased and decreased as I poked at my spine, while the pain in no way altered when I poked at my left leg.

The causal theory failed because it could not account for the sort of problematic cases mentioned above. The present theory, however, can account for such cases by recognizing that P1-sentences play a dual role. On the one hand, P1-sentences may be used to assert that a person has a pain of a certain sort, or of a certain quality, e.g. a 'leg-achy'
sensation. On the other hand, PL-sentences may be used to assert that a person has a pain and the place of the first bodily link in the causal chain producing the pain is in his left leg.

PL-sentences used in the first sense may be based on the way in which a person has learned to describe his sensations. A pain in the leg is described as a 'leg-achy' sensation, because we learn to describe pains with reference to the place where the injury is. Normally a pain is felt at the place where the injury is. There are, however, exceptions: referred pains, for example. If a PL-sentence about a referred pain is interpreted in the first sense, viz., as asserting that the person has a 'leg-achy' sensation, then its truth value is unaffected by the fact that the injury may not be located where the pain is felt to be. A PL-sentence, used in the first sense, asserts nothing about the location of the pain, or the location of the injury causing the pain.

PL-sentences used in the second sense are inferences based on the fact that normally when a person feels a 'leg-achy' sensation there is an injury in his leg. If a PL-sentence about a referred pain is interpreted in this second sense, then when a person discovers that his inference was mistaken he comes to realize that the PL-sentence, used in the second sense, is false, i.e. attempts but fails to refer to the place of the cause of the pain.

Thus, according to the local sign theory, PL-sentences play a dual role. In the case of referred pain, a PL-sentence,
used in the first sense, is still a perfectly acceptable report of what kind of pain the person feels. While a PL-sentence, used in the second sense, would have to be recognized as asserting something false, if it was used to assert that the injury causing a referred pain was in the leg when in fact the injury was in the spine. The basic assumption of the local sign theory is that PL-sentences used in the first and in the second sense do not conflict, because they are used for different purposes.

The local sign theory is capable of handling objections based on the body-linkedness of pains, on phantom limb sensations, and on referred pains. The theory, however, fails to account for what Baier called the 'facts of the pin-prick.' The facts of the pin-prick show that PL-sentences may conflict in their two uses, and so the solution provided by the local sign theory does not work in all cases. It will be remembered that in certain problematic cases, e.g. referred pain, the two pain-locating games yielded different conclusions. On the basis of the 'feel' of the pain the pain was located at one place, while on the basis of poking the pain was located at another place. The local sign theory resolved this problem by arguing that PL-sentences based on the 'feel' of the pain do not refer to the spatial location of the pain at all, hence the apparent conflict between the two uses of PL-sentences is not a real one.

---

10Baier, pp. 142-143.
The implication of this solution is that it can never be the case that a PL-sentence used in the first sense is withdrawn or altered if it apparently conflicts with a PL-sentence used in the second sense. The facts of the pin-prick, however, show that this is exactly what occasionally has to be done. Imagine a situation in which a person is pricked by a pin, say, on his back. He is not allowed to explore the area by touch, he is forbidden to look, or to play the wound-locating game in any way. He is asked to locate the place of the injury caused by the pin-prick. Now it could easily happen that going by the 'feel' of the pain alone the person mislocates the place of the injury. His mistake need not be drastic; he may be only a few inches off the place of the pin-prick. If, however, the person is subsequently allowed to play the wound-locating game, then he may well come to change his mind. He may then withdraw the first PL-sentence, viz., the one that was based on the 'feel' of the pain only, and propose a second PL-sentence, in place of the first, based on having played the wound-locating game. But if the local sign theory were correct, this could not have happened, since PL-sentences based on the 'feel' of the pain are not supposed to be used to locate pains, so they cannot be used to mis-locate pains either. The facts of the pin-prick show, however, that this is not the case.

(7) The modified local sign theory

PL-sentences are used to assert that

(F) The place of a pain is the place where playing the wound-locating game and going by the local sign we judge that
the first bodily link in the causal chain producing the pain is located relative to the location of other pains and possible other sensations.

This view is advanced by Taylor\textsuperscript{11} in order to meet Baier's objection against the local sign view. Taylor\textsuperscript{12} writes:

The general interest in the analysis of pain location statements lies in attempting to remove the ontologically disturbing idea of a sensation, apparently something non-physical, being at a particular place. Two very general ways of doing this are on the one hand to analyze the concept of pain at a place into two elements, a physical event felt to be going on at a place, and some attitude to this event, and on the other hand, to reduce the place of a pain to the place of the cause of a pain.

The local sign theory is an attempt at doing the former; the modified local sign theory is Taylor's attempt at doing the latter.\textsuperscript{13}

Taylor's theory differs from the local sign theory, because of the denial implicit in it that PL-sentences play a dual role. It seems that Taylor acknowledges that PL-sentences can be based either on local signs or on playing the wound-locating game, but whatever PL-sentences are based on, they are used to assert that a person has a pain and the pain is located at a certain place. The disagreement between the local sign theory and Taylor's modification of it is that

\textsuperscript{11}D. M. Taylor, 'The Location of Pain,' Phil. Quart., 15 (1965), pp. 53-62. Hereafter abbreviated as 'Taylor.'

\textsuperscript{12}Taylor, p. 53.

\textsuperscript{13}I think that Armstrong's account of the location of bodily sensations is substantially similar. See D. M. Armstrong, Bodily Sensations, Studies in Philosophical Psychology, (London: Routledge, 1962).
on the former view some PL-sentences are not used to refer to the place of a pain, while on the latter view all PL-sentences are, at least attempted to be, used to refer to the place of a pain.

The difficulty for Taylor's theory is to give an account of PL-sentences in terms of the place of the cause of the pain and avoid objections raised against the causal theory as well as the local sign theory. Taylor's way of overcoming the difficulty is to attempt to combine the best features of these two theories.

In accordance with (F), 'I have a pain in my left leg' is interpreted as asserting that I have a pain and the place of the pain is the cause of the pain as ascertained by (i) the local sign, i.e. by having a 'leg-achy' sensation, (ii) playing the wound-locating game, and (iii) locating the place of the cause of the pain in relation to other pains and other sensations, e.g. the place of the wound is my left leg, because the pain I feel is very near to the place where I feel the itch in my left leg.

Taylor meets the pin-prick objection by agreeing that a person may be mistaken about the place of the pin-prick if he relies on local sign alone. But if he relies on other evidence as well, viz., the wound-locating game and the place of other sensations as compared to the place of the pin-prick, then the chances are that he will not be mistaken. Taylor, it seems to me, acknowledges that any PL-sentence may be mistaken in attempting but failing to refer to the place of the cause.
of the pain. But this, he seems to hold, is no cause for alarm; it merely shows that the pain-locating game could be played better.

Taylor's view can also deal with the referred pains whose existence refuted the causal theory. For if one plays the wound-locating game and locates the pain in relation to the place of other sensations, then referred pain in the leg will be located in the spine. For it is by poking around the spine that the pain in the leg increases or decreases; the referred pain remains unaltered if one pokes around the leg, where the pain is supposed to be.

There is, however, a question that generates a dilemma for the modified local sign theory. The question is this: is a pain identical with the physiological and/or behavioural events that are normally thought of as causing the pain? If the identity is denied, then the theory has not accomplished its purpose. For it set out to analyse PL-sentences which were taken to refer to the place of a pain. It was concluded that PL-sentences refer to the place of the cause of the pain. But if a pain is not identical with what is normally thought of as its cause, then the question of where pains are remains. It is the spatial location of pains that is a puzzling problem; the spatial location of the cause of the pain has little theoretical import. Taylor may have solved the latter, but the former still stands.

If, however, Taylor opts for the view that a pain is identical with the physiological and/or behavioural events
that are normally thought of as the causes of pain, so that the place of the pain is the place of the cause of the pain, then another difficulty arises. It seems to be one thing to claim to have a pain at a certain place in the body, e.g. in the leg, and another to have a pain of a certain quality, e.g. a 'leg-achy' sensation. These seem to be different, for in certain circumstances, e.g. if it is a phantom limb sensation, it is reasonable to change one's mind about the place of the pain, while it seems that under no circumstances would one change one's mind about the quality of the pain felt. But if a pain were identical with what is normally thought of as its cause, then the feeling that the cause of the pain is in my left leg, and the feeling that I have a 'leg-achy' sensation should always coincide. For if pains were identical with what are normally regarded as their causes, then the only way a 'leg-achy' sensation could occur would be to have something wrong with one's leg. As the cases of phantom limb, referred pain, and pin-prick show, however, this is not the case.

Taylor's modified local sign theory does not meet this objection. I do not think, however, that the objection is insurmountable. The theory I shall develop is not altogether unlike Taylor's; I shall meet this objection in the course of developing the identity theory's account of PL-sentences.

III

(8) The identity theory's account of PL-sentences

In giving an analysis of PL-sentences the identity
theory has to do two things. First, it must reconcile the conflicting tendencies in the interpretation of PL-sentences. There is the tendency, on the one hand, to interpret PL-sentences as referring to the place of a pain. The opposing tendency, on the other hand, is not to recognize the falsity of some PL-sentences if it is shown that the place of the pain cannot be the place to which the PL-sentences seem to have been used to refer. The first tendency implies that pains are in space, the second tendency implies that pains are either non-spatial or that spatial considerations are irrelevant to them. The identity theory can be successful, in this respect, only if it can show that the second tendency does not stand in the way of the spatial localization of pains.

The second task of the identity theory is to give such an account of the spatial localization of pains as to allow for the identity of sensations and BP. The conditions of theoretical identity require not only that sensations should be spatially localizable, but also that the spatial location of a sensation should be precisely the same as the spatial location of those BP with which the sensation is alleged to be theoretically identical. So the identity theory would be refuted if it were found that the place of the pain in the leg is the leg, just as much as if it were found that pains have no place at all.

(9) Critics of the identity theory, however, hold that it makes no sense to talk about the spatial localizability of sensations.
One of the main objections against the identity theory is based on just this assumption. Shaffer\textsuperscript{14} writes:

The time has come to attack the theory directly, by bringing up the major difficulty in any theory of this sort, the location problem. \ldots\ Since brain events occur somewhere in the brain but mental events do not, it follows that they cannot be identical.

Or consider Malcolm's\textsuperscript{15} criticism:

It is clear that a brain process has spatial location. A brain process would be a mechanical, chemical or electrical process in the brain substance, or an electrical discharge from the brain mass, or something of the sort. \ldots\ I think that in our ordinary use of the terms "thoughts" and "thinking," we attach no meaning to the notion of determining the bodily location of a thought. We do not seriously debate whether someone's sudden thought occurred in his heart, or his throat, or his brain. Indeed, we should not know what the question meant.

Or as Kim\textsuperscript{16} states the objection:

The so-called location problem for mental states and events has perhaps been the strongest obstacle to the Identity Theory; the alleged difficulty raised by it seems to have persuaded more philosophers against the theory than any other single difficulty. As formulated by the critics of the theory, the objection runs as follows. If a mental state is to be identical with a physical state, the two must share all properties in common. But here is one property, spatial localizability, that is not so shared; that is physical states and events are located in space, whereas mental states and events are not. Hence, mental events and states are different from physical ones.

\textsuperscript{14} J. A. Shaffer, 'Recent Work on the Mind-Body Problem,' \textit{Am. Phil. Quart.}, 2 (1965), pp. 81-104. Hereafter abbreviated as 'Shaffer, APQ.'


\textsuperscript{16} Kim, p. 233.
It is a noteworthy feature of these criticisms that they assume that physical states, events and processes have spatial location; and that it is equally clear that mental states, processes, and events do not. I question both assumptions. It is, I think, none too clear that physical processes like maturation or dying, and physical states like weighing 150 pounds or mental fatigue, and physical events like scoring a goal or increasing velocity, have spatial location, and if they have spatial location just what exactly it is. Nor is it clear to me that it is absurd to say that the sense in which some physical processes, states, or events have spatial location could not be the sense in which some mental processes, states, or events have spatial location. I do not think that there is more prima facie absurdity in supposing that the mental state of having a pain in my leg has spatial location than there is in supposing that the physical state of having a hand has spatial location.

(10) One might attempt to dispose of the spatial localizability objection against the identity theory the way Smart\(^\text{17}\) tried to do it. He writes:

The experience, it will be said, is not in physical space, whereas the brain process is. Hence the experience is not a brain process. This objection seems to beg the question. If my view is correct the experience is in physical space: in my brain.

But if the spatial localizability objection against the identity theory is question-begging, so is Smart's way of meeting

\(^{17}\)Smart, PSR.
it. The problem after all is: do sensations have spatial location? Dogmatic assertions, affirmative or negative, do not solve the problem.

The point supporting Smart's argument is that if the identity theory were shown to be correct, and the only obstacle to its general acceptance were the fact that ordinary language makes us reluctant to ascribe spatial location to sensations, then perhaps it would be justifiable to adjust ordinary language so that it would accommodate theoretical advances.

The point which goes against Smart's way of meeting the objection is this: since the identity theory is very far from being free of all objections other than the spatial localizability one, it is necessary to examine each objection in turn. If one did not do this, and followed Smart's procedure instead, one might well reply to all objections against the identity theory by saying that if the theory were correct, then the objection would not apply. But in doing this one would assume the truth of what is at issue, viz., the correctness of the theory. It is necessary, therefore, to argue for the theory by meeting each and every objection against it.

The way in which I shall attempt to meet the spatial localizability objection is by showing that it does make sense to ascribe spatial location to sensations. I shall develop the identity theory's account of PL-sentences in response to what is perhaps the most detailed criticism of it.
Shaffer argues that mental states (Shaffer's "C-states") cannot be identical with brain processes (Shaffer's "B-processes") because they do not occur at the same place. Shaffer notes that one of the requirements of the identity of C-states and B-processes is that "both must be located in the same place." Then he goes on to ask:

Is the spatial requirement met by C-states and B-processes? Do they occur in the same place? No. B-processes are, in a perfectly clear sense, located where the brain is, in a particular region of physical space. But it is not true that C-states occur in the brain, or inside the body at all, for that matter. To be sure, I may have a pain in my leg or in my head; we do locate sensations in the body. But that is not to say that we give location to the state of consciousness that I have when I am having a sensation. The pain is in my leg, but it is not the case that my state of being-aware-of-a-pain-in-my-leg is also in my leg. Neither is it in my head. . . . In fact, it makes no sense at all to talk about C-states as being located somewhere in the body. . . . The fact that it makes no sense to speak of C-states occurring in a volume occupied by a brain means that the Identity Theory cannot be correct.

Shaffer then goes on to take up a suggestion offered by Smart to the effect that we could adopt a convention "whereby it would make sense to talk of an experience in terms of appropriate physical processes." Shaffer explains what such a convention might be:

---

18J. A. Shaffer, 'Could Mental States Be Brain Processes?', J. of Phil., 58 (1961), pp. 813-322. Hereafter abbreviated as 'Shaffer, JP.'

19Shaffer, JP., p. 815.

20Smart, SEP.

21Shaffer, JP., p. 816.
A convention we might adopt would be something like this: for any C-state, if it has a corresponding B-process, it will be said to be located in that place where its corresponding B-process is located. Given this convention, it then becomes a matter for empirical investigation whether any C-state has location in space and where that location, if any, is. The outcome of such an investigation could be settled, at least in principle, with as much exactness as we like.

I shall consider three different objections against this proposal. The first one is given by Shaffer himself:

The question whether the conceptual revision proposed by Identity theorists should be accepted or not cannot be determined solely by philosophers. It is in part an empirical question, to be judged in terms of future discoveries in neurology and psychology. . . . In the case of the Identity theory, the linguistic innovation consists in modifying our concept of C-states by giving criteria for the spatial location of C-states. Only future discoveries in neuro-physiology can tell how fruitful this innovation might be.

As things are now, Shaffer argues, C-states and B-processes, cannot be identical because one of the conditions of identity, viz., spatial localizability, is lacking. So the identity theory has to await empirical evidence which may or may not be forthcoming. If evidence is forthcoming, then, and only then, it may be useful to institute conceptual revision. But until then the identity theory has not satisfied at least one of the conditions of the identity of C-states and B-processes.

A second objection against the conceptual revision proposed by Smart and elaborated by Shaffer is advanced by

Shaffer, JP., pp. 821-822.
Let us suppose that certain neural events are discovered to be (causally) necessary and sufficient conditions for the occurrence of such and such a type of pain experience. And let us further suppose that the convention is adopted of "locating" such experiences at the point(s) in the brain where these neural events take place. Now let us ask what consequences such a convention might carry. One such consequence would be, I submit, that pain experiences of the type in question would be rendered "public" entities in the sense that no person would be in any better position essentially than any other for determining with certainty whether such an experience was occurring. . . . But if "publicizing" all mental states in the sense indicated is a consequence of adopting spatial conventions for the expressions that refer to such states, surely it is false that adopting such conventions is consistent with the rules which govern such expressions under current semantic conditions, and a fortiori false that the sort of conceptual changes Shaffer insists are necessary to keep the Identity theory from lapsing into a priori falsehood would leave our mental concepts intact. For surely it is an essential feature of (e.g.) the language of pain that pain experiences are not "public" in the above sense.

In other words, the adoption of the conceptual revision would produce much more far reaching changes than Smart envisaged. The acceptance of these changes merely on the ground that the identity theory needs them would amount to making the theory true by definition. And this is clearly not Smart's intent.

The third objection against the proposed conceptual revision comes from Malcolm.  

23 R. C. Coburn, 'Shaffer on the Identity of Mental States and Brain Processes,' J. of Phil., 60 (1963), pp. 89-92. Hereafter abbreviated as 'Coburn, 1963.'

24 Malcolm, SMIT, p. 119.
it makes no sense to ascribe spatial location to mental states, then he writes:

> It might be replied that as things are the bodily location of thoughts is not meaningful notion; but if massive correlations were discovered between thoughts and brain processes, then we might begin to locate thoughts in the head. To this I must answer that our philosophical problem is about how things are. It is a question about our present concepts of thinking and thought, not about some conjectural future concepts.

All these objections tell, I think, against Smart's proposed conceptual revision. I think, however, that one does not need to bring about a conceptual revolution in order to show that one can talk meaningfully about the spatial location of sensations.

(12) Let us return to Shaffer's original objection, viz., that it makes no sense to say that sensations have spatial location, while it makes sense to say that brain processes have spatial location, and hence that they cannot be identical.

The first step toward the formulation of a satisfactory account of the spatial location of sensations is to recognize that it is not the spatial location of an entity, but rather that of a process that we have to establish. This is the force of Smart's remarks:

> If we take the example of the after-image it will be objected that the after-image is patently not in physical space... but in some sort of psychological space. This objection falls down because I am not arguing that the after-image is a brain process. On my view there are no such entities as after-images,

---

25 Smart, PSR., p. 97.
but there are processes of having-an-after-image ... I am arguing that the experience of having an after-image is in fact a brain process, and hence is in physical space.

Thus, if sensations are in space, they are in space as physical processes like maturation or dying are in space, and not as physical entities like marbles are in space.

Shaffer\(^{26}\) is willing to recognize this point, but he argues that it still makes no sense to ascribe spatial location to sensations—not even if they are regarded as processes:

Such things as temperature, potential energy, solubility, electrical resistance, and viscosity do not have a location, yet they may still be identical with micro-states of a physical object. Since we are considering mental events, let us consider comparable events, changes in temperature, potential energy, solubility, etc. Surely we can ask where in the object these changes occurred. Certainly if we identify temperature with some property of molecules, perhaps their average kinetic energy, then the change in temperature must occur where the change in the molecules occur. If we are not willing to locate the change in temperature there, we must say that it is not identical with the change that does occur there. To take another example, when a person goes into a shock, we do not locate the shock in the vascular system (nor anywhere else in the body). That shows it is wrong to identify shock with, say, a collapse in blood pressure; the latter is, at best, the cause of the shock.

In other words, if a change in temperature is identical with a change in the mean kinetic energy of molecules in a physical system, then it must make sense to say that the change in temperature is located at the same place as the change in the mean kinetic energy of the molecules in a physical system is

\(^{26}\)Shaffer, APQ., p. 97.
located. Shaffer points out, and I am in complete agreement with him, that the change in temperature is located where the change in the mean kinetic energy of the molecules in a physical system is located. The question is: where is that located?

It seems to me that the only sensible answer is that since the molecules are located all over the system, indeed they are the system, so the change in the mean kinetic energy of the molecules is located all over the system. That is, the spatial location of the change in temperature is the spatial location of the system. It is a mistake to ask, therefore, where in the system the change in temperature is located, since it is not located in any one part of the system, but rather in the system as a whole.

Now Shaffer, after having postulated an analogy between change in temperature and mental events and processes, goes on to demand an answer about the spatial location of mental processes, an answer that would be a mistake to demand about the spatial location of physical processes. Shaffer asks what part of the body houses the state of shock, and when no answer is forthcoming, he takes this as showing that shock has no spatial location. But if the state of shock has none, neither has the change in temperature. For it is equally non-sensical to ask: where in the system is the change in temperature located? If, however, Shaffer is prepared to say that the spatial location of the change in temperature is the same as the spatial location of the system as a whole, then he should
be prepared to say also that shock is located where the body that is in shock is located.

(13) From the discussion of Shaffer's objection we can derive the way in which the identity theory's account of the spatial location of sensations should be further developed. Let us take the argument step by step.

The identity theory asserts that sensations are identical with BP. It is objected that sensations are not, while BP are, spatially localizable, and that consequently they cannot be identical. The reply to this objection is that it is not sensations-as-entities, but rather sensations-as-processes whose spatial localizability is in question. The objection, however, is merely reiterated: it makes no sense to talk about the spatial localizability of sensations-as-processes, i.e. the having of sensations, either.

Kim27 replies to this objection by arguing that having a hand, weighing 145 pounds, having a temperature of 97 degrees, and other so-called physical states and events have no clear spatial location either. A hand can be located in space, but having a hand cannot; a brain may be located in space, but not a brain state; my body can be located in space, but not my body's weighing 145 pounds. Thus the inconclusiveness and weakness of this objection to the Identity Theory stems not so much from the possible localizability of mental states and events as from the vagueness of the general concept of spatial location for events and states.

The objection now becomes that while it is true that the concept of spatial location of events and states is unclear,

27Kim, pp. 233-234.
mental events and states still cannot be located, while physical states and events can be. The validity of this objection, however, depends upon the outcome of the analysis of the concept of spatial location of events and states in general.

Kim offers the following analysis:

We locate events by locating the particulars that "undergo" them. Something explodes in an explosion, and the explosion is located where the thing that explodes is located; when there is fire, something burns, and the fire is where the burning thing is; and similarly, a death takes place where the dying man is located. Particulars are located first; events and states are located relatively to particulars. Or, we may say, particulars are the primary localizable entities; events and states are localizable only derivatively.

Thus, according to this analysis of the spatial location of sensations, i.e. as events or processes, they are localizable derivatively, i.e. with reference to the particulars which undergo them. Consequently, to assert that sensations are spatially localizable is to assert that the particular that undergoes them is spatially localizable, and to deny that sensations are spatially localizable is to deny that the particular in question is spatially localizable. It seems, on the face of it, that there are only two candidates that can be offered here. However, if either of them is accepted, the identity theory cannot be held in the form in which I want to hold it. I shall argue therefore that there is a third candidate.

---

28 Kim, p. 234.
Kim goes on to discuss the first candidate:

In order to show that Socrates' being in pain is not spatially localizable it must be shown that, on this construal of the location of an event, Socrates to whom the property of being in pain is attributed is not a spatially localizable entity. But in order to show this one must show or assume that the subjects of mental properties --or the subjects of mental events and states-- are immaterial souls or mental substances in the full-fledged Cartesian sense.

If this alternative were accepted, then, of course, the identity theory would not be correct.

Kim then concludes his argument:

The situation, therefore, seems to be this. Insofar as the notion of the location of an event is unclear and vague, it is not clear that all physical events and states have location; and insofar as it is made clear--in terms of the location of particulars--the assertion that mental events and states lack spatial location implies the Cartesian thesis of the immaterial soul and unextended mental substance. Hence the objection based on the location problem is unclear and therefore inconclusive, or it begs the question at issue.

It would be tempting to stop here with Kim, and conclude that the identity theory has been defended against the spatial localizability objection. I think, however, that one can do more, and perhaps more needs to be done. For if spatial localizability of sensations and BP is a necessary condition of their identity, then the argument that the concept of spatial location of events or processes is unclear hurts the identity theory as much as it hurts the objection against it.

---

29 Kim, p. 234.
30 Kim, p. 234.
So it seems to be necessary to proceed to show how the particular of which sensations are predicated can be said to be spatially localizable. I turn thus to another candidate proposed by Nagel in reply to Malcolm's criticisms. Nagel suggests that

if the two sides of the identity are not sensations and brain processes, but my having a certain sensation or thought and my body being in a certain physical state, then they will both be going on in the same place—namely, wherever I (and my body) happen to be. It is important that the physical side of the identity is not a brain process, but rather my body's being in that state which may be specified as "having the relevant process going on in the brain." That state is not located in the brain; it has been located as precisely as it can be when we have been told the precise location of that of which it is a state—namely, my body. The same is true of my having a sensation: that is going on whenever I happen to be at the time, and its location cannot be specified more precisely than mine can.

Thus, according to Nagel, the particular with reference to which sensations are spatially localizable is the body or the person.

If Nagel's suggestion were accepted, then the objection that sensations do not, while brain processes do, have spatial location, would be met. But the acceptance of the suggestion would amount to a denial that sensations are identical with B2. Sensations, on this analysis, would be states or processes of the body or of the person, not of the brain. As Nagel puts it:

---

32Nagel, p. 341.
Instead of identifying thoughts, sensations, after images, and so forth with brain processes, I propose to identify a person's having the sensation with his body's being in a physical state or undergoing a physical process. The subject are the person and his body (not his brain).

In order to maintain the version of the identity theory I hold, it has to be shown that the having of sensations is spatially localizable with reference to the brain rather than the body. It is to this task that I shall now turn.

(14) It should be noted that the claim that sensations are spatially localizable with reference to the body, and the claim that sensations are spatially localizable with reference to the brain are not incompatible: the brain is part of the body. The difference between the two claims is that the latter is much more specific than the former. The problem is to justify the more specific claim.

The justification, briefly, is this: it is a necessary condition for adopting a particular with reference to which processes are spatially localized, that there must be something going on in the particular such that it goes on if and only if the process in question is occurring. For example, dying is correctly localized with reference to an organism if and only if irreversible cellular damage of a certain magnitude is going on in the particular. Or singing is correctly localized with reference to a person if and only if his vocal chords vibrate in a certain way. My contention is that the only thing that satisfies this requirement in the case of
It is important to realize that I am not begging the question. My contention is that sensations are spatially localizable with reference to the brain, not because sensations are identical with BP but because BP, and only BP, satisfy the requirement of occurring if and only if sensations occur. Sensations may or may not be identical with BP, just as dying or singing may or may not be identical with irreversible cellular damages or with certain ways in which the vocal chords vibrate.

(15) The identity theory offers the following analysis of the spatial location of sensations. The place of a pain is established derivatively. It is the place of the particular, i.e. the brain, of which the process of having a sensation is predicated. On this view, PL-sentences are used to assert that the person is aware of something untoward going on at a certain place in his body; what the person is aware of is a stimulus that he interprets, and possibly misinterprets, as indicating the place of the untoward happening.

The place of the pain thus is not the place where the pain is felt to be; nor is it the place where the stimulus is correctly or incorrectly located; nor is it the place where the cause of the pain is; nor is it the place where, going by the local sign and by playing the wound-locating game, the

\[33\] I have argued for this in Chapter Three, Section (7).
pain is judged to be. The place of the pain is the place of the particular that undergoes the process of having a pain. This particular is neither the dualistic mind, nor is it the body. It is rather the brain. So the place of a pain, in short, is the brain.

Of course, the brain is in the body, and the body is, or is part of a person; so, in a sense, the place of a pain is the place of the body or of the person. But the sense in which this is true is derivative. The reason why the place of a pain, in this sense, is the body or the person is that the brain is part of the body or of the person.

When a person says 'I have a pain in my leg' what is happening is that a stimulus, of the sort that usually indicates that something untoward is going on in the leg, is scanned by the brain. The presence of the scanned stimulus is at least a necessary condition for the person's having a sensation. The identity theory holds that a scanned stimulus of a certain sort is identical with the having of a sensation. This radical claim, however, need not be correct for my argument in this chapter to hold. It is enough for my purposes here that the scanned stimulus be a necessary condition of the having of a sensation.

Given this analysis of the place of a pain, the question arises of how a person knows that when he is aware of a particular stimulus, it is his leg that hurts. In other words, if PL-sentences do not indicate in any straightforward sense the place of pain, how is it that the person knows what part of
his body is injured? In answer to this I shall appeal to the modified local sign theory. I disagree with that theory's analysis of PL-sentences, but I agree with its account of the wound-locating game. That is, I think that the modified local sign theory gives a mistaken account of the place of a pain, but that its account of how we learn to identify the place of an injury is correct, viz., that a person learns to identify the place of an injury by playing the wound-locating game.

By playing the wound-locating game a person may learn to identify the place of the injury. The place of the injury, however, need not be the place of the pain. The place of the injury is the place at which the stimulus (eventually scanned by the brain) originates. But conclusions arrived at on the basis of the wound-locating game are not necessarily correct, as the cases of the phantom limb, referred pain and pin-pricks demonstrate. In other words, according to the theory I hold, PL-sentences are corrigible. That is, when a person says 'I have a pain in my leg' he may be mistaken in the sense that while he has a 'leg-achy' sensation, it is perhaps not his leg that hurts, or it may be that it is his leg that hurts but the injury is elsewhere, or perhaps he has no leg at all. If the place of the pain is the brain, then while the presence of a scanned stimulus is at least a necessary condition of the having of a pain, as far as the location of the pain is concerned, it makes no difference whether or not the stimulus is correctly located. The reason for this is that my analysis
of the place of a pain in no way depends on the place at which the stimulus originates. Thus the identity theory's account of the spatial location of sensations is not open to any of the objections that proved fatal to formulations (A)-(E).

Nor does the objection against (F), i.e. Taylor's theory, refute the present analysis. The objection was that Taylor seemed to be forced to accept that there is a difference between having a pain in the leg and having a 'leg-achy' sensation. The difference is that under certain circumstances I may change my mind about having a pain in the leg, e.g. if I am told that unknown to me my leg has been amputated, but that I need never change my mind about having a 'leg-achy' sensation. If Taylor's account were correct, i.e. if a pain were identical with what is ordinarily thought of as its bodily cause, then there should be no difference between the place of what is ordinarily regarded as the cause of the pain and the place where the pain is felt to be.

My way of meeting this objection is to point out that the sentence 'I have a pain in my leg' can indeed be interpreted at least in two different ways. First, it may be taken to assert that 'I have a pain, and the place where the cause of the pain is is in my leg.' This is a corrigible statement; one may be wrong about the place of the cause of one's pain. Second, it may be taken to assert that 'I have a "leg-achy" sensation, regardless of where its cause is.' Whether this is corrigible or not depends on which account of avowals is correct. I argued that avowals are corrigible, but my reply here
does not depend on the correctness of my view on avowals. My reply is based on the need for recognizing that these two interpretations do not conflict. The first states where a person believes the pain to have originated; the second states what sort of pain it is taken to be.

The theory I support, unlike Taylor's, is capable of handling this, because it does not analyse the place of a pain (of a certain sort) in terms of the cause of the pain. Of course, I agree with Taylor that one learns to recognize what sort of pain one has by referring to the place where the injury causing the pain is. But once one has learned to identify all the many different sorts of pain, i.e. once one has learned to say that the pain I now have is like the pain I normally have when my leg hurts, then one can allow for abnormal cases, cases where there is a 'leg-achy' sensation without there being anything wrong with one's legs.
(1) The purpose of this chapter is to attempt to clarify the nature of that identity, which according to the identity theory, holds between the process of having a sensation and BP. I shall call that particular identity 'theoretical identity.'

I shall show what theoretical identity is by stating its necessary conditions which are: identity of properties (see Section [2]); empirical identity (Section [3]); a particular kind of extensional identity (Section [4]); heterogeneous identity (Section [5]); reductive identity (Section [6]).

Another way of expressing the purpose of this chapter is to say that it is an attempt to specify the conditions under which one is justified in passing from the truth of the correlation statement to the truth of the identity statement.

(2) Identity of properties

The identity statement is being used here to assert that two processes are identical. However, two processes
are identical only if the properties of the two processes are identical. Thus, the truth of the identity statement seems to depend on whether or not each and every property of the process of having a sensation is also a property of BP.

It is objected, however, that if two processes, $P_1$ and $P_2$, are asserted to be theoretically identical, then before the identity could be justifiably asserted both $P_1$ and $P_2$ would have to be identified independently of each other. However, if $P_1$ and $P_2$ were identifiable independently, runs the objection, each would need to have properties the other did not have. For, identifiability is based on the properties of the processes which are to be identified. If two processes had different properties, they could not be identical, while if they had all the same properties they could not be identified independently. It seems, therefore, that either there are two distinct processes, or there is only one process, logically identical with itself. Neither alternative leaves room for the sort of empirical identity that the identity theory requires.

The special relevance of this objection to the theoretical identity of sensations and BP follows from the alleged fact that two processes may be identified on the basis of what are taken to be irreducibly different properties. Sensations may be identified on the basis of psychic

---

1This is a version of the objection reportedly advanced by Max Black. Cf. Smart, PSR, p. 94.
properties introspectable by the subject and publicly observable by no one. BP, on the other hand, may be identified on the basis of physical properties observable, in principle, by anyone and introspectable by no one. It follows that if these properties are indeed irreducibly different, then BP and sensations cannot be theoretically identical.

The objection stated in the last two paragraphs is, thus, double-barrelled. First, it is a general objection against the theoretical identity of processes. Second, it is a particular objection against the theoretical identity of sensations with BP.

The first part of the objection is that if two processes had different properties they could not be identical; if, on the other hand, they had identical properties, they could not be identified independently. The objection is based on a failure to realize that, first, when a process is identified, one need not know or use all the properties it has when making the identifying reference. Second, if two processes have identical properties, they may be identified independently on the basis of different properties.

To illustrate these points, consider the following counter-examples. It is possible to make an identifying reference to a process even though one may be ignorant of many of its properties. For example, I may identify a particular sensation of pain by saying: 'The pain I feel in my left toe now.' The sensation, of course, has many properties beside the one used for identifying it. However, my knowledge
or ignorance of physiology, or my expertise in giving introspective reports, are normally irrelevant to the achievement of identifying reference. Though there undoubtedly are cases when only a skilled diagnostician, or an expert self-scrutinizer could identify a particular sensation.

Secondly, two processes may be theoretically identical, and yet be identified independently on the basis of different properties. For example, we may identify a natural phenomenon, e.g. lightning, on the basis of its position in the sky, speed, colour, etc. And we may identify a natural phenomenon, e.g. the collision of ionized particles in the atmosphere, on the basis of its electric charge. The fact that lightning and the collision of ionized particles in the atmosphere are identifiable by means of different properties, does not constitute an objection against their theoretical identity.

Analogously, from the fact that sensations may be identified on the basis of introspectable properties, while BP may be identified on the basis of observable properties, it does not follow that sensations and BP cannot be theoretically identical. They may be theoretically identical, and they may be identified on the basis of different properties which, in fact, they have in common.

To generalize the point: a process \( P_1 \), may be constituted only of the properties \( p, q, r, s \), and the process, \( P_2 \), may also be constituted solely of the properties \( p, q, r, s \); \( P_1 \) and \( P_2 \) may be then identical. \( P_1 \), however, may be
independently identified on the basis of \( p \), while \( P_2 \) may be identified independently on the basis of \( q \).

The truth of the statement asserting the identity of processes depends on the truth of the statement asserting the identity of their properties. The independent identification of two processes leaves it an open question whether or not they are identical. It may be the case that it simply is not known whether the independently identified processes are identical. Or, it may be the case that even though sensations are identified on the basis of psychic properties they are, as further investigation might show, theoretically identical with \( \text{BP} \). The onus of showing that this is the case, and of explaining how psychic properties could be physical properties is upon the identity theorist.

We are thus led back to the second point of the double-barrelled objection mentioned above, viz., that \( \text{BP} \) and sensations cannot be theoretically identical because they are identified on the basis of irreducibly different properties. It has to be granted, of course, that if sensations and \( \text{BP} \) do have irreducibly different properties, then they cannot be theoretically identical. It has to be granted also that a phenomenological description of sensations does support the view that they may have different properties. The crucial question is: are these properties irreducibly different? The identity theory is tenable only if the difference in the properties of sensation and \( \text{BP} \) is not irreducible. If their properties are not irreducibly
different, this may allow that what a person knows introspectively is theoretically identical with what physiologists observe instrumentally.

Two arguments will be offered in favour of this possibility. First, if a process, $P_1$, has the property $q$, and if a process, $P_2$, has the property $r$, then it is possible that $P_1$ and $P_2$ are one and the same process, having the properties $q$ and $r$. That is, just because sensations have introspectable properties and BP have instrumentally observable properties, it does not follow that sensations and BP are distinct. It may well be that there is only one process, having both introspectable and instrumentally observable properties, yet identified by the use of two descriptions.

This consideration is sufficient to refute the a priori objections, to the theoretical identity of sensations and BP discussed above. The refutation of these a priori objections, however, falls short of providing one of the necessary conditions for the truth of the identity statement. For, while a dualism of processes can no longer be maintained on these a priori grounds, dualism of properties remains.

It is necessary, therefore, to advance a second argument against these a priori objections to the theoretical identity of sensations and BP. The argument is based on the possibility that one and the same property of a given process may appear to be of one sort when subjected to one type of observation, and of quite a different sort when subjected to
another type of observation. This possibility rests on actual cases, e.g. the properties of lightning appear to be radically different from the properties of the corresponding collision of ionized particles in the atmosphere; change in the temperature of gases appears to have quite different properties from the properties that the change in the mean kinetic energy of corresponding molecules appears to have. Yet these appearances are deceptive; members of these pairs are theoretically identical.

The fact that in some cases the apparent difference between properties, which subsequently turn out to be theoretically identical, is explicable in terms of the distinctions between different types of observation, may possibly apply to sensations and BP. Thus an apparent difference of properties is not a priori incompatible with both the identity and the distinctness of the processes of which they are properties. This shows that the apparent differences between the properties of sensations and BP is not a ground for an a priori objection against the identity theory.

(3) Empirical identity

The identity of sensations and BP is a kind of empirical, as opposed to logical, identity. That is, the identity is not asserted to hold between the meanings of the expressions 'sensations' and 'BP.' The force of the identity is that these expressions denote the same process.\(^2\) The claim

\(^2\) 'Denotation' is used in the sense that a word is said to denote a particular if the word truly applies to the particular. W denotes P = Df. W truly applies to P.
is not that the expressions 'sensations' and 'BP' are interchangeable in any or all contexts; the claim is rather that these expressions denote one and only one process.

The identity theory recognizes that 'sensations' and 'BP' are not interchangeable in all contexts without loss of meaning. It is clear, for example, that the assertion 'I have a tingling sensation in my toe' is not interchangeable, without loss of meaning, with the assertion 'I have a BP in my toe.'

That sensations and BP are identical is partly an empirical or factual claim. The extent to which the truth of the correlation statement is an empirical or factual claim is the extent to which the assertion of the identity is an empirical or factual claim. The truth of the empirical or factual claim can be determined by testing but not, of course, solely by an analysis of the expressions 'sensations' and 'BP.'

Not all cases of empirical identity, however, are cases of theoretical identity. An empirical identity may also be, what I shall call, 'accidental.' In statements used for expressing the accidental identity of two or more particulars, the particulars are denoted either by proper names, e.g. 'Cicero = Tully,' or by definite descriptions, e.g. 'The tallest man in the room = the heaviest man in world.' An identity which is expressed by the use of proper names is the result of the christening convention whereby a particular may have several proper names. If the conventions were such that
each particular had but one proper name, accidental identity expressed by the use of proper names would be eliminated. As Quine\(^3\) points out: "To rid language of redundant nomenclature of the simple type, e.g. 'Tully' and 'Cicero,' would be no radical departure."

The other type of accidental identity is expressed by the use of definite descriptions. The identity in this case is purely coincidental. The force of 'coincidental' is that the identity holds either in isolated, unique cases, or, if it holds in more than one case, it does so irregularly and unpredictably. For instance, the tallest man in a certain room may indeed be the heaviest man in the world. But from the fact that these two definite descriptions happen to denote the same person, it in no way follows that the heaviest man in the world will regularly be the tallest man in that room. Indeed, if this identity held regularly, and if each time one walked into a certain room the tallest man there was also the heaviest man in the world, we would find this most queer and suspect that someone is playing a clever trick on us. The general point is that accidental identity holds irregularly in particular situations only and there is no reason to expect the recurrence of the coincidence upon which the identity is based.

Of course what is taken to be an accidental identity

---

at one time may be found to be a theoretical identity when more information comes in, and *vice versa*. The point is that a justifiable claim that two processes are theoretically identical requires that one do something more than show that two expressions denote the same process.

Accidental identity, of course, is not the identity that is alleged to hold between sensations and BP. The claim of the identity theory is that sensations and BP are regularly and invariably and predictably identical. Thus the identity claim made by the theory is too strong to be satisfied by accidental identity. This stronger claim would be partly satisfied if each and every member of the class of sensations were also a member of the class of BP, and *vice versa*.

(4) Extensional identity

Theoretical identity is a kind of extensional identity. That is, theoretical identity does not hold solely in virtue of the meanings of the expressions used for ascribing allegedly identical properties to particulars. Nor does theoretical identity hold on the basis that any expression that can be used for ascribing a property to one member of the putative identity can also be used for ascribing the same property to the other member. Theoretical identity holds between all the extensional properties of one particular and some of the non-extensional properties of another particular. Some of the typical extensional properties that
a particular may have are: having shape, having size, having weight, having duration, having location, being in motion, interacting with the environment, etc.

Predicates standing for intensional and modal properties are excluded from theoretical identity. If these predicates were not excluded, it would be an easy matter to deny many well-established theoretical identity statements. Consider, for example, the true theoretical identity statement 'Chemical bond = electromagnetic forces between atoms within a molecule.' It would be true to say of a person who knows little science that 'He knows that "Chemical bond = chemical bond,"' but false to say that 'He knows that "Chemical bond = electromagnetic forces between atoms within a molecule."'

Modal predicates have to be excluded also. For if they were not excluded, modal predicates, as well as intensional ones, could be used to deny true theoretical identity statements on purely a priori grounds. For example, it is true that 'Water = H₂O,' but while it is true that 'It is logically necessary that "Water = water,"' it is false that 'It is logically necessary that "Water = H₂O."'

The justification for excluding predicates standing for intensional and modal properties from the concept of theoretical identity stems from the fact that those predicates can be used to give different characterizations of particulars that are in fact identical, simply because of the meaning of the predicate or the linguistic role the predicate plays in a given context.
If it is recognized that the theoretical identity of sensation and BP excludes identity with regards to intensional and modal properties, then several standard objections against the identity theory will turn out to be ignoratio elenchi arguments.

Sensations, it is argued against the identity theory, are necessarily mental, BP are not. Sensations are described by the use of predicates appropriate for the description of the mental, while BP are described by the use of predicates appropriate for the description of the physical. Sensations are described as being, for example, necessarily private, the person experiencing them is said to have a necessarily privileged access to them, it is taken to be necessarily the case that one is non-cognitively aware of the having of a sensation, avowals of sensations are supposed to be logically incorrigible, descriptions of sensations are said to be necessarily intentional. None of these can be said of BP. Consequently, it is argued, sensations and BP cannot be identical.

The reply to this cluster of objections is that insofar as the above properties are non-extensional, i.e. properties that sensations have in virtue of being examined in a particular context, the objection does not apply. It is perfectly compatible with the theoretical identity of

---

\(^4\) E.g. K. Baier, 'Smart on Sensations,' *A. J. P.*, 40 (1962), pp. 57-68. Also Coburn, 1963.
sensations and BP that they have quite different non-extensional properties.

The problem now is to provide a method for distinguishing extensional and non-extensional properties. In my view, extensional properties are, in a sense, context-independent, while non-extensional properties are, in a sense, context-dependent. I shall now attempt to clarify what these senses are.

It seems to me that it is true of all normal human beings that when they perceive particulars with their unaided senses, they describe these particulars in certain ways, e.g. as having spatial dimensions, as having duration, as being in motion or being stationary, as interacting with the environment, etc. Not only are these ways of perceiving particulars common to all normal human beings, they are, I hold, criterial\(^5\) for being a normal human being. That is, one way of telling whether a creature is a normal human being is by determining whether when he perceives the world through his unaided senses he describes it in roughly the same way as we do.

Normal human beings, describing the world on the basis of unaided sense perception, have a particular world view; they have a rudimentary theoretical structure. I shall call this primitive theory 'the common sense view of the

\(^5\)See my discussion of the concept of criterion, Chapter Three, Section (6).
world.' The language of the common sense view of the world is ordinary language. It is, I hold, criterial for being a normal human being that one uses the common sense view of the world when one describes the world on the basis of unaided sense perception.

By a property being 'context-independent' I mean that it belongs to that class of properties of which it is true to say that their being perceived and described in certain ways is criterial for being a normal human being. In other words, a property is context-independent if it is part of the common sense view of the world. By a property being 'extensional' I mean that it is context-independent.

The justification for calling these properties 'context-independent' is not that particulars must be perceived and described as having them no matter what the context is, but rather that no matter in what context the particulars are perceived and described, the context of the common sense view of the world is presupposed. This latter context is presupposed in the sense that even the most sophisticated scientific theory has as its starting point the way in which we ordinarily perceive and describe the world. It is always within the framework of the common sense view of the world that the initial problems, that later bloom into full-fledged branches of science, are stated. Ultimately it is always with reference to unaided sense perception that instruments are designed and calibrated.

There are properties, apart from intensional and
modal ones, that are context-dependent. The context-dependency relevant for my purpose is that of being dependent on the context of a scientific theory. For example, a normal human being perceiving the world while using the framework of physics may describe particulars as having the property of being comprised of molecules, having a certain number of electrons, etc., or if he were perceiving the world while using the framework of chemistry, he may describe particulars as having the property of being soluble in acid, forming a stable compound with H₂O, etc.

By a property being 'context-dependent' I mean that the property is perceived and described the way it is because it is perceived and described while using the framework of a scientific theory. By a property being 'non-extensional' I mean that it is context-dependent.

Thus the way in which I propose to distinguish exten-sional and non-extensional properties is by reference to the theoretical framework which is being used when the particulars of which they are properties are being perceived and described. Extensional properties are those that particulars are perceived and described as having when using the framework of the common sense view of the world.

Ancient Greek peasants, Phoenician priests, Eskimos, astronauts, and scientists outside of their laboratories, all perceive and describe particulars as having shape, size, duration, weight, as being in motion or being stationary, as interacting with the environment, etc. And if the
Neanderthal Man had no sense of duration, or if he could not distinguish between stationary particulars and particulars in motion, then, I think, we would have reason for saying that he was not a normal human being, in the sense in which we use that epithet. Non-extensional properties are those that particulars are perceived and described as having when operating within the framework of a scientific theory.

It is vitally important to realize that I am not advocating here the logical or ontological primacy of the common sense view of the world. In other words, I do not subscribe to the view that properties as perceived by the unaided senses are the 'given.' I do not, that is, hold the view that the common sense view of the world is the view of the world as it really is. On the contrary, I hold that one may come to a much more precise and profound understanding of the particulars that are components of the common sense view of the world if they are described in terms of scientific theories. I shall argue in the next section that extensional properties are reducible to non-extensional properties.

On the other hand, it ought to be accepted, I think, that even the most sophisticated scientific account must start with the common sense view of the world, with extensional properties. The common sense view of the world is the ultimate starting point of all enquiry. The reason for this is that in order to have a sophisticated view of the world, normal human beings, as things are now, must first have a simple view. And the simple view is the common sense
view of the world, obtained through the unaided senses. And if it is asked: 'Why is this so?' the only answer is that this is the way human beings are.

Let us return now to sensations and BP. The identity between them should be construed as holding in virtue of all the extensional properties of sensations and some of the non-extensional properties of BP. In other words, the two processes are identical because all context-independent properties of sensations are identical with some of the context-dependent properties of BP, i.e. some of those context-dependent properties that are neither intensional nor modal. In the next two sections I shall propose conditions under which extensional and non-extensional properties could be said to be identical. I shall argue that the identity is asymmetrical and reductive.

It might be objected, however, that the notion of theoretical identity is incoherent, because it is a mistake to suppose that the theoretical identity of two particulars can ever be established. It is, after all, always in a certain context, be it the common sense view of the world or a scientific theory, that we perceive and describe any and all particulars. But if particulars are always perceived and described in a context, what one has to do is to establish that a particular described in one context, e.g. in the

---

6If I understand it correctly, this is the objection raised by W. Sellars in 'The Identity Approach to the Mind-Body Problem,' Rev. of Meta., 18 (1965), pp. 430-451.
context of the common sense view of the world, can be identical with a particular described in another context, e.g. in the context of a scientific theory. But how could this be done?

One way of showing that putatively identical particulars, perceived and described in different contexts, are theoretically identical, is by showing that the contexts in terms of which the particulars are perceived and described are not really different. A possible way of doing this is by arguing that there is a reductive relation between all veridical ways of describing the world. That is, the theoretical identity of particulars perceived and described in apparently different contexts could be established if all veridical descriptions offered in one context could be shown to be reducible to veridical descriptions offered in the other context. The program of showing that this is the case is the program of the unity of science. I shall discuss this shortly.

(5) Heterogeneous identity

Theoretical identity is an empirical identity holding between all the extensional properties of one member of the putative identity and some of the non-extensional properties of the other member. I argued in the previous section that

---

the non-extensional properties of sensations should be excluded from the consideration of theoretical identity. I want to note now that BP have no extensional properties at all. This somewhat unusual conclusion follows from the analysis of 'extensional property' in terms of 'context-independent property.' All properties of BP are context-dependent, and therefore non-extensional, because BP are perceived and described in terms of a sophisticated scientific theory, viz. neurophysiology.

In this section I shall argue for the claim that it is not a necessary condition of theoretical identity that each and every property one member of the putative identity has must be shared by the other member. I have argued for this claim already (in Section [4]) in respect to intensional and modal properties. In this section I shall argue that it holds for other non-extensional properties as well.

Consider, for example, cases of single cell mutations. In such cases the cell survives under conditions that were previously lethal. The following identity statement holds: 'Viability under previously lethal conditions = certain specifiable changes in DNA.' Now it is true of the process for which the expression on the left hand side of the '=' sign stands that it can be observed sometimes with the naked eye, sometimes only microscopically, while it would be false to say this of the process for which the expression appearing on the right hand side of the '=' sign stands; certain specifiable changes in DNA can be observed only by the use of
highly sophisticated chemical or physical techniques. So, while the two processes do not share all their properties, they can still be regarded as being identical.

In order to characterize this further feature of theoretical identity it is important to distinguish between heterogeneous and homogeneous identity statements.

Some examples of homogeneous identity holding between processes are:

(HO_1): The Morning Star twinkling = Venus twinkling
(HO_2): Scott dying = the author of Waverley dying
(HO_3): Cicero being eloquent = Tully being eloquent

Some examples of heterogeneous identity holding between processes are:

(HE_1): Chemical bond between two atoms being established = change in the reduction potential of the two atoms
(HE_2): Change in viscosity = change in the mean distance between molecules
(HE_3): Change in the temperature of a gas = change in the mean kinetic energy of the molecules of the gas.

The relevant difference between HO and HE identity is that if HO identity holds, all expressions that stand for the properties of the process denoted by the expression on the left hand side of '=' sign can also be used to stand for the properties of the process denoted by the expression on the right hand side of the '=' sign. But in the cases of HE identity this is not the case.

The explanation of this difference holding between HO and HE identity is that the processes that are HO identical occur in one and the same, albeit rudimentary, theory, viz., common sense, while the processes that are HE identical occur
in different theories. HO identity is intra-theoretical, while HE identity is inter-theoretical.

The 'Sensations = BP' identity statement is of the HE type. The theory in which 'sensation' occurs is the common sense view of the world, while the theory in which 'BP' occurs is neurophysiology. The problem is: What is the force of saying that the denotations of the expressions appearing on opposite sides of the '=' sign in the case of HE identity are identical? That is, what are the conditions in which two processes, perceived and described in different theories, can be said to be theoretically identical?

(6) Reductive identity

The general answer is that HE identity holds between particulars denoted by different expressions if the theory \( T_2 \) in which one of these expressions occurs is micro-reducible to the theory \( T_1 \) in which the other expression occurs. I shall refer to \( T_2 \) as the 'reduced' theory, while \( T_1 \) will be called the 'reducing' theory. The problem of testing for identity of the HE type becomes: What are the conditions for the micro-reducibility of \( T_2 \) to \( T_1 \)?

\( T_2 \) is micro-reducible to \( T_1 \) if and only if:

(1) The vocabulary of \( T_2 \) contains terms that

---

(a) are not in the vocabulary of $T_1$, and
(b) are not definable in the vocabulary of $T_1$, and
(ii) Any observational data explainable by $T_2$ is explainable by $T_1$, and
(iii) $T_1$ is at least as well systematized as $T_2$.

$T_1$ is as well systematized as $T_2$ if the reducing theory, $T_1$, is at least as simple as the reduced theory, $T_2$, assuming $T_1$ and $T_2$ to be of like strength. However, if $T_1$ is a much stronger theory than $T_2$, then it seems reasonable to allow $T_1$ additional complexity over $T_2$. Thus the degree of systematization of a given theory depends on the balance between the strength and the simplicity of the theory in comparison with other theories. $T_1$ is stronger than $T_2$ if $T_1$ explains everything that $T_2$ does, and in addition $T_1$ explains more than $T_2$ does; and/or if $T_1$ allows a greater number of significant predictions than does $T_2$.

The micro-reductive relation outlined above may hold not only between theories, but also between branches ($B$) of science, e.g. chemistry, physics, etc. I shall assume that a given B of science is constituted solely of several theories. If these theories are micro-reducible, then the B, constituted of these theories, can also be micro-reduced. Accordingly, $B_2$ is micro-reducible to $B_1$ if and only if for any theory or theories in $B_2$ ($T_{B2}$) there is a theory or theories in $B_1$ ($T_{B1}$) such that $T_{B2}$ is micro-reducible to $T_{B1}$. The conditions for the micro-reduction of $T_{B2}$ to $T_{B1}$ are exactly the same as are the conditions for the micro-reduction.
of $T_2$ to $T_1$, outlined above.

The micro-reduction required by an identity of the HE type is defined by Oppenheim and Putnam\(^9\) as follows:

The essential feature of a micro-reduction is that the branch $B_1$ deals with the part of the objects dealt with by $B_2$. We must suppose that corresponding to each branch we have a specific universe of discourse . . . , and that we have a part-whole relation. . . . Under the following conditions we shall say that the reduction of $B_2$ to $B_1$ is a micro-reduction: $B_2$ is reduced to $B_1$; and the objects in the universe of discourse of $B_2$ are wholes which possess a decomposition into proper parts all of which belong to the universe of discourse of $B_1$. For example, let us suppose that $B_2$ is a branch of science which has multicellular living things as its universe of discourse. Let $B_1$ be a branch with cells as its universe of discourse. Then the things in the universe of discourse of $B_2$ can be decomposed into proper parts belonging to the universe of discourse of $B_1$. If, in addition, it is the case that $B_1$ reduces $B_2$ at the time $t$, we shall say that $B_1$ micro-reduces $B_2$ at time $t$.

The concept of 'decomposition,' used by Oppenheim and Putnam, has been defined by Rescher and Oppenheim\(^10\) in the following way:

To specify some concept of part we must state the conditions under which an object $x$ is to be considered as a part (in this sense) of an object, or whole, $y$. The specification of a particular part-whole relation thus determines for a given object, or whole, $w$, just which objects are its parts, i.e. for what objects $x$ it is the case that $x$Ptw. It is only when the context makes clear which specific part relation is intended that we may speak of 'parts' and 'wholes' without further ado. Now, given a particular object, $w$,

---

\(^{9}\)Oppenheim and Putnam, p. 6.

\(^{10}\)N. Rescher and P. Oppenheim, 'Logical Analysis of Gestalt Concepts,' Brit. J. for the Phil. of Science, 6 (1955), pp. 89-106, see p. 91.
and some specific part relation, \( P_t \), we will say that a class \( D \) of \( P_t \)-parts of \( w \) (i.e. parts as specified by \( P_t \)) is a decomposition of \( w \) if every \( P_t \)-part of \( w \) has some \( P_t \)-part in common—i.e. overlaps—with at least one element of \( D \). Clearly, the class of all \( P_t \)-parts of an object is a decomposition in this sense, but the object may have other decompositions which do not include all of its \( P_t \)-parts.

The relation of micro-reduction has the following logical features:

(i) transivity, i.e. if \( B_3 \) micro-reduces to \( B_2 \) and if \( B_2 \) micro-reduces to \( B_1 \), then \( B_3 \) micro-reduces to \( B_1 \); and

(ii) irreflexivity, i.e. no \( B \) micro-reduces to itself; and

(iii) asymmetry, i.e. if \( B_1 \) micro-reduces \( B_2 \), then \( B_2 \) never micro-reduces \( B_1 \).

Transivity and asymmetry can be explained on the basis of the part-whole relationship holding between the objects of the respective universes of discourse of \( B_1 \), \( B_2 \), and \( B_3 \). For if the objects of \( B_2 \) are decomposable into the objects of \( B_2 \) and if the objects of \( B_2 \) are decomposable into the objects of \( B_1 \), then the objects of \( B_3 \) are decomposable into the objects of \( B_1 \); hence transivity. Further, if the objects in the universe of discourse of \( B_1 \) are parts of objects in the universe of discourse of \( B_2 \), then while the objects of \( B_2 \) can be decomposed into the objects of \( B_1 \), the objects of \( B_1 \) cannot possibly be decomposed into the objects of \( B_2 \); hence asymmetry. Irreflexivity follows simply from the fact that no object can be part of itself.

(7) Theoretical identity and physicalism

The consequent interpretation of theoretical
identity, i.e. the identity that according to the identity theory holds between sensations and BP, is the following: the identity is theoretical if and only if it is
(i) identity of properties,
(ii) empirical, as opposed to logical, identity,
(iii) identity of all the extensional properties of sensation and some of the non-extensional properties of BP
(iv) heterogeneous, as opposed to homogeneous, identity,
(v) reductive identity of the micro-reducing type.

The identity theory can be, though need not be, regarded as a special theory within physicalism. Physicalism is the metaphysical view according to which the unity of science is possible, and it will be achieved if all sciences are micro-reduced, directly or indirectly, to physics. The micro-reduction of all sciences implies that, if the micro-reduction is achieved; there will be a unity of vocabulary of all sciences, and also that there will be a unity of laws of all sciences.

Oppenheim and Putnam\textsuperscript{11} offer the following definition:

The term 'Unity of Science' will be used in two senses to refer, first, to an ideal state of science, and, second, to a pervasive trend within science, seeking the attainment of that ideal. In the first sense 'Unity of Science' means the state of unitary science. It involves . . . unity of vocabulary, or 'Unity of Language'; and unity of explanatory principles, or 'Unity of Laws.' That Unity of Science, in this sense can be fully

\textsuperscript{11}Oppenheim and Putnam, p. 4.
realized constitutes an over-arching meta-scientific hypothesis which enables one to see a unity in scientific activities that might otherwise appear disconnected or unrelated. In the second sense Unity of Science exists as a trend within scientific inquiry, whether or not unitary science is ever attained, and notwithstanding the simultaneous existence, (and, of course, legitimacy) of other, even incompatible trends.

Physicalism, or the view that all sciences are micro-reducible to physics, is a highly controversial metaphysical view that I do not propose to defend here, although I am very sympathetic to it. However, the identity theory may be held independently of physicalism. It may be the case that the common sense view of sensations can be reduced to a neurophysiological view, which would guarantee the success of the identity theory, and yet physicalism may still fail due to the non-micro-reducibility of some other science. As far as the identity theory in its present form is concerned, it would be sufficient for its success if it could be shown that sensations and BP are theoretically identical.

(8) The theoretical identity of sensation and BP

We are now in a position to specify the conditions that must obtain so that the theoretical identity of sensations and BP can be established. These conditions are: first, that the correlation statement be true, and second, that the identity statement be true. The truth of the correlation statement is a necessary, but not sufficient, condition of the truth of the identity statement.

The correlation statement has empirical and
philosophical components. The philosophical component is the requirement that both terms of the correlation, viz., sensations and BP, be independently identifiable. Sensations, it was argued, are identifiable through avowals and non-verbal behaviour, while BP are identifiable through brain-readings. It is also a philosophical question whether sensations are the sorts of processes that can be correlated with BP. It was argued that since sensations and BP are spatially and temporally determinable their correlation can be tested.

The empirical component of the correlation statement is the actual testing of the assertion that there is a perfect correlation between sensations and BP. There is a perfect correlation between sensations and BP if the following conditions are fulfilled:

(i) Sensations and BP occur invariably in spatio-temporal coincidence, that is, at any given time, if a particular sensation, e.g. pain occurred, there also occurs a particular BP. Further, the particular sensation must have exactly the same spatial location as the particular BP. Sensations and BP, being processes, can be spatially located with reference to the particular of which they are processes. It was argued

12 Chapter Three, Sections (4)-(6), and Chapters Four and Five.
13 Chapter Three, Section (3).
14 Chapter Five.
15 Chapter Five.
that the particular is the brain.

(ii) Sensations and BP are non-separable
That is, for each and every sensation there is a particular BP such that when and only when the sensation occurs the particular BP also occurs, and vice versa.

(iii) Sensations and BP are not manipulable with independent or unrelated, i.e. non-corresponding, effects.
That is, if due to circumstances either sensations or BP were to undergo alterations there would be corresponding alterations in the other term of the correlation. For example, if due to brain damage a person's BP were abnormal, there would be a corresponding decrease in the person's sensitivity to certain stimuli. Or, if due to mental illness a person's capacity for having certain sensations were impaired, there would be a corresponding alteration in the functioning of his brain.

If these conditions are fulfilled, then one would be justified in asserting that sensations and BP are perfectly correlated. It is, however, compatible with this state of affairs that sensations and BP are two distinct processes. The identity theory, therefore, needs to fulfill additional requirements for the justification of the claim that 'sensations' and 'BP' denote one and the same process.

The additional claim is that sensations and BP are theoretically identical because of the micro-reducibility of all the extensional properties of sensations to some of the non-extensional properties of BP. This claim is justified
if the following requirements are met: the part of the common sense view of the world which is concerned with sensations is reducible to a branch of science, viz., physiology in general, neurophysiology in particular. This is accomplished if

(i) The vocabulary of the common sense view of sensations contains terms that

(a) are not in the vocabulary of physiology. 'Pain' and 'awareness' are examples of such terms.
(b) are not definable in the vocabulary of physiology. 'Pain' and 'awareness' are not so definable, due to the fact that these terms may occur in intensional and modal contexts.

(ii) Any observational data concerning sensations that are explainable by the common sense view are explainable by physiology.
For example, if there is a common sense explanation for the fact that pain normally causes avoidance behaviour, physiology must be capable of providing a physiological explanation of the same phenomenon.

(iii) Physiology is at least as well systematized as is the common sense view of sensations.
That is, physiology must be capable of explaining everything that the common sense view of sensations does.

If these conditions are fulfilled, then the identity theory is justified in claiming that sensations and BP are theoretically identical. It has to be admitted immediately
that the conditions have not been satisfied. However, the point the identity theory aims at establishing is that these conditions could be satisfied. That is, one of the main purposes of the identity theory is to show that there are no valid philosophical or logical objections against undertaking an investigation a result of which may be the conclusion that sensations are theoretically identical with BP.

There is, however, an objection against this program. The objection is that even if the full program of the identity theory is carried out, one would still not be justified in eliminating the possibility that sensations and BP are two distinct processes. The purpose of the next chapter is to meet this objection.
CHAPTER EIGHT

THE IDENTITY THEORY, DUALISM, AND THE CRITERIOLOGICAL VIEW

I

(1) The purpose of the first part of this chapter is to examine the objection that theoretical identity is still compatible with the distinctness of sensations and BP. The objection is stated by Kim\(^1\) in the following way. In considering the reduction of one theory to another he writes:\(^2\)

> If the reduction is to be genuinely intertheoretic, the reduced theory will contain concepts not included in the vocabulary of the reducing theory, and these concepts will occur essentially in the laws of the reduced theory. Hence, if these laws are to be derived from the laws of the reducing theory in which these concepts do not occur, we shall need, as auxiliary premises of derivation, certain statements in which concepts of both theories occur. We may refer to those statements as "connecting principles."

In accordance with this, an identity statement, used to assert the theoretical identity of sensations and BP, may be thought of as a connecting principle. As Kim writes:\(^3\)

---

\(^1\) Kim, pp. 228-229.

\(^2\) Kim, p. 228.

\(^3\) Kim, p. 229.
Psycho-physical identity statements can serve as such connecting principles, just as statements like "gas is a collection of molecules" and "temperature is the mean kinetic energy of molecules" serve as connecting principles in the reduction of classical thermodynamics to statistical mechanics.

Kim then argues that while this seems to be a plausible view, it does not support the assertion that 'sensations' and 'EP' denote one and the same process.

What needs to be shown is that unless the identification is made, the derivation of mentalistic laws from neuro-physiological laws is impossible. That is, it has to be shown that nothing less than psycho-physical identity statements will do as psycho-physical connecting principles. But it is dubious that this can be shown; in fact, psycho-physical correlation statements seem sufficiently strong to function as the requisite connecting principles.

Kim's contention is that the correlation statement may be true, and psychology may be reducible to physiology, and yet the corresponding identity statement may not hold.

In attempting to meet this objection we have to ask: what are the conditions in which the identity statement would be falsified? One of these conditions would be if sensations have properties that failed to fit into the space-time-causal network, and consequently could not be reduced to properties that do fit into that network. If sensations do have such properties, then not all properties of sensations are reducible, and thus one of the conditions of theoretical identity, viz., property identity, would not be fulfilled.

The result of this, it seems to me, would be the abandonment of the identity theory. For, if sensations have properties that do not fit into the space-time-causal network,
then sensations must be thought of as having non-physical, psychic, properties. This conclusion is precisely the one that Feigl appears to adopt at the end of his long monograph.\textsuperscript{4}

In distinguishing his position from epiphenomenalism Feigl\textsuperscript{5} writes:

Our solution of the mind-body problem differs quite fundamentally from materialistic epiphenomenalism in that: (1) it is monistic, whereas epiphenomenalism is a form of dualistic parallelism; (2) the "physical" is interpreted as a conceptual system (or as the realities described by it), but not as the primary kind of existence to which the mental is appended as a causally inefficacious luxury, or "shadowy" secondary kind of existence; (3) quite to the contrary, mental states experiences and/or knowable by acquaintance are interpreted as the very realities which are also denoted by a (very small) subset of physical concepts. . . . Speaking "ontologically" for the moment, the identity theory regards sentience (qualities experienced, and in human beings knowable by acquaintance) and other qualities (unexperienced and knowable only by description) the basic reality.

It seems then that 'basic reality' is made up of two kinds of qualities: those that fit into the space-time-causal network, and those that do not. Sensations have both kinds of qualities or properties. Insofar as they have the former kind, they are reducible to BP, insofar as they have the second kind, they are non-reducible. Feigl calls those qualities that do not fit into the space-time-causal network, collectively, 'raw feels.'

\textsuperscript{4}H. Feigl, 'The "Mental" and the "Physical,"

\textsuperscript{5}Feigl, p. 474.
Let us go back now to Kim's objection and restate it. It may be argued from the putative fact that the assertion of a perfect correlation between sensations and BP is true, and from the putative fact that the common sense view of sensations is reducible to physiology it does not follow that all the properties sensations have are also properties that BP have. For sensations do, while BP do not, have a 'raw feel' component. Consequently, sensations and BP cannot be identical. It is necessary to enquire, therefore, whether or not sensations have an irreducible raw feel component.

(2) As far as I know the expression 'raw feel' was first used by Tolman. Presenting, with the view of criticising, an argument against his own position Tolman writes:

Sensations, says the orthodox mentalist, are more than discriminanda-expectations. .... They are in addition immediate mental givens, "raw feels." They are unique subjective suffusions in the mind. And it is these "raw feels," these suffusions, which constitute the ultimate entities in which as psychologists we are, or should be, interested.

Tolman does not deny the existence of raw feels, nor for that matter, does he enquire what "subjective suffusions" are. He does go on to say, however, that even if raw feels exist, scientists qua scientists are justified in disregarding them.

If there be "raw feels" .... these "raw feels" are by very definition "private" and not capable


7Tolman, pp. 250-251.
of scientific treatment. And we may leave the question as to whether they exist, and what to do about them, if they do exist, to other disciplines than psychology—e.g. to logic, epistemology, and metaphysics. And whatever the answers of these other disciplines, we, as mere psychologists need not be concerned.

In support of this position Tolman quotes Lewis: 10

In the end, the supposition of a difference in immediate experience ("i.e. our raw feels," says Tolman) which is not to be detected through divergence in discrimination and relations, is a notion very difficult to handle. Because such difference would, ex hypothesi, be ineffable. We can have no language for discussing what no language or behavior could discriminate. And a difference which no language or behavior could convey is, for purposes of communication, as good as non-existent.

Accordingly, Tolman proposes to discard raw feels onto one of several "scrap-heaps." We could ignore them; or, we could admit their existence, but recognize that their existence "makes no difference"; or, we could proceed on the assumption that "raw feels may be the way physical realities are intrinsically, i.e. in and for themselves," or, with a slight variation, we could argue following Russell, "that experienced qualities are the intrinsic nature of a nervous process."

Tolman qua psychologist is perhaps justified in disregarding raw feels. The identity theory, however, cannot afford the same luxury. One of the objections against the

8 Tolman, p. 253.
9 Tolman, p. 426.
11 Tolman, pp. 426-427.
theoretical identity of sensations and BP is based on the alleged existence of raw feels.

(3) It is extremely difficult to state precisely what the objection is, because it is not easy to discover what raw feels are supposed to be. To say that raw feels are 'ineffable' or 'subjective suffusions,' or that raw feels 'may be the way physical realities are intrinsically,' is to obscure the problem even more. The difficulty in stating what raw feels are supposed to be is due to the fact that even if they exist we talk about them only in terms of their alleged causes and effects.

We might approximate, however, what raw feels are thought to be by the following illustration. Imagine a millenial neurophysiologist. Suppose that he can attach electrodes to the 'pain center' of the brain. If he sends mild electro-shocks through the electrodes the person may exclaim 'Ouch!', his blood pressure may increase, his pupils may dilate; in a word, we may detect the typical causes and effects of pain. If he remains conscious during the experiment, we can ask him: 'Do you feel pain?'; he may reply: 'No.' Something seems to be lacking. It is a raw feel that is lacking. Normally, the presence of the indicators of pain justify the assumption that the person feels pain. In this case the person felt no pain. The objection against the theoretical identification of sensations and BP is that if raw feels are not always present when BP are present, but are always present when sensations are present, then sensations
cannot be theoretically identical with BP.

In order to appreciate the force of this objection it is necessary to step outside, as it were, of the identity theory. If one does not do this, the objection will be felt to beg the question. Identity theorists would merely assume that the difference between the situation created by the millenial neurophysiologist's manipulations and ordinary pain-producing situations is the absence of certain BP, viz., those identical with raw feels. The immediately experienced qualities of sensations, i.e. those raw feels with which we are all supposed to be directly acquainted, are really nothing but BP. What we apprehend introspectively is identical with what neurophysiologists observe instrumentally. Talk about raw feels, argues the identity theorist, does nothing to invalidate this argument.

The disagreement between the identity theorist and the champion of irreducible raw feels is a disagreement about the question of proof. The latter feels that the former ought either to show how raw feels are reducible to BP, or abandon the identity theory. The former feels that unless the notion of irreducible raw feels is given cognitive significance no objection has been made. Accordingly, I shall enquire whether the assertion that there are irreducible raw feel components of sensations has cognitive significance.

Let us assume that the correlation statement is true. Not only the identity theory, but also different forms of dualism, viz., parallelism, epiphenomenalism, some forms of
interactionism, and the double aspect theory, are consistent with the truth of the correlation statement. The raw feel argument is in support of dualistic theories, and against the identity theory. Its purpose may be regarded as showing that it is unjustified to pass from the truth of the correlation statement to the assertion of the identity statement. I shall consider the raw feel objection as it was formulated by Feigl. The objection comes to this:

If the synthetic (i.e. contingent) element in the relation that we have stressed throughout is admitted, then there is something which purely physical theory does not and cannot account for. Is there then a kind of 'brute fact' which our monistic theory has to accept but for which there is possibly no explanation, in the same sense as there can be (within naturalistic empiricism) no explanation for the fact that our world is what it is. . . .?

Feigl's objection is, I think, based on the fact that he is impressed by the testimony of immediate experience. It is undoubtedly true that immediately experienced raw feels appear to have no relation of identity to any instrumentally observed BP. If this appearance turns out to be no more than that, i.e. if raw feels merely correspond to, but are not identical with, some BP, then perhaps we would have to regard raw feels as irreducible 'brute facts.' In consequence, the identity theory would have to be abandoned.

(4) In reply to the objection let us first agree that the testimony of immediate experience establishes the

---

12 Feigl, pp. 390-391.
existence of raw feels. The identity theory does not deny their existence. The point at issue is the correct analysis or interpretation of the testimony of immediate experience. It is agreed by both sides that we learn introspectively that we have raw feels, the dispute is about what raw feels are. Feigl, in his objection, holds that raw feels are irreducible brute facts that cannot be accommodated by a physicalistic identity theory, I hold that they can be.

How can this dispute be settled? It is clear that empirical evidence is no longer of any help. The conflict arises only after all the empirical evidence is available. The dispute is about the interpretation of the empirical evidence. A difference, as James said, to be a difference must make a difference. Let us therefore see what difference it makes which interpretation we adopt.

If, per impossibile, an identity theorist were to draw up a list containing all the known classes of events occurring in the universe, the list would contain 'n' items. If their opponents, who are being considered here, were to do the same, the list would contain at least 'n + 1' items; the additional item being raw feels. Some identity theorists deny, while their opponents assert, that raw feels are part of the 'basic furniture' of the universe.

What reason is there for supposing that either side is correct in this dispute? How do we settle questions about the possible existence of basic, unanalyzable entities in the universe? I shall discuss several possible lines of
argument in favor of dualism and Feigl's position. The view that raw feels are basic, unanalyzable parts of the universe could not be held on empirical grounds. Not only is there no empirical evidence for this view, there ex hypothesi could be none. The reason for this is that according to the initial assumption all the empirical evidence for the existence of raw feels is already available, viz., in the correlation statements whose truth is assumed; the conflict is about the interpretation of this evidence.

It might be argued that the testimony of immediate experience is evidence for the conclusion that raw feels cannot be given a reductive account. It has to be admitted, of course, that the properties predicated of raw feels on the basis of direct acquaintance appear to be quite different from the properties predicated of BP on the basis of instrumental observation. However, the fact that phenomenon (A) appears to have certain characteristics under certain conditions, and a phenomenon (B) appears to have quite different characteristics under different conditions licenses neither the conclusion that A and B are distinct nor that they are irreducible. The apparent difference between them may be due to the different observational conditions. ¹³ If dualists and Feigl wish to argue for the irreducibility of raw feels, they cannot by relying upon introspective evidence alone; additional arguments are required.

¹³This point was argued in Chapter Seven.
Another argument sometimes offered for the irreducibility of raw feels is the doctrine of the given. Feigl argues that epistemology, in order to provide an adequate reconstruction of the confirmation of knowledge claims must employ the notion of immediate experience as a confirmation basis; (the "given" cannot be entirely a myth).

Let us tentatively accept Feigl's argument. Its natural extension to the present case is the identification of the given with raw feels. The argument supporting Feigl's and the dualist's view of raw feels then becomes the following: through introspection we have evidence of the existence of raw feels. Raw feels are the given. The given is the confirmation basis of all of our empirical knowledge claims; hence its existence cannot be questioned. Nor, if raw feels are the given, can they be explained in terms of BP, for it is in terms of the given that everything else, including BP, is ultimately explained.

My criticism of this argument is that it conceals a confusion between two ways of regarding the given: the epistemological and the ontological. The view that there is an epistemological given may be justified, but it does not support the conclusion that raw feels are 'inexplicable.' The view that there is an ontological given is unjustified and rather well-refuted, hence it cannot be used to support the 'inexplicability' of raw feels. I shall not argue

\[\text{Feigl, p. 392.}\]
against there being an ontological given, I accept the criticisms of this view presented by Feyerabend, Popper, Quine, Sellars, Wittgenstein, and others. That is to say, I deny that if we penetrated 'deeply' enough into the nature of reality we would there find a level of existence in terms of which everything else that exists can be explained, but which itself cannot be explained.

On the other hand, I accept the need for the epistemological given. I take the claim that there must be an epistemological given to mean that there must be primitive observational terms in any scientific theory. These observational terms are, of course, inexplicable in terms of the theory in which they function as primitives. Their inexplicability, however, is not due to some peculiar ineffable property that the referents of the observational terms have; their inexplicability is due rather to the way in which the theory has come to be constructed. Other, alternative, constructions are possible in which what is given is different. In the light of this, one can accept Feigl's claim that raw feels are the given in a particular theory and hold, at the same time, that in another theory raw feels need not be the given, i.e. need not be the primitive observational terms. It is plausible to regard raw feels as being unanalyzable primitives in the common sense framework. Nothing that Feigl or dualists say, however, prevents one from attempting to construct a theory in which raw feels are not unanalyzable primitives. And this is precisely what the identity
theory is trying to do. It is unsatisfactory, therefore, to object to the identity theory by pointing out that in another theory raw feels play a different role.

(5) The dispute between the identity theorist and his critics now stands in the following way: according to the identity theory sensations and BP are theoretically identical. This assertion is criticized by saying that the theoretical identity of sensations and BP is compatible with their distinctness. That is, the truth of the correlation statement is sufficient for the fulfillment of the requirements of the theoretical identity; there is neither need nor justification for the assertion of the corresponding identity statement. The identity theorist meets the objection by arguing that if sensations and BP are to be regarded as distinct, even though they are acknowledged to be theoretically identical, there must be a difference between them. This difference, it is argued, against the identity theory, is that sensations do, while BP do not, have a raw feel component. Further, not only must it be true that only sensations have a raw feel component, but also that this raw feel component must be irreducible. The two arguments offered for the irreducibility of raw feels, viz., one based on the testimony of immediate experience, the other based on the doctrine of the given, were shown not to support the irreducibility of raw feels. The situation now is that both the identity theorist and his critics accept the testimony of immediate experience concerning the existence of raw feels. The
disagreement is about the interpretation of the testimony of immediate experience. I have been arguing up to now for the view that the interpretation offered by Feigl and dualists is mistaken. I shall argue that the interpretation offered by the identity theory should be accepted.

The success of an interpretation, or rather, of a theory, should be judged by examining whether or not the theory fulfills certain criteria. The difficulty is, however, that the criteria of success that the identity theory would accept are not the same as those that its critics would accept. It is thus possible that there is complete agreement between the identity theorist and his critic as to what the facts are, what precisely the theory is trying to do, how it does it, and yet there be disagreement as to whether the theory is successful.

In order to decide, in the last analysis, between the identity theory and competing theories of mind, one would have to decide which criteria of success should be accepted. Now I do not know how to justify my acceptance of certain criteria of success except by saying why I accept them. And saying this will ultimately come down to saying that they seem plausible to me. If this is less than a good justification, I can find consolation in two things. First, my primary concern is to show that given certain reasonable criteria of success, the identity theory is successful. So it is, at the very least, a competitor among theories of mind. Second, not only the identity theory, but all theories are
judged by criteria of success whose adoption is not justified any better than is the identity theory's.

I shall offer two criteria of success: simplicity and unity of explanation. These two criteria are not criteria of success of theories, they are rather criteria of success of a theory relatively to other theories. That is, I assume that any theory competing with the identity theory, and the identity theory itself, fulfills such requirements as explaining all the relevant facts, being internally consistent, and being, at least partially, testable. My argument is that given the fulfillment of these conditions, one should accept the theory that is simpler and gives a unitary explanation. Let us consider these criteria.

(6) The criterion of simplicity favoring the identity theory may be stated in this way. Given the truth of the correlation statement, we should accept the corresponding identity statement because it is simpler to suppose that there is only one process, rather than there being two processes. The argument for simplicity then takes the following form; as Kim\textsuperscript{15} puts it:

A correlation statement cries out for an explanation: Why is it that whenever and wherever there is water, there is \( \text{H}_2\text{O} \)? Why is it that wherever and only whenever a person has pain he is in some specific brain state? \ldots We can answer these questions if, and perhaps only if, we accept the corresponding identity statements. That is, we

\textsuperscript{15}Kim, p. 229.
shall answer: Because water is H₂O, because pain is brain state.

The question is what sort of simplicity is achieved by moving from the correlation statement to the corresponding identity statement?

Kim⁶ argues that "the simplicity thus achieved is rather trivial and of minimal significance from a scientific point of view." The reason for this conclusion is that:

The factual cash value of the identity is simply the correlation, and in terms of factual information we are simply repeating in the explanans what is supposed to be explained. This is a far cry from the usual kind of scientific explanation in which a fact or regularity is explained by involving more general and more comprehensive laws and theoretical principles.¹⁷

The replacement of the correlation statement by the identity statement does not reduce the number of primitive concepts, nor does it reduce the number of primitive assumptions. The replacement yields "neither economy of concepts, nor economy of assumptions."¹⁸

We have to argue, it seems to me, that the simplicity achieved by the identity theory is almost completely irrelevant from the scientific point of view. However, scientific simplicity is not the only kind of simplicity there is. I shall call the relevant kind of simplicity 'ontological.'

If we assume an ontological framework that has as

---

16Kim, p. 230.
17Kim, p. 230.
18Kim, p. 230.
basic entities particulars, e.g. persons, animals, material objects, and the like, and properties, e.g. solidity, brittleness, being opaque, and the like, then we can explain the identity of processes on the basis of identity of particulars and identity of properties. In such a framework the distinctness of processes would entail that there are different particulars and properties involved in the occurrence of processes.

Given this ontological framework, it would be ontologically simpler to assume that sensations and BP are identical, rather than distinct, processes. Their distinctness entails the existence of a greater number of particulars and properties than does their identity. Therefore, if all other factors are equal, the identity theory is to be preferred, because it is ontologically simpler. The simplicity of the identity theory is based on its postulation of the existence of a smaller number of particulars and properties than do competing theories.

There are at least two different ways in which dualistic theories could conflict with the identity theory. The identity theory assumes that there is only one kind of particular, viz., the body or the brain, and only one kind of property, viz., physical, involved in giving a psychological description of a person. Radical dualism disagrees: according to this view there are two different kinds of particulars, viz., the mind and the body, and two different kinds of properties, viz., mental and physical, involved in giving a
psychological description of a person. A more moderate dualistic theory may agree that there is only one kind of particular, yet hold that two different sets of properties, mental and physical, should be predicated of the one particular. The identity theory is simpler, from the ontological point of view, than is either radical or moderate dualism.

A dualist may agree with everything I said, but may still disagree with my conclusion. It might be asked: why should one prefer ontological simplicity to ontological complexity? In reply to this, it seems to me, one can only say: if the assumption that there exists only one kind of particular with only one kind of property will explain all the facts, then the assumption that there exists two kinds of particulars or two sets of properties must be vacuous. One can conclude absurdity from vacuity, however, only if a narrow verificationist view is taken. The inescapable conclusion seems to be that while dualists may have a valid claim, vis-a-vis ontological simplicity, the claim has not been made out. In order to make it out, dualists have to specify what mind and mental properties are, apart from the brain or the body and their physical properties.

(7) In this section I shall discuss the second criterion whose fulfillment should lead us to prefer the identity theory vis-a-vis other theories of mind. The criterion is that of giving a unitary explanation of all explananda. If this criterion is accepted, then the identity theory is to
be preferred to other theories, because it fits into and contributes to a unitary explanatory framework.

Smart\(^{19}\) explains his view of philosophy partly as an attempt to fulfill this criterion:

As I propose to use the word 'philosophy' it will stand primarily for an attempt to think clearly and comprehensively about: (a) the nature of the universe, and (b) the principles of conduct. . . . Notice that I have said both 'to think clearly' and 'to think comprehensively.' The former expression ties up with the prevailing conception of philosophy as linguistic or conceptual analysis, and the latter ties up with another conception of philosophy as the rational reconstruction of language so as to provide a medium for the expression of total science.\(^{20}\)

Smart then goes on to explain what he means by 'thinking comprehensively':

A man might analyse biology in a certain way. He might argue, as I shall do, that living organisms, including human beings, are simply very complicated physico-chemical mechanisms. This man might also analyse physics as the ordering and predicting of sense experiences. . . . But though he might be thinking clearly, he would not be thinking comprehensively.\(^{21}\)

As a biologist, this man would postulate one kind of relation between man and nature; as a physicist he would postulate another. Thinking as a biologist he would say that man is part of nature; thinking as a physicist he would say that what nature is thought to be depends on the experiences of

\(^{19}\)Smart, PSR., Chapter One.

\(^{20}\)Smart, pp. 1-2.

\(^{21}\)Smart, p. 2.
man. Smart writes: 22

To think comprehensively he would have to discover a way of thought which enabled him to think both as a biologist and as a physicist. Presumably a comprehensive way of thought would be one which brought all intellectual disciplines into a harmonious relationship with one another.

But more needs to be said. Identity theorists as well as their critics may have claim to comprehensive thinking. What needs to be shown is that the identity theory, as opposed to other theories, supports a particular kind of comprehensive thinking.

This requirement is fulfilled if one accepts the criterion of unitary explanation in addition to comprehensiveness. That is, this criterion is fulfilled if the same type of explanation is provided for all explananda. The same type of explanation would be given if the program of the unity of science were possible. That is, if all explananda fitted into one science or another, and if all sciences were reducible to one fundamental science, then all explananda would be given a unitary, viz., scientific, explanation.

The program of the unity of science may be conceived in several different ways: 23

First, Unity of Science in the weakest sense is attained to the extent to which all terms of science are reduced to the terms of one discipline, (e.g. physics or psychology).

22 Smart, p. 2.

This conception of the unity of science is based on the unity of language:

Second, Unity of Science in a stronger sense (because it implies Unity of Language, whereas the reverse is not the case) is represented by Unity of Laws. It is attained to the extent to which the laws of science become reduced to the laws of some one discipline.

So the program of the unity of science, or the achievement of a unitary explanatory framework, depends on the unity of language or on the unity of laws.

Now my point is that if one accepts the desirability of a unitary explanatory framework, and if there are two or more competing theories that differ only in respect of one thing, viz., whether or not they fit into a unitary explanatory framework, then one should accept the theory that does fit into the framework.

It seems to me that of the identity theory and competing dualistic theories only the identity theory fulfills this requirement. For while dualistic theories may provide comprehensive explanation, they do not provide unitary ones. The reason for this is that dualistic theories assume the existence of either both mental particulars and mental properties, or only of mental properties. It is essential for dualistic theories that mental particulars and properties be irreducible to physical particulars and properties. So if there is a comprehensive dualistic explanation it cannot be unitary, since the explanation of mental particulars and properties must differ in kind from the explanation of physical particulars and properties.
So, if the only difference between the identity theory and its dualistic competitors were the giving of unitary explanations, then given this criterion, the identity theory should be accepted.

II

(8) The term 'The Criteriological View' was invented, as far as I know, by Chappell. He uses it to refer collectively to certain views of Wittgenstein, Malcolm, and Strawson; Shoemaker, I think, could be safely added to the list. The basic assumption of the criteriological view is that there is a criterial relation between mental states and behaviour, and simpliciter between sensations and behaviour. Chappell writes:

This relation is a logical relation: it is not just a fact that certain things are said and done by a man who has a pain. The relation is logical, because as Malcolm puts it . . . "the satisfaction of the criterion of establishes the existence of beyond question," or because to use Strawson's phrase . . . the presence or satisfaction of a criterion constitutes a "logically adequate basis" for asserting the existence of the thing of which it is a criterion,


26Chappell, p. 17.
or for ascribing that thing (if that is the sort of thing it is) to something. Behavior is a criterion of pain, according to those who hold this view, because an appeal to a man's behavior establishes beyond question that he has pain, because if he is behaving in a certain way he must have a pain.

The criteriological view may be understood as an attempt to explain what justification there is for describing others and ourselves as having certain experiences. The justification is that the presence of criteria for the experience in question logically entitles one to describe a person as having the experience.

Strawson argues that there is an ambiguity of purpose in Wittgenstein's version of the criteriological view. Strawson writes:

> Wittgenstein seems to me to oscillate in his discussion of this subject between a stronger and a weaker thesis, of which the first is false and the second true. These may be described, rather than formulated, as follows. . . . The stronger thesis says that no words name sensations (or 'private experiences'); and in particular the word 'pain' does not (cf. 293). The weaker thesis says that certain conditions must be satisfied for the existence of a common language in which sensations are ascribed to those who have them; and that certain confusions about sensations arise from a failure to appreciate this, and consequently to appreciate the way in which the language of sensations functions.

The strong thesis is reminiscent of behaviourism; sensation-predicates, according to this view, do not refer to mental . . .

---

28Strawson, pp. 83-84.
states, they refer rather to actual and potential behaviour. I agree with Strawson that the strong thesis is false, but I shall not argue for this point.

The purpose of this part of the chapter is to shed further light on the identity theory by showing that it is compatible with the weaker thesis. Indeed, the identity theory may be regarded as an attempt to go beyond the weaker criteriological view.

(9) Strawson argues that in order to ascribe sensations to others it is necessary to have a common sensation-language; he then goes on to say:

The only possible common pain-language is the language in which pain is ascribed to those who talk the language, the criteria for its ascription being (mainly) pain-behaviour. And because of this fact it is necessarily empty and pointless (I will not say meaningless) either (a) to speculate about the ascription of pain to anything which does not exhibit behaviour comparable in relevant respects with human behaviour, or (b) to raise generalized doubts about other people's experience of pain, or about one's own knowledge of this. It is the above points which I take Wittgenstein essentially to be making.

That is, the weak criteriological view holds that it is a necessary condition of having knowledge of other minds that there be a common sensation-language. Sensations are ascribed to others in this common language on the basis of criteria, and the criteria are the many different kinds of pain behaviour.

Strawson, p. 88.
Strawson agrees with Wittgenstein up to this point. Strawson, however, claims that the way in which Wittgenstein argues is misleading: 30

For from none of those facts does it follow that 'pain' is not the name of a sensation. On the contrary. It is only in the light of the fact that 'pain' is the name of a sensation that those facts are intelligible; or better, to say that 'pain' is the name of a sensation is (or ought to be) just to begin to draw attention to those facts. One could say: that pain is a sensation (or, that sensations have the special status they have) is a fact of nature which dictates the logic of 'pain.'

Strawson's interpretation of Wittgenstein points, it seems to me, to the following conclusion. The weak criteriological view is engaged in an investigation of what Strawson calls "the logic of 'pain.'" I take this to be an attempt to discover the conditions in which 'pain' is used correctly, to discover what are the criteria for the correct use of 'pain,' to investigate the "depth grammar" of 'pain' and other sensation-predicates. In other words, the weak criteriological view attempts to clarify the way in which sensation predicates are used in common, ordinary language.

However, 'pain' and other sensation-predicates are terms not only in the vocabulary of ordinary language, but also in the vocabulary of scientific languages, e.g. the language of neurophysiology. Neurophysiologists may be taken to be investigating partly that "fact of nature which," according to Strawson, "dictates the logic of pain."

30 Strawson, p. 88.
The identity theory may be regarded as an attempt to bring together the philosophical analysis of the use of sensation-predicates in ordinary language and the scientific analysis of the phenomenon of sensation. If the identity theory is successful in combining the philosophical and the scientific analyses of sensations, then it will have taken a useful step in the direction of achieving unity of explanation. This will result, we may hope, in the eventual achievement of the goal that Smart described as bringing "all intellectual disciplines into a harmonious relationship with one another." 31

(10) Some of Shoemaker's 32 arguments can be used in support of my claim that the identity theory performs an important task in bringing together philosophical and scientific analyses. Shoemaker 33 argues that corrigible first-person psychological statements have both a necessary and a contingent component. 34 The necessary component of these statements is that, according to Shoemaker, it is a necessary

31 Smart, p. 2.
33 Shoemaker, pp. 215-239.
34 Shoemaker denies that all first-person psychological statements are corrigible. I argued against this claim in Chapters Four and Five. Here I shall assume that most of those statements that Shoemaker regards as incorrigible are, in fact, corrigible.
truth that corrigible first-person psychological statements are generally true. Or, to put it differently, if corrigible first-person psychological statements are made at all, most sincere and confident statements are true. Shoemaker argues:

It is a necessary truth that if a group of human beings . . . make [corrigible first-person] statements, there will exist correlations [(e.g.)] . . . between the uttering of certain sounds (or the making of certain gestures) and its being the case that the speaker's eyes are open and directed toward an object of a certain kind. . . . But that such correlations do exist in the behaviour of any group of human beings will be a contingent fact.

The necessary component in first-person statements is due to the criterial relation that is asserted to hold between the ascription of the experience to the person and the person's behaviour. Shoemaker holds that the existence of this necessary truth requires no explanation. The reason for this is, I suppose, that Shoemaker holds that if there are human beings to speak a language, then they necessarily require criteria.

On the other hand, the existence of contingent correlations of the sort discussed above is something for which an explanation may be demanded. But the explanation of the fact that these correlations exist is hardly to be found in any epistemological theory . . . what is to be explained is an empirical phenomenon . . . and it is to the natural sciences (perhaps to physiology), not to epistemology that we must turn if we want an explanation of it.36

There is much in this passage with which an identity

35Shoemaker, p. 238.
36Shoemaker, p. 239.
theorist could agree. I argued that the identity theory has a philosophical and an empirical component. The empirical component, it will be remembered, is a statement asserting that there is a perfect correlation between the process of having a sensation and the occurrence of EP. The philosophical component is made up: first, of the clarification of how sensations could be identified independently, and second, of a justification for arguing from the truth of the correlation statement to the assertion of the identity statement.

The identity theory and Shoemaker are in complete agreement about the first philosophical component. That is, the identity theory and Shoemaker (or, rather the weak criteriological view) agree that in order to give any sort of analysis of sensations it is necessary to identify them independently. There is the further agreement that sensations are identified independently through behaviour; and that what makes the identification possible is the existence of criterial and symptomatic relations between sensations and behaviour.

There is also a measure of agreement between the identity theory and the weak criteriological view concerning the empirical component of the identity theory. That is, while Shoemaker claims that the correlation is purely a scientific matter, the identity theory holds that it is

---

37Chapter Three, Section (1).
largely a scientific matter. There are philosophical and logical considerations involved in establishing the correlation. It is not a matter of indifference to a philosophical analysis of sensations whether or not there is a perfect correlation between the criteria of sensations and sensations; nor is it a philosophically neutral question whether or not the correlation can be tested by considering the spatial location of the correlated items.

The disagreement between the identity theory and Shoemaker's version of the weak criteriological view arises in connection with the second philosophical component of the identity theory, viz., the justification for passing from the correlation statement to the identity statement. Shoemaker's attempt at separating the necessary and the contingent components of first-person corrigible psychological statement, and his insistence upon the philosophical relevance of the necessary component only, stems from a view of the task of philosophy that the identity theory does not share.

The disagreement may be expressed in terms of the distinction between two conceptual frameworks. One is the everyday, common sense framework populated by persons, animals, plants, and by middle-sized material objects, i.e. rocks and houses, pebbles and mountains, but not by electrons and galaxies. These entities are observable, and normally are observed, through the unaided senses. The language of the common sense framework is ordinary language. The other
framework is the framework of science. The entities of this framework are normally not observable by the unaided senses, but only by instruments, if indeed they are observable at all. Molecules, electrons, kinetic energy, bacteria, quasars, and the like, constitute the field of investigation of the scientific framework. The language of the scientific framework is sophisticated, precise, quantifiable, technical language.

Now it seems to me that Shoemaker and the weak criteriological view regard the common sense framework and ordinary language as the proper field of philosophical analysis. The identity theory, however, considers the proper task of philosophy to be the investigation of the relation between the common sense and the scientific frameworks, as well as an examination of the common sense framework. This requires, of course, some understanding of the scientific framework. Insofar as the identity theory concerns itself with sensations as they appear in the common sense framework, there is no disparity of purpose between it and the weak criteriological view. However, when the identity theory leaves the common sense framework and attempts to see what bearing scientific investigations of sensations have on the ordinary, common sense view of sensations, it parts company with the weak criteriological view.

It seems to me that there is a need for bridging the ever-widening gap between the scientific world-view and the
sophisticated common sense view of the world. Ryle argues the same way; he writes:

We often worry ourselves about the relations between what we call 'the world of science' and 'the world of common sense'. When we are in a certain intellectual mood, we seem to find clashes between the things that scientists tell us about our furniture, clothes and limbs and the things we tell about them. We are apt to express these felt rivalries by saying that the world whose parts and members are described by scientists is different from the world whose parts and members we describe ourselves, and yet, since there can be only one world, one of these seeming worlds must be a dummy-world. Moreover, as no one nowadays is hardy enough to say 'Bo' to science, it must be the world that we ourselves describe which is the dummy-world.

The identity theory may be regarded as an attempt to show that there are not two worlds. It is the case, rather, that we describe our one world in two different ways. When we regard the world as scientists, we express ourselves in a terminology suitable for the purpose of science. When we express ourselves during the routines of daily life, we use a different terminology, because our purposes in expressing ourselves in day to day living are different from our purposes as scientists.

Part of the purpose of the identity theory is to show that when we talk about sensations in ordinary language, we talk about the same things as we do when we talk about

---


39Ryle, p. 68.
sensations in scientific language. Neither the world of
science nor the world of common sense is a "dummy-world."
They are the results of different ways of observing and
thinking about one and the same world.
BIBLIOGRAPHY

(All and only those works are included that are mentioned in the text.)

Anscombe, G. E. M.  
Intention  

Armstrong, D. M.  
Bodily Sensations  
Studies in Philosophical Psychology  

Aune, B.  
'Is Introspective Knowledge Incorrigible?'  

Ayer, A. J.  
'Privacy'  
The Concept of a Person  

Baier, K.  
'Smart on Sensations'  

Bradley, M. C.  
'Critical Notice--Philosophy and Scientific Realism'  

Buck, R. C.  
'Non-Other Minds'  
Analytical Philosophy, First Series, ed. R. J. Butler  

Campbell, C. A.  
On Selfhood and Godhood  
The Muirhead Library of Philosophy  
Coburn, R. C.  
'Sheffer on the Identity of Mental States and Brain Processes'  
J. of Phil., 60 (1963), pp. 89-92.

Cornman, J. W.  
'Pains and Space'  

Erwin, F. R. and Sternbach, R. A.  
'The Identity of Mind and Body'  

Farrell, B. A.  
'Hereditary Insensitivity to Pain'  

Feigl, H.  
'Experience'  

Feigl, H.  
'Physicalism, Unity of Science and the Foundations of Psychology'  
The Philosophy of Rudolf Carnap, ed. P. A. Schilpp  

Feyerabend, P. K.  
'Mind-Body, Not a Pseudoproblem'  

French, J. D.  
'Materialism and the Mind-Body Problem'  
Rev. of Meta., 17 (1963), pp. 49-66.

Garnett, A. C.  
'The Reticular Formation'  

Gasking, D.  
'Avowals'  

Hebb, D. O.  


<table>
<thead>
<tr>
<th>Author(s)</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oppenheim, P. and Putnam, H.</td>
<td>'Unity of Science as a Working Hypothesis'</td>
</tr>
<tr>
<td></td>
<td><em>Minnesota Studies in the Philosophy of Science</em></td>
</tr>
<tr>
<td></td>
<td>Minneapolis: Univ. of Minnesota Press, 1955.</td>
</tr>
<tr>
<td>Penfield, W. and Jasper, H.</td>
<td>'Functional Localization in the Cerebral Cortex'</td>
</tr>
<tr>
<td></td>
<td><em>Epilepsy and the Functional Autonomy of the Human Brain</em></td>
</tr>
<tr>
<td>Pitcher, G.</td>
<td>'Sensations and Brain Processes: A Reply to Prof. Smart'</td>
</tr>
<tr>
<td>Place, U. T.</td>
<td>'Is Consciousness a Brain Process?'</td>
</tr>
<tr>
<td></td>
<td>'Materialism as a Scientific Hypothesis'</td>
</tr>
<tr>
<td>Pratt, C. C.</td>
<td>The Logic of Modern Psychology</td>
</tr>
<tr>
<td>Quine, W. V. O.</td>
<td>Methods of Logic</td>
</tr>
<tr>
<td>Rescher, N. and Oppenheim, P.</td>
<td>'Logical Analysis of Gestalt Concepts'</td>
</tr>
<tr>
<td>Routley, R. and Macrea, V.</td>
<td>'On the Identity of Sensations and Physiological Processes'</td>
</tr>
<tr>
<td>Ryle, G.</td>
<td>The Concept of Mind</td>
</tr>
<tr>
<td></td>
<td>Dilemmas</td>
</tr>
<tr>
<td>Sellars, W.</td>
<td>'The Identity Approach to the Mind-Body Problem'</td>
</tr>
</tbody>
</table>
Shaffer, J.  
"Could Mental States Be Brain Processes?"  

Shoemaker, S.  
Self-Knowledge and Self-Identity  

Skinner, B. F.  
Verbal Behavior  

Smart, J. J. C.  
"Sensations and Brain Processes"  

Smart, J. J. C.  
"Sensations and Brain Processes: A Rejoinder to Dr. Pitcher and Mr. Joske"  

Smart, J. J. C.  
"Further Remarks on Sensations and Brain Processes"  

Smart, J. J. C.  
"Brain Processes and Incorrigibility--Reply to Baier"  

Smart, J. J. C.  
Philosophy and Scientific Realism  
International Library of Philosophy and Scientific Method  

Smart, J. J. C.  
"Materialism"  
J. of Phil., 60 (1963), pp. 651-652.

Smart, J. J. C.  
"The Identity Thesis--A Reply to Prof. Garnett"  
A. J. P., 43 (1965), pp. 82-83.

Spence, K. W.  
"The Emphasis on Basic Function"  
Psychological Theory, ed. M. H. Marx  

Sternbach, R. A.  
"Congenital Insensitivity to Pain"  
Psychological Bulletin, 60 (1963), pp. 252-264.


