MIND-BODY IDENTITY AND IRREDUCIBLE PROPERTIES

Neil Lubow

July 1, 1977

ABSTRACT

The identity theory, advocated as a solution to the mind-body problem by materialists such as Feigl and Smart, has been criticized for implying the existence of irreducible properties (i.e. properties incompatible with materialism). After summarizing the relevant theses of materialism, I consider several versions of the irreducible properties objection, and argue that they are all unsuccessful.

Mind-Body Identity and Irreducible Properties

I: The Irreducible Properties Objection

Most objections to the identity theory have the following form: mental states have (or lack) property Ø; brain states lack (or have) property Ø; things are identical just in case they share all the same properties; therefore mental states are not identical with brain states. The objection I shall consider in this paper does not fit this familiar pattern. Instead, this objection contends that if the identity theory is true, there are irreducible properties, i.e., properties which are inconsistent with materialism.

Despite the way it has been presented in the literature, the irreducible properties objection is not an objection to the identity theory per se. I take the identity theory to be the thesis that (at least some) mental states (or processes or events) are identical with brain states (or processes or events). The identity theory has been proposed by such philosophers as J. J. C. Smart and Herbert Feigl as an adjunct to the more basic theses of materialism or physicalism (I shall use 'physicalism' and 'materialism' synonymously). The irreducible properties objection essentially claims that the identity theory and materialism are incompatible--not that the identity theory is wrong. Proponents of this objection argue that the identity theory does not support materialism; instead, it leads to the conclusion that materialism is false.

Smart himself states the irreducible properties objection, first in "Sensations and Brain Processes,"[1] and again

[1] J. J. C. Smart, "Sensations and Brain Processes," in The Philosophy of Mind, ed. V. C. Chappell (Englewood Cliffs, N. J.: Prentice Hall, 1962), pp. 160-72. Hereafter cited as "S & BP."

in Philosophy and Scientific Realism.[2] Many critics, following Smart, have used this objection in one form or another to attack the identity theory. In fact, Smart's inability to adequately reply to these attacks at one time caused him to give up the identity theory, although he still maintained his materialism.[3]

According to Smart's first statement of this objection, the contingent nature of the psycho-physical identity statements shows that

> the qualities of sensations are something over and above the qualities of brain processes. That is, it may be possible to get out of asserting the existence of irreducibly psychic processes, but not out of asserting the existence of irreducibly psychic properties. For suppose we identify the Morning Star with the Evening Star. Then there must be some properties which logically imply that of being the Morning Star, and quite distinct properties which entail that of being the Evening Star. Again, there must be some properties ... which are logically distinct from those in the physicalist story.[4]

Smart again states this objection, in a slightly more intelligible form, in P & SR. He writes:

> It may be objected that even if I can get out of saying that sensations or experiences are entities over and above brain processes I can do so only at the cost of

[2]J. J. C. Smart, Philosophy and Scientific Realism (London: Routledge & Kegan Paul, 1963), pp. 94-96. Hereafter cited as P & SR.

[3]J. J. C. Smart, "Comments on the Papers," The Identity Theory of Mind, ed. C. F. Presley (St. Lucia, Queensland: University of Queensland Press, 1967), p. 91. Smart has once again become a full-fledged identity theorist. He attributes his temporary surrender to mental exhaustion; see Smart, "Further Thoughts on the Identity Theory," The Monist, 56 (April, 1972), 149-62, esp. n. 3.

[4]Smart, "S & BP," p. 166.

> admitting 'emergent' or irreducibly non-physical properties of brain processes. If this objection can be substantiated it is a very serious one, for, as we saw in Chapter IV, emergent properties would be just as much objectionable nomological danglers as psychical entities are. I think I can rebut the objection, and I wish to make it quite clear that I wish to rebut it ... Now it may be said that if we identify an experience and a brain process and if this identification is, as I hold it is, a contingent or factual one, then the experience must be identified as having some property not logically deducible from the properties whereby we identify the brain process. To return to our analogy of the contingent identification of the author of Waverley with the author of Ivanhoe. If the property of being the author of Waverley is the analogue of the neurophysiological properties of a brain process, what is the analogue of the property of being the author of Ivanhoe? There is an inclination to say: 'an irreducible, emergent, introspectible property.'[5]

It is evident that materialism is incompatible with there being such properties. Smart says that such properties would be outside "the physicalist framework,"[6] "outside the physicalist picture,"[7] and "logically distinct from those in the physicalist story."[8] Properties which are "irreducible," "emergent," or "purely phenomenal or introspectible"[9] are outside the physicalist framework. The existence of these properties is incompatible with materialism in the sense that if something actually has such a property, then materialism is false. Properties of mental states (psychic properties) are not the only properties purported to be outside the physicalist framework. In Chapter 4 of

[5]Page 94.

[6]"S & BP," p. 166.

[7]"S & BP," p. 161.

[8]"S & BP," p. 166.

[9]"S & BP," p. 166.

P & SR, Smart attempts to show that secondary qualities, such as colors, are not outside the physicalist framework.

It is not immediately clear from the above passages (or from anything in Smart's writings) exactly what the irreducible properties objection is. Consequently I will try to provide a detailed and comprehensive account of this objection which will include statements of the objection taken from the writings of its major proponents.

II: Materialism

In order to clearly state and properly evaluate the irreducible properties objection, it will be necessary to examine in detail certain aspects of materialism, especially those parts which are purportedly incompatible with the identity theory. Many different philosophical views have been called 'materialism' or 'physicalism', and it is not my intention to consider every position that has been characterized in this way. The only versions of materialism relevant here are those held by modern identity theorists, particularly Smart and Feigl; and it is only a portion of these philosophers' materialist views which will be considered here. For example, both Smart and Feigl are realists about theoretical entities, but the issues of realism vs. instrumentalism are not involved in the irreducible properties objection, and consequently this part of the materialists' position will be ignored.

Feigl's Materialism

The basic tenets of Feigl's materialist views are summarized in the two theses of physicalism. "The first thesis of physicalism or the thesis of the unity of the language of science is essentially the proposal of a criterion of scientific meaningfulness in terms of intersubjective confirmability."[10] According to this first thesis,

[10]Herbert Feigl, "Physicalism, Unity of Science and the Foundations of Psychology," in The Philosophy of Rudolph Carnap, ed. P. A. Schilpp (La Salle, Ill.: Open Court, 1963), p. 227.

a factual statement is scientifically meaningful if and only if it is, in principle, capable of intersubjective test and confirmation. A sentence is intersubjectively confirmable just in case the state of affairs it describes can be known about via the senses. Any phenomenon which is detectable by the senses, either directly (e.g. it can be seen) or indirectly (e.g. it is causally related to some phenomenon which can be seen) is intersubjectively knowable.

The demand that the factual statements of science be intersubjectively confirmable rests

> on the belief that there is nothing in heaven or on earth (or even beyond both) that could not possibly be known, i.e., there are no assertions about reality which could not conceivably be confirmed or disconfirmed on the basis of sense perception .... whatever there is in any shape or form, things, events, states, anywhere, at any time, "inorganic," "organic," "mental," "social," etc., can be causally related--even if only very indirectly by complicated chains--to the sense organs of human organisms.[11]

Thus Feigl believes that the intersubjective language is capable of describing everything which exists in the world, without leaving anything out, because all the phenomena in the world are causally interrelated. The universal or statistical laws which describe these causal interrelations make up what Feigl calls the 'nomological net.' If there were some phenomenon which was causally inefficacious in the physical realm (for brevity, simply 'inefficacious'), it could not be the subject of scientific study. It would not be describable in an intersubjective language, nor could it be known intersubjectively. The first thesis of physicalism denies that there are any such phenomena.

The second thesis of physicalism asserts that there will be no emergence or irreducibility in the biological, psychological, and social sciences. "This second thesis of physicalism claims that the facts and laws of the natural and the social sciences can all be derived--at least in principle--

[11]Feigl, "Physicalism," p. 239.

from the theoretical assumptions of physics. We may formulate this second thesis as the belief in the possibility of a unitary explanatory system."[12] The second thesis is stronger than the first, and is more easily falsified.

The two theses of physicalism correspond to two senses of 'physical' which Feigl explicates in "The 'Mental' and the 'Physical'," [13] his major work on the identity theory.

Feigl defines the first sense of 'physical' as follows:

> By "physical$_1$ terms" I mean all (empirical) terms whose specification of meaning essentially involves logical (necessary or, more usually, probabilistic) connections with the intersubjective observation language, as well as the terms of this observation language itself. Theoretical concepts in physics, biology, psychology, and the social sciences hence are all--at least--physical$_1$ concepts.[14]

The notion of physical$_1$ terms is intended to complement and help explain the first thesis of physicalism. Both the first thesis of physicalism and the notion of physical$_1$ are meant to characterize the empirical (in contrast to the mathematical or logical) language of science. Feigl is here working in the tradition of logical positivism and its now abandoned verification principle of meaning.[15]

[12]Feigl, "Physicalism," pp. 227-28.

[13]Herbert Feigl, "The 'Mental' and the 'Physical'," in Minnesota Studies in the Philosophy of Science: Concepts, Theories, and the Mind-Body Problem, II, ed. H. Feigl, M. Scriven, and G. Maxwell. (Minneapolis: University of Minnesota Press, 1958), pp. 370-497. Hereafter cited as "M & P."

[14]Feigl, "M & P," p. 424.

[15]See Carl Hempel, "Empiricist Criteria of Cognitive Significance: Problems and Changes," in Aspects of Scientific Explanation and Other Essays in the Philosophy of Science (New York: The Free Press, 1965), pp. 101-19.

The first thesis of physicalism provides a criterion for determining which sentences are scientifically meaningful. The definition of 'physical$_1$' provides a criterion for determining which empirical terms are scientifically meaningful. However the relationship between the two criteria is not obvious, for the former criterion is described in terms of intersubjective confirmability, while the latter criterion is described in terms of "connections" with the "intersubjective observation language".

The first thesis of physicalism states that factual sentences are scientifically meaningful if and only if they are intersubjectively confirmable. The test of intersubjective confirmability is suitable for sentences, but it is not directly applicable to terms. For terms are not true or false, and therefore they are neither confirmable nor unconfirmable. Consequently Feigl introduces the notion of physical$_1$ terms in order to characterize as scientifically meaningful the empirical terms used in intersubjectively confirmable sentences.

Although the criterion of intersubjective confirmability is relatively clear, the criterion described in the definition of 'physical$_1$' is considerably more mysterious. To begin with, Feigl is not very specific as to what constitutes the intersubjective observation language. Apparently he has in mind the kind of language used in every day life and science to describe objects, properties, and relations which are more or less directly observable by the senses. Thus, such expressions as 'tree', 'is taller than', and 'red', seem to be in the observation language. Feigl clearly intends to exclude from the observation language such terms as 'magnetic field', 'electron', 'pain', etc.

However, the vagueness of the definition of 'physical$_1$' is not entirely due to the notion of an observation language (whatever it may be). According to the definition, an empirical term which is not itself in the observation language is a physical$_1$ term if and only if its "specification of meaning" involves "logical (necessary or, more usually, probabilistic) connections" with the observation language. Based on this remark it is difficult to determine just what the connection is supposed to be between the meaning of physical$_1$ terms and the observation language. Nevertheless,

this much is certain: Feigl is not demanding that all physical$_1$ terms be explicitly definable using only observation terms. For example 'magnetic field' is certainly a physical$_1$ term, yet Feigl believes that it is not explicitly definable in observation terms.[16] This is because 'magnetic field' is a theoretical term, and Feigl says that all such terms are "not explicitly definable on the basis of observation terms ... [but instead are] specified by postulates and by correspondence rules relating them to the terms of the observation language".[17]

Feigl's current view of theoretical terms is worth dwelling upon here, for it will help explain the connection between physical$_1$ terms and the observation language. For a detailed discussion of the meaning of theoretical terms, Feigl refers his readers to an article by Carnap with which Feigl is in agreement.[18] In this article, Carnap also contends that theoretical terms are not explicitly definable using only observation terms. This means that no logically necessary and sufficient condition, described in observation terms, can be provided for the use of theoretical terms. Indeed, the connection between theoretical terms and observation terms is much weaker, for Carnap does not even require that there be an empirically sufficient condition (described in the observation vocabulary) for the use of theoretical terms. Since theoretical terms are clearly physical$_1$ terms, it follows that Feigl does not require, by his definition of 'physical$_1$', that there be even an empirically sufficient observational condition for every physical$_1$ term.

Having ruled out explicit definability, what then is the connection between physical$_1$ terms and the observation language? Feigl, unfortunately, provides no clear and precise answer to this question. Nevertheless, from his

[16]"M & P," p. 425.

[17]"M & P," p. 425.

[18]"M & P," pp. 394, 425. The Carnap article referred to is "The Methodological Character of Theoretical Concepts," in Minnesota Studies, I, pp. 38-76.

writings on the subject the following somewhat vague picture emerges: if Ø is an empirical term, not itself in the observation language, then Ø is physical$_1$ if and only if the meaning of Ø is such that some observation statements provide good reason for correctly using Ø. This falls far short of a logically (or empirically) sufficient condition for the correct application of Ø. In some cases, observation statements at best give only a justification for the probabilistic application of Ø. For example if Ø is some predicate '$\underline{F}$ _', observation statements might only provide good reason to say that there is such-and-such a probability that $\underline{Fa}$.

Physical$_1$ terms then, are any empirical terms in the total intersubjective language. In addition "the sort of objects or processes which can be described (and possibly explained or predicted) in the concepts of a language with an intersubjective observation basis"[19] are also characterized as physical$_1$.

The second sense of 'physical' is as follows: "By 'physical$_2$' I mean the kind of theoretical concepts (and statements) which are sufficient for the $\underline{\text{explanation}}$, i.e., the deductive or probabilistic derivation, of the observation statements regarding the inorganic (lifeless) domain of nature".[20] Physical$_2$ terms are the theoretical terms of a utopian physics-chemistry whose theory can explain all inorganic phenomena.

Using the notions of physical$_1$ and physical$_2$, the first thesis of physicalism asserts that everything in the world can be described in physical$_1$ terms. The second thesis of physicalism claims that all the phenomena in the world can be explained by theories which contain only physical$_2$ theoretical terms.

Nomological Danglers

In "The 'Mental' and the 'Physical'" Feigl introduced

[19]"M & P," p. 421.

[20]"M & P," p. 424.

the phrase 'nomological dangler.' Subsequently this phrase has been used (and misused) by many authors in their writings about materialism and the identity theory. Therefore it is important to explain just what a nomological dangler is and to describe the context in which the notion is introduced.

Feigl claims that there are empirically testable differences among various theories concerning the relationship between mind and body. For example, let us assume that there are in fact one-one correlations between the occurrences of mental states (Ψs) and brain states (Øs), and in addition, assume that $physical_2$ determinism (predictability) holds for the Ø series. Given these assumptions, Feigl argues that mind-body interactionism is ruled out, since it would involve a breach in $physical_2$ determinism. However, epiphenomenalism and the identity theory are each consistent with these empirical assumptions. Epiphenomenalism claims that the Ψs are not causally efficacious. If mental states are causally inefficacious, then they are not $physical_1$, i.e., their existence is not intersubjectively confirmable. In discussing the Ψ - Ø correlation laws, (while hypothesizing epiphenomenalism is true) Feigl says,

> these correlation laws are utterly different from any other laws of ($physical_2$) science in that, first, they are nomological "danglers," i.e., relations which connect intersubjectively confirmable events with events which ex hypothesi are in principle not intersubjectively and independently confirmable ... And second, these correlation laws would, unlike other correlation laws in the natural sciences, be (again ex hypothesi) absolutely underivable from the premises of even the most inclusive and enriched set of postulates of any future theoretical physics or biology.[21]

From the above passage, we can see that a nomological dangler is a law which connects events that are causally efficacious (hence intersubjectively confirmable) with events that are causally inefficacious (non-intersubjectively confirmable). If epiphenomenalism were true, then

[21]"M & P," p. 428.

the Ψ - Ø correlation laws would be nomological danglers. In addition, Feigl says they will not be reducible (derivable) from even a utopian physics or biology. All nomological danglers will be irreducible[22] to physics, but not all irreducible laws will be nomological danglers.

In "Sensations and Brain Processes," Smart applies the phrase 'nomological dangler' to sensations, instead of to the laws which relate sensations to brain processes. Although Smart corrects this particular mistake in later publications, he then often leaves the impression that any irreducible or unexplainable law is a nomological dangler. This is, however, also a mistake, since only laws which relate the intersubjectively confirmable with the non-intersubjectively confirmable are nomological danglers.

J. T. Stevenson, following Smart, also uses the phrase 'nomological danglers' to describe sensations rather than the laws correlating sensations with brain processes. Stevenson tries to criticize the identity theory by attempting to show that even if the identity theory were true, sensations would still be danglers, and thus still be objectionable to materialists. Stevenson writes:

> sensations were nomological danglers in virtue of certain properties which they had, and we have in no way eliminated these properties. In fact, by insisting (1) that 'sensation' is not synonymous with 'brain process', (2) that sensations are strictly identical with brain processes, and (3) that there are brain processes which are identical with sensations, we have ensured that there are just as many danglers as there were before we accepted the strict identity of sensations and brain processes. Indeed, on Smart's thesis it turns out that brain processes are danglers, for now brain processes have all those properties that made sensations danglers.[23]

[22]Apparently in the sense of N-reduction. See pp. 14-15.

[23]J. T. Stevenson, "'Sensations and Brain Processes': A Reply to J. J. C. Smart," The Philosophical Review, 69 (1960), rpt. in The Mind/Brain Identity Theory, ed. C. V. Borst (London: Macmillan, 1970), p. 89.

If we momentarily accept the Smart-Stevenson terminology, 'nomological danglers' describes phenomena that are causally inefficacious, rather than the laws relating efficacious phenomena to inefficacious phenomena. Stevenson is mistaken when he claims that given the identity theory, sensations will still be danglers. Sensations are danglers only if epiphenomenalism is true, for then they would be causally inefficacious. But if the identity theory is true, epiphenomenalism is false, and sensations are efficacious. Consequently, if the identity theory is true, sensations are not nomological danglers (in the Smart-Stevenson sense of that expression) nor are the laws correlating sensations with brain processes nomological danglers (in Feigl's sense of the expression). Thus the identity theory does have the advantage over epiphenomenalism of eliminating nomological danglers (in both senses).

Smart's Materialism

Smart, in explaining the general principles of his own materialism, unfortunately goes into less detail and is more vague than Feigl. Nevertheless, the similarity between the materialist views of Smart and Feigl will be obvious. Smart offers this definition: "By 'materialism' I mean the theory that there is nothing in the world over and above those entities which are postulated by physics (or, of course, those entities which will be postulated by future and more adequate physical theories)."[24] Among the "entities postulated by physics," Smart means to include not just sub-atomic particles, atoms, and molecules, but aggregates of these entities as well. Smart believes that all organisms, including men, are aggregates of these physical entities. He writes, "my thesis is that man is a physical mechanism, and I frequently express this loosely in the form 'man is a machine'."[25]

When he says that man is a machine, Smart is not only asserting that human bodies are composed of physical particles; he is also asserting that the behavior of men will

[24]J. J. C. Smart, "Materialism," The Journal of Philosophy, 60 (1963), 651.

[25]P & SR, p. 107.

be explicable by the laws governing those physical constituents. Smart says, "I wish to lay down that it is incompatible with materialism that there should be any irreducibly 'emergent' laws or properties, say in biology or psychology."[26]

Smart's materialism thus has two basic principles which correspond loosely with Feigl's two theses of physicalism. The first of Smart's principles is an ontological claim, stating what kinds of things exist in the world; and this claim corresponds to the first thesis of physicalism. Smart's second principle specifies what kinds of laws will be needed for explaining the behavior of all that exists; and this principle corresponds to the second thesis of physicalism.

Reduction

In order to evaluate the irreducible properties objection, it will be necessary to examine in detail such notions as reduction, emergence, and the unity of science. Neither Feigl nor Smart provides a detailed discussion of these notions; instead, they refer their readers to the writings of various philosophers who have written extensively on these subjects. It is in this direction that we must now turn.

Like 'materialism', 'reduction' has many meanings. The sense of 'reduction' employed by modern materialists is exemplified when it is said that the science of thermodynamics has been reduced to statistical mechanics. The logical analysis of this kind of reduction is currently a topic of great interest and controversy among philosophers of science.[27]

There are two standard analyses of reduction in the contemporary literature. They are somewhat different, and

[26]"Materialism," p. 652.

[27]For a summary of this controversy, see Chapter 3 of J. J. C. Smart, Between Science and Philosophy (New York: Random House, 1968).

since Smart and Feigl do not express an explicit preference for one or the other, I shall describe both. In doing so, it will be helpful to make the standard logical-empiricist assumptions about the structure of science, and to adopt the following terminology. Let us suppose that a science may contain the following kinds of statements: (1) observation statements, which are singular statements used to describe experimental conditions and observations; and (2) theoretical statements, which are general (or universal) in form. The set of theoretical statements (call it 'the theory') may include: (a) theoretical postulates, and theorems derivable from them; (b) coordinating definitions, (correspondence rules) which connect the theoretical postulates with experimental concepts and procedures; and (c) experimental laws. The descriptive (non-logical) expressions contained in these statements are divided into two distinct classes: (1) theoretical terms, which occur only in the theory, and never in the observation statements; and (2) observation terms.

In reductions of the relevant type, the reduced (secondary) science contains descriptive expressions in its theory which do not occur in the theory of the reducing (primary) science. For example, the word 'temperature', which occurs in thermodynamic theory, does not occur in the theory of statistical mechanics.

The first model of reduction to be described is provided by Ernest Nagel.[28] Consider two sciences $\underline{S}_1$ and $\underline{S}_2$, along with their respective theories $\underline{T}_1$ and $\underline{T}_2$. $\underline{T}_2$ is $\underline{N}$-reduced to $\underline{T}_1$ only if: $\underline{T}_2$ contains expressions not contained in $\underline{T}_1$, and (1) there is a set, $\underline{K}$, of statements which link each expression occurring in $\underline{T}_2$ but not in $\underline{T}_1$ with some theoretical term of $\underline{T}_1$, such that (2) $\underline{T}_2$ can be derived (deduced) from $\underline{K}$ and $\underline{T}_1$. (1) is called the condition of connectability, and (2) is the condition of derivability.

The $\underline{N}$-reduction of one science to another essentially involves the deduction of the theory of the secondary science from the theory of the primary science along with

[28]Ernest Nagel, The Structure of Science: Problems in the Logic of Scientific Explanation (New York: Harcourt, Brace & World, 1961), esp. Chapter 11.

certain other connecting statements (K). As Nagel points out, this deduction would be impossible without the addition of these connecting statements, since the theory of the secondary science contains terms which do not appear in the theory of the primary science. The impossibility arises because of "the familiar logical canon that, save for some essentially irrelevant exceptions, no term can appear in the conclusion of a formal demonstration unless the term also appears in the premises." [29] Nagel argues persuasively that this logical canon does apply to the kinds of conclusions (i.e., scientific theories) which we are considering.

To illustrate this type of reduction, Nagel describes part of the procedure involved in the reduction of thermodynamics (the secondary science) to statistical mechanics and the kinetic theory of matter (the primary science). One part of thermodynamic theory is the Boyle-Charles' law for ideal gases, which can be written:

$$\text{(i)}\quad pV = kT.$$

The gas occupies a container such that V is the volume of the container, p is the instantaneous pressure on the walls of the container, T is the absolute temperature of the gas, and k is a constant for a given mass of gas. The reduction of thermodynamics will require the deduction of (i) from the theory of the primary science. The theory of the primary science contains as a postulate the statement

$$\text{(ii)}\quad pV = 2/3E$$

where p and V are the same as in (i) and E is the mean kinetic energy of the gas molecules in the container. However, the theory of the primary science does not contain the expression 'temperature' ('T'), and so we cannot deduce (i) from (ii), or from any other part of the primary theory unless we add some premise connecting 'T' with some theoretical term of the primary science. In this case we add the premise

$$\text{(iii)}\quad 2/3E = kT$$

[29] Nagel, pp. 352-53.

From (ii) and (iii) we can deduce (i), and consequently we have N-reduced a part of thermodynamic theory, the Boyle-Charles' law, (i). Similar deductions can be carried out for the other statements contained in the theory of thermodynamics, resulting in the N-reduction of that science to statistical mechanics and the kinetic theory of matter.

The status of the added premises, such as (iii), which connect the terms unique to the secondary science with theoretical terms of the primary science, is of great interest. Nagel proposes three alternative characterizations of these connecting statements. First, they may be considered "logical connections between established meanings of expression."[30] That is, the statement connecting the terms may be a descriptive definition, such that the terms are synonymous; or the statement may follow from some such definition, and thus involve a one-way logical entailment. If this is the case, then for any term 'A', occurring only in the secondary science, the meaning of 'A', "as fixed by the rules or habits of usage of the secondary science must be explicable in terms of the established meanings of theoretical primitives in the primary discipline."[31] The second alternative is "that the linkages are conventions, created by deliberate fiat."[32] That is, the connecting statements are coordinating definitions (rules of correspondence) which connect the theoretical terms "to experimental concepts and observational procedures,"[33] in a manner consistent with other coordinating definitions. The third alternative

> is that the linkages are factual or material. The assumptions then are physical hypotheses, asserting that the occurrence of the state of affairs signified by a certain theoretical expression 'B' in the primary science is a sufficient (or necessary and sufficient) condition for the state of affairs designated by 'A'.

[30]Nagel, p. 354.

[31]Nagel, p. 354.

[32]Nagel, p. 354.

[33]Nagel, p. 97.

> It will be evident that in this case independent evidence must in principle be obtainable for the occurrence of each of the two states of affairs, so that the expressions designating the two states must have identifiably different meanings ... the meaning of 'A' is not related analytically to the meaning of 'B'.[34]

Consequently, the truth of such connecting statements cannot be established solely by conceptual analysis; they must be supported by empirical evidence as well.

Which of these three alternatives describes the connecting statement ('$2/3\underline{E} = \underline{kT}$') in our example? Nagel claims that the first alternative does not apply. He points out that the descriptive expressions of a science are often at least partially defined by the theory or procedures of that science. The word 'temperature' has one meaning as employed by the layman, and another quite distinct (although related) meaning as defined by thermodynamic theory. Nagel asserts that

> 'temperature', in the sense the word is employed in classical thermodynamics, is not synonymous with 'mean kinetic energy of molecules,' nor can its meaning be extracted from the meaning of the latter expression. Certainly no standard exposition of the kinetic theory of gases pretends to establish the postulate by analyzing the meanings of the terms occurring in it. The linkage stipulated by the postulate cannot therefore be plausibly regarded as a logical one.[35]

This is an important point which has significant implications for the irreducible properties objection. Here is a paradigm case of reduction, such that an expression ('temperature') occurring only in the secondary theory is not synonymous with any word or words in the theoretical vocabulary of the primary theory. Consequently some non-analytic statement must be postulated which links this expression with some theoretical term ('mean kinetic energy of molecules') of the primary theory in order for the re-

[34]Nagel, pp. 354-55.

[35]Nagel, p. 355.

duction to occur. As Nagel says, "the essential point in this discussion is that in the reduction of thermodynamics to mechanics a postulate connecting temperature and mean kinetic energy of gas molecules must be introduced, and that this postulate cannot be warranted by simply explicating the meanings of the expressions contained in it."[36]

This point might be disputed as follows. Just as 'temperature' was once defined by the theory and procedures of thermodynamics, it is now defined by statistical mechanics so that 'temperature' is synonymous with 'mean kinetic energy of molecules'. Thus temperature is now identical, by definition, with kinetic energy, and consequently the connecting postulate ('$2/3\underline{E} = \underline{kT}$') is analytically true. Nagel counters this objection by pointing out that 'temperature' now has two different meanings, which we can distinguish by writing 'temperature_a' or '$\underline{T}_a$', defined by thermodynamics, and 'temperature_b' or '$\underline{T}_b$' defined by statistical mechanics. However from (ii) ('$\underline{pV} = 2/3\underline{E}$'), which is part of the theory of statistical mechanics, and (iii') ('$2/3\underline{E} = \underline{kT}_b$'), which is analytically true, we <u>cannot</u> deduce the Boyle-Charles' law, (i) ('$\underline{pV} = \underline{kT}_a$'). That argument would be invalid, because it equivocates between '$\underline{T}_a$' and '$\underline{T}_b$'. From (ii) and (iii') we can deduce

$$\text{(i')}\quad \underline{pV} = \underline{kT}_b.$$

However, (i') is not the Boyle-Charles' law, (i), and consequently we have not yet <u>N</u>-reduced that part of thermodynamics to mechanics. If we introduce a new postulate

$$\text{(iv)}\quad \underline{T}_b = \underline{T}_a$$

we could then deduce the Boyle-Charles' law, (i), from (ii), (iii'), and (iv). However, (iv) is not true as a matter of definition; consequently the <u>N</u>-reduction still employs a non-analytic premise, namely (iv).[37]

The question still remains as to whether the connecting postulate in our example, '$2/3\underline{E} = \underline{kT}$', is a rule of cor-

[36] Nagel, p. 357.

[37] Nagel, pp. 357-58.

respondence (the second alternative) or a straight factual hypothesis (the third alternative). Nagel claims that the answer to this question will vary according to the particular way the reduction is explicated. There are several possible explications, and the answer is determined by the order in which the connecting postulates are asserted.[38]

The second model of reduction is provided by Kemeny and Oppenheim.[39] They assume, as does Nagel, that the descriptive expressions of a theory can be divided into two groups, the observation terms and the theoretical terms. The theoretical terms are those expressions which are used only in the theory, and never in the observation statements of the science in question. Consider two theories T_1 and T_2. T_2 is K & O-reduced to T_1 if and only if: T_2 contains theoretical terms that are not theoretical terms in T_1, and all observational data explainable by T_2 is also explainable by T_1. Kemeny and Oppenheim add the condition that T_1 is as "well systematized" as T_2, but this condition is vague and is not relevant to the issues at hand, and so it will be ignored. They also stipulate that the theoretical terms unique to T_2 are not definable using the theoretical terms of T_1.

The concept of explanation employed is the standard deductive-nomological model.[40] The observational phenomenon to be explained is described by an observation statement (the explanandum) O_1. O_1 is explained by some theory, T, if and only if there is some observation statement O_2 (specifying initial conditions) such that the conjunction

[38]Nagel, pp. 356-57.

[39]John Kemeny and Paul Oppenheim, "On Reduction," Philosophical Studies, 7 (1956), 6-19.

[40]Kemeny and Oppenheim refer the reader to this model of explanation as described in Carl Hempel and Paul Oppenheim, "Studies in the Logic of Explanation," Philosophy of Science, 15 (1948), 133-75. They express some reservations about the adequacy of this definition, and indicate that their model of reduction would be consistent with an improved definition of 'explain'.

of $\underline{T}$ and $\underline{O}_2$ (the explanans) entails $\underline{O}_1$, although $\underline{O}_2$ alone does not entail $\underline{O}_1$.

For our purposes, the most important difference between $\underline{K}$ & $\underline{O}$-reduction and $\underline{N}$-reduction is that the latter requires statements which connect the terms unique to $\underline{T}_2$ with the theoretical terms of $\underline{T}_1$, while the former seems to require no such connecting statements. One reason for the absence of connecting statements is that $\underline{K}$ & $\underline{O}$-reduction, unlike $\underline{N}$-reduction, is applicable only when the primary and secondary sciences have the same observational vocabulary.[41] A further restriction on $\underline{K}$ & $\underline{O}$-reduction is that the terms unique to the secondary theory must be theoretical terms. No such requirements are involved in $\underline{N}$-reduction. These restrictions eliminate the need for connecting statements, but severely limit the applicability of the $\underline{K}$ & $\underline{O}$-reduction model.

Consider, for example, the reduction of psychology to physics-chemistry. We might expect an adequate psychological theory to explain among other things, the occurrence of certain mental states. Thus it should explain why Jones was in pain all night, and why Smith hates his neighbor. Since the mentalistic predicates ' __ is in pain' and ' __ hates __' appear in an explanandum, they are part of the observational vocabulary of mentalistic psychology. However, these predicates are not in the observational (nor of course, the theoretical) vocabulary of present day physics-chemistry. Consequently we cannot hope to $\underline{K}$ & $\underline{O}$-reduce mentalistic psychology to physics-chemistry.

Such $\underline{K}$ & $\underline{O}$-reduction is impossible because it requires that a deductive explanation be given, using physics-chemistry, for every phenomenon that can be explained by our imagined mentalistic psychology. But, as Jaegwon Kim has pointed out,[42] a deductive explanation of some phenomenon

[41] Paul Oppenheim and Hilary Putnam, "The Unity of Science as a Working Hypothesis," in Minnesota Studies in the Philosophy of Science, II, pp. 3-36, esp. p. 6.

[42] Jaegwon Kim, "Reduction, Correspondence, and Identity," The Monist, 52 (1968), 424-38.

described by mentalistic predicates cannot be given if the theory employed in the explanans does not contain those predicates (or any predicates which are logically equivalent to them). Most modern materialists assume that mentalistic predicates are not definable using only the predicates of physics-chemistry. If this assumption is correct, it seems reasonable to believe that a mentalistic observation statement (O_1) cannot be deductively explained by physics-chemistry (T_1), regardless of what the initial conditions, (O_2) may be.

This point is related to Nagel's argument that no theory can be explained by (deduced from) another theory if the conclusion contains terms not contained in the premises. The issue is somewhat more complicated in the present case, however, because O_2 may well contain the terms which are missing from T_1. In fact, according to the account of 'explanation' given above, we can deduce a mentalistic explanandum (O_1) from the theory of physics-chemistry (T_1). To use Kim's example: "take as your physical law 'Copper expands upon heating,' take as your singular premise 'This object is a piece of copper which does not expand upon heating, or else Jones is angry at his wife,' and finally take as your mentalistic explanandum 'Jones is angry at his wife.' Clearly this sort of 'deductive-nomological' inference can hardly be taken to constitute an explanatory inference."[43]

This example does not show that it is possible to explain mentalistic phenomena using physics-chemistry. Instead it points out a weakness in the standard definition of 'explain'. It seems reasonable to assert that, given an adequate definition of 'explain' (which is not presently available), it is impossible to deduce an explanandum which contains expressions not contained in (nor definable using expressions from) the theory of the explanans. Consequently,

[43]Kim, "Reduction," n. 7. Kim also refers the reader to R. Eberle, D. Kaplan, and R. Montague, "Hempel and Oppenheim on Explanation," Philosophy of Science, 28 (1961), 418-28; and J. Kim, "On the Logical Conditions of Deductive Explanation," Philosophy of Science, 30, (1963), 286-91.

if we assume the indefinability of mentalistic expressions using the expressions from physics-chemistry, it is impossible to explain mentalistic phenomena using only the theory of physics-chemistry.

One science cannot be K & O-reduced to another if the secondary theory contains observation terms not contained in the primary theory. In order to make the K & O-reduction model applicable in such cases, we must modify either the K & O-reduction thesis or the definition of 'explain'. The modification must allow the addition of some statements which connect the observation terms unique to the secondary theory with the theoretical terms of the primary theory. In our example, such connecting statements might take the form of conditional or bi-conditional psycho-physical laws. Similarly, in K & O-reducing biology to physics-chemistry we need bio-physical or bio-chemical connecting statements whenever there are observation terms in biology which are not found in physics-chemistry and which are not definable using expressions from physics-chemistry.

Thus both K & O-reduction and N-reduction require psycho-physical connecting statements for the reduction of mentalistic psychology to physics-chemistry. If we assume that mentalistic expressions are not definable using only the theoretical terms of physics-chemistry, we can conclude that the psycho-physical connecting statements will not be analytically true. In this respect they are like the connecting statement '$2/3E = kT$,' used in N-reducing thermodynamics to statistical mechanics. The psycho-physical connecting statements can be correspondence rules, supplying additional connections between the theoretical terms of physics-chemistry and observation terms; or they can be empirical laws.

It is important to distinguish the reduction models described above from the notion of reduction that was once advocated by logical positivists.[44] The early positivist model for reduction claimed that expressions unique to the

[44]See Carl G. Hempel, "The Logical Analysis of Psychology," Revue de Synthese (1935), trans. Wilfrid Sellars and rpt. in Readings in Philosophical Analysis, ed. H. Feigl and W. Sellars (New York: Appleton, 1949), pp. 373-84.

reduced theory were translatable (definable) into expressions from the reducing theory. For example, the reduction of psychology to physical science was thought to require the explicit definition of mentalistic expressions (e.g., 'is angry', 'is in pain') in terms of behavior or dispositions to behave. The two contemporary reduction models, in contrast, require no translations or definitions. Thus, the N-reduction or K & O-reduction of biology and psychology will not be hampered if biological and psychological expressions are not synonymous with expressions from the physical sciences.

Both the reduction models described above have certain general features which should be stated explicitly.[45] Reduction essentially involves a relation between two theories. Consequently saying simply that a certain theory is irreducible is as incomplete as saying that a certain dog is smaller. One theory (or branch of science) is reducible or irreducible only relative to some other theory (or branch of science). Science, however, is continually changing. As time passes not only does the body of experimental data increase, but also the theories themselves are modified or replaced by other theories. The contemporary theory of physics is quite different from the theory of physics two hundred years ago. Thus questions concerning reducibility must be understood relative to a certain time. Prior to 1850, thermodynamics was not reducible to the other branches of physics, such as classical mechanics. However, with the addition of the kinetic theory of matter as well as certain statistical assumptions to the body of physical theory, it became possible to reduce thermodynamics to the other branches of physics. Consequently, the irreducibility of one science to another at a given time does not demonstrate some ultimate fact about the fundamental nature of the universe. Instead it indicates certain logical relations between the theories of those sciences at that time. Changes in either or both theories might eventually bring about the reduction at some later date.

When Smart and Feigl say that there are no irreducible laws in psychology and biology, they are not claiming that, at this time, those sciences are reducible to physics-chem-

[45] Nagel, Structure of Science, Chapter 11.

istry. However, they do believe that at some future date the reductions will take place, and as materialists they attempt to defend this belief against philosophical objections.

Emergence

Thus far I have discussed the reduction of one theory or science to some other theory or science. The notion of an irreducible property, however, has not yet been explained. Unfortunately, neither Smart nor his critics have clarified just what an irreducible property is, and how the notion relates to inter-theoretic reduction. However, Smart does discuss emergent properties, and from his discussion it is quite clear that all emergent properties are irreducible properties.

The notion of emergence arises when considering the relations between the properties of complex wholes and the properties of their constituent parts. For example, it has been claimed that the occurrence of the properties of chemical compounds cannot be explained or predicted solely from a knowledge of the properties of the elements which combine to form those compounds. If this were true, the properties of the compounds would be emergent. This is not to say that any property of a complex entity which is not a property of any of its parts is an emergent property. As Smart says,

> there is a trivial sense in which new qualities emerge when simples are put together to form a complex .... while Boscovitch's point masses do not possess shape, a cloud of them could. Even four point masses have (in general) the property of determining a tetrahedron, but it would be absurd to say that each one of them determined a tetrahedron. The theory of emergence, if it is to say anything interesting, clearly must assert emergence in some sense other than this trivial one.[46]

In a footnote, Smart says that "an extended discussion of

[46] P & SR, pp. 51-52

the notion of emergence is given by E. Nagel,[47] and so we will use Nagel's characterization of a non-trivially emergent property.

Nagel describes the background conditions usually assumed by the proponents of emergence:

> let O be some object that is constituted out of certain elements a_1, ..., a_n standing to each other in some complex relation R; and suppose that O possesses a definite class of properties P, while the elements of O possess properties belonging to the classes A_1 ..., A_n respectively .... the occurrence of the elements a_1, ..., a_n in the relation R is by hypothesis the necessary and sufficient condition for the occurrence of O characterized by the properties P.[48]

Furthermore we are to suppose that we have "complete knowledge" of the elements of O. That is, we know all the properties of these elements, when existing either in isolation or in complexes other than O. Suppose that "there is at least one property P_e in the class P such that, despite complete knowledge of the elements, it is impossible to predict [deduce] from this knowledge that, if the elements stand to each other in relation R, then an object O possessing P_e will be formed."[49] In such a case O would be an emergent object, and P_e would be an emergent property of O.

For example, suppose we had complete knowledge (in the specified sense) of hydrogen, and complete knowledge of oxygen, but we could not deduce from this knowledge that water is translucent. Then translucency would be an emergent property of water.

Nagel points out that the above description of emergence embodies certain confusions. The notion of complete knowledge is misleading because it implies that we have, or

[47] P & SR, p. 51.

[48] Nagel, p. 367.

[49] Nagel, pp. 367-68.

could have, a complete list of all the properties of the constituent parts, $\underline{a}_1, \ldots, \underline{a}_n$, which specifies their "inherent natures". But when a constituent is a theoretical entity, such as a sub-microscopic particle, its properties will be determined largely by some theory which postulates these entities. Thus the properties of hydrogen atoms will vary from one atomic theory to another. Furthermore

> statements about the properties of complex wholes can be deduced from statements about their constituents only if the premises contain a suitable theory concerning these constituents--one which makes it possible to analyze the behavior of such wholes as "resultants" of the assumed behaviors of the constituents. Accordingly, all descriptive expressions occurring in a statement that is allegedly deducible from the theory must also occur among the expressions used to formulate the theory or the assumptions adjoined to the theory when it is applied to specialized circumstances.[50]

Consequently, the statement 'Water is translucent' will not be deducible from a theory concerning hydrogen and oxygen atoms unless that theory (or the accompanying special assumptions) contains the expressions 'water' and 'translucent'.

Consider another example. Gases have the property, $\underline{P}$, of increasing in pressure in direct proportion to increases in temperature when volume remains constant. It should be obvious that a statement describing the occurrence of this property cannot be deduced solely from classical Newtonian mechanics, which does not contain the expression 'temperature'. Suppose we add as a premise a molecular theory $\underline{T}_m$, which states among other things that gases are composed of perfectly elastic spherical molecules, whose dimensions are very small when compared to the average distances between them. As long as $\underline{T}_m$ does not contain the term 'temperature', we will not be able to deduce the occurrence of $\underline{P}$, from $\underline{T}_m$ and mechanics. We could then say that $\underline{P}$ was an emergent property of gases, relative to mechanics and $\underline{T}_m$; since we cannot deduce that gases have $\underline{P}$ solely from our knowledge of

[50] Nagel, pp. 368-369.

the properties of their parts (the molecules) as specified by mechanics and $\underline{T}_m$. However, if we add to our explanans the assertion, $\underline{S}$, that temperature is directly proportional to mean kinetic energy of molecules, we can deduce a statement describing the occurrence of $\underline{P}$. Thus $\underline{P}$ is not an emergent property relative to mechanics, $\underline{T}_m$, and $\underline{S}$.

It is not clear whether these additional assumptions, which relate micro-entities or properties (such as molecular energy) to macro-entities or properties (such as temperature) are to be considered as a part of the micro-theory, or as a separate auxiliary theory. In any case, a property will be emergent relative to an explanans, if the explanans lacks the necessary micro-macro laws.

As is the case with reduction, the concept of emergence, as explicated above, is relational. It makes no sense to speak of an absolutely emergent property, for whether or not a property is emergent depends upon what theory and additional special assumptions we use to predict (deduce) its occurrence. Thus, although certain chemical properties of compounds are emergent relative to the atomic theory of 1900, those properties are not emergent relative to contemporary quantum theory.

It is in this light that we must understand the materialists' rejection of emergence on the biological and psychological levels. When Smart and Feigl say that there are no emergent properties in biology or psychology, they obviously mean that none of the properties which these sciences attribute to living organisms (including man) are emergent with respect to some version of physics, or some combined science, physics-chemistry. Living organisms are considered to be complex arrangements of molecules, atoms, electrons, etc. Consequently the materialist believes that statements describing the occurrence of psychological and biological properties will be deducible from some theory concerning the micro-entities of which organisms consist. It is clear that relative to present day physics and chemistry, most biological and psychological properties are emergent, even if we include in the explanans the known micro-macro laws of biochemistry and biophysics. Consequently the materialist must be understood as predicting the deducibility of the occurrence of biological and psycho-

logical properties from some future physics-chemistry, conjoined with a greatly enlarged set of micro-macro laws.

The Unity of Science

One tenet of Smart's materialism is "that science is increasingly giving us a viewpoint whereby organisms are able to be seen as physio-chemical mechanisms."[51] This remark is accompanied by a footnote which refers the reader to the paper "The Unity of Science as a Working Hypothesis" by Oppenheim and Putnam (see n. 41). For a description of how micro-laws can explain macro-phenomena, Feigl also refers his readers to this paper.[52] Consequently we can assume that Smart and Feigl agree with at least the basic theses advanced in this paper by Oppenheim and Putnam.

Unity of science, according to Oppenheim and Putnam, "is attained to the extent to which the laws of science become reduced to the laws of some one discipline. If the ideal of such an all-comprehensive explanatory system were realized, one could call it Unitary Science."[53] Oppenheim and Putnam explicate and argue for the hypothesis that the social and biological sciences will eventually be reduced to the physical sciences. They adopt the Kemeny and Oppenheim model of inter-theoretic reduction. In addition, they utilize the notions of inter-branch reduction and micro-reduction, which they explicate as follows.

One branch of science, B_2, is reduced to another branch, B_1, at time t, if and only if there is some theory T_1 in B_1 at t such that T_1 reduces T_2 (where T_2 is the theory of B_2 at t).

For each branch of science and its corresponding theory there is supposed to be a specific universe of discourse. For most of the elements in the various universes of discourse there will be a specified part-whole relation.

[51] Smart, "S & BP," p. 161.

[52] "M & P," p. 376.

[53] "Unity of Science," p. 4.

Given these conditions, $\underline{B}_2$ is micro-reduced to $\underline{B}_1$ at $\underline{t}$ if and only if $\underline{B}_2$ is reduced to $\underline{B}_1$ at $\underline{t}$ and the elements of the universe of discourse of $\underline{B}_2$ have a decomposition into proper parts, all of which belong to the universe of discourse of $\underline{B}_1$. For example, suppose $\underline{B}_2$ is the branch containing molecules in its universe of discourse, and suppose we specify atoms to be the proper parts of molecules. If atoms are in the universe of discourse of $\underline{B}_1$, and if $\underline{B}_1$ reduces $\underline{B}_2$ at $\underline{t}$, then $\underline{B}_1$ micro-reduces $\underline{B}_2$ at $\underline{t}$.

A branch $\underline{B}_1$ is a potential micro-reducer of a branch $\underline{B}_2$ if and only if the elements of the universe of discourse of $\underline{B}_1$ are proper parts of the elements of the universe of discourse of $\underline{B}_2$. The relations 'micro-reduces' and 'potential micro-reducer' are transitive, asymmetric, and irreflexive.

Oppenheim and Putnam argue that unity of science will be achieved only via micro-reductions. They specify six reductive levels, each individuated by its respective universe of discourse:

6 . . . Social Groups
5 . . . (Multicellular) living things
4 . . . Cells
3 . . . Molecules
2 . . . Atoms
1 . . . Elementary particles

The objects of levels 1-5 are considered to be proper parts of the objects at the next highest level. Consequently each branch of science whose universe of discourse is at a given level will always be a potential micro-reducer of any branch whose universe of discourse is at the next highest level (if there is one).

There is supposed to be associated with each level a theory which contains a set of predicates normally used to describe things on that level. For example, '__ is hungry' would be a predicate in the theory of level 5, and '__ has a positive charge' would be a predicate in the theory of level 1. Unitary science will be achieved when for levels 1-5, the theory of level $\underline{n}$ micro-reduces the theory of level $\underline{n} + 1$.

Jerry Fodor claims that the identity theory, when held in conjunction with the views of Oppenheim and Putnam on the unity of Science, leads to absurd consequences. Fodor points out that, according to Oppenheim and Putnam, the micro-reduction of psychology to neurology is an important step towards achieving the unity of science. Fodor then writes:

> On the Oppenheim-Putnam account, "the essential feature of micro-reduction is that the branch [of science] B_1 [which provides the micro-reduction of B_2] deals with the parts of the objects dealt with by B_2."
>
> Our present point is that it is difficult to understand how this could be the correct model for the relation between psychological and neurological theories. Psychological entities (sensations, for example) are not readily thought of as capable of being microanalyzed into anything, least of all neurons or states of neurons. Pains do not have parts, so brain cells are not parts of pains.
>
> It is, in short, conceivable that there may be true psycho-physical identity statements, but it seems inconceivable that such statements are properly analyzed as expressing . . . relations between wholes and their parts.[54]

It is apparent that Fodor has misunderstood the Oppenheim-Putnam account of the micro-reduction of psychology to neurology. On their account, the objects of a branch of science are the elements in that branch's universe of discourse, as specified by one of the six reductive levels. Psychology is a branch dealing with level five; consequently the objects of psychology are not sensations, or any other mental states. The objects in question are living organisms (people, rats, etc.) and the parts of those objects are cells. Thus there is no need to assert that pains have parts, and that their parts are brain cells; those absurdities are not involved in the micro-reduction of psychology.

Fodor is correct in claiming a disanalogy between the

[54]Jerry A. Fodor, Psychological Explanation: An Introduction to the Philosophy of Psychology (New York: Random House, 1968), p. 112.

identification of mental states with brain states and the identifications usually associated with micro-reductions (e.g. water is H_2O). These commonly cited and relatively uncontroversial examples involve the identification of some whole (e.g. a water molecule) with the aggregate of its micro-parts (e.g. two hydrogen atoms and one oxygen atom). In such cases, an object on one reductive level is identified with a complex of objects from a lower reductive level. Mental states and brain states, however, are not themselves objects on any of the six reductive levels. They are states <u>of</u> certain objects on such levels.

Some scientists and philosophers assert that heat (or temperature) is identical with the mean kinetic energy of molecules. This claim is much more controversial, on conceptual grounds, than the assertion that water is H_2O. Heat (or temperature) and mean kinetic energy are not objects on any Oppenheim-Putnam reductive level; they are attributes or states of objects on certain levels. In this respect the purported identification of heat with mean kinetic energy is analogous to the identification of mental states with brain states.

Irreducible Properties and Unwelcome Terminology

At the beginning of this paper I pointed out that materialism is incompatible with certain kinds of properties. Having discussed materialism in some detail, it is now possible to characterize specifically the kinds of properties excluded from the materialist framework. To begin with, all properties which are not inter-subjectively knowable (i.e., not physical$_1$) are excluded by materialism. This would include what Smart calls "<u>purely</u> phenomenal or introspectible"[55] properties. In addition, any emergent properties on the biological, psychological, or social levels would be outside of the physicalist picture. If the identity theory implies the existence of properties from either of these two categories, then the identity theory is incompatible with materialism.

Corresponding to their ontological preferences, materialists also have linguistic preferences. All words

[55]"S & BP," p. 166. Emphasis added.

which are not part of the intersubjective ($physical_1$) language are excluded from the materialist's preferred vocabulary. In addition since materialists deny that there are any emergent properties, they cannot use words in describing the world which would commit them to the existence of such properties. For example, the materialist who is also an identity theorist wants to say that sensations exist, i.e. that people feel pains, have itches, etc. But a materialist cannot make such assertions if being in pain is an emergent property. Thus words which would commit the materialist to the existence of emergent properties cannot be part of his preferred vocabulary.

III: The Irreducible Properties Objection Again

The Semantic Argument

Having discussed materialism in some detail, it is now possible to examine specific versions of the irreducible properties objection. As will become evident, the irreducible properties objection is actually a family of interrelated objections, all of which argue from the identity theory to the falsity of materialism. One version of the objection has been proposed by J. T. Stevenson,[56] and can be represented by the following argument.

A

(1) The identity theory is true.

(2) If the identity theory is true, there are brain processes which are identical with sensations.

(3) The defining properties of sensations are also properties of those brain processes with which the sensations are identical.

(4) If the defining properties of sensations are not M-properties, then some brain processes have properties which lie outside the physicalist framework (M-properties are all and only those properties within the

[56]Stevenson, pp. 87-94.

physicalist framework).

(5) If there are brain processes with properties outside the physicalist framework, then materialism is false.

(6) If the defining properties of sensations are M-properties, then 'sensation' is definable in terms of M-properties.

(7) "If 'sensation' is definable in terms of M-properties, then 'sensation' is synonymous with some word or words in the materialist's preferred vocabulary."[57]

(8) "'Sensation' is not synonymous with 'brain process' or any other word in the materialist's preferred vocabulary."[58]

(9) Therefore, the defining properties of sensations are not M-properties (from (6), (7), & (8)).

(10) Therefore, materialism is false (from (4), (5), & (9)).

Premise (2) is obviously true, as long as we include as part of the identity theory the assertion that sensations exist. It is worth mentioning this apparently trivial point, for it is just this assertion that is denied by eliminative materialists, such as Feyerabend and Rorty.[59]

The notion of a 'defining property' in (3) needs some explanation. Stevenson assumes that words can be explicitly defined by providing an appropriate list of properties. These are the properties which the definiendum connotes. For example, Stevenson says, "'Evening Star' connotes among other things 'appearing in the evening'."[60]

[57]Stevenson, p. 90.

[58]Stevenson, p. 89.

[59]P. K. Feyerabend, "Mental Events and the Brain," The Journal of Philosophy, 60 (1963), 295-96; and Richard Rorty, "Mind-Body Identity, Privacy, and Categories," The Review of Metaphysics, 19 (1965), 24-54.

[60]Stevenson, p. 88.

Apparently Stevenson has something like this in mind. In general, a definition of some predicate '$\underline{P}$' would be:

$$\underline{\text{Def}}: \quad \Lambda \underline{x}(\underline{P}\underline{x} \leftrightarrow \underline{F}_1\underline{x} \wedge \underline{F}_2\underline{x} \ldots . \underline{F}_n\underline{x})$$

A definition of some name '$\underline{N}$' would be:

$$\underline{\text{Def}}: \quad \underline{N} = \iota \underline{x}(\underline{F}_1\underline{x} \wedge \underline{F}_2\underline{x} \ldots . \underline{F}_n\underline{x})$$

These definition sentences are supposed to be necessarily true. The properties expressed by the predicates '$\underline{F}_1$' ... '$\underline{F}_n$' are the defining properties of the expression being defined. In a derivative sense, we can say that the defining properties of the name, '$\underline{N}$' are also defining properties of the thing named by '$\underline{N}$'. Similarly, we can say that the defining properties of the predicate '$\underline{P}$' are also defining properties of the property expressed by '$\underline{P}$', as well as the things having that property. (This explanation of 'defining properties' is oversimplified, but should be adequate in the present context.) As Stevenson points out, if $\underline{A} = \underline{B}$, then $\underline{A}$ and $\underline{B}$ share all the same properties, including each other's defining properties. Thus (3) follows from (1) and (2).

Premise (4) follows from (3). (5) is entailed by materialism. Premises (6) and (7) follow from the theory of definitions outlined above. The kinds of expressions excluded from the materialist's vocabulary were described in Section II. The assumption here is that the words constituting the predicates which express $\underline{M}$-properties are in the preferred vocabulary. Thus to get (7), Stevenson is probably using some principle like this: if '$\underline{A}$' is definable in terms of $\underline{M}$-properties, using predicates $\underline{F}_1$ ... F_n, then those predicates consist of words from the materialist's preferred vocabulary which together are synonymous with '$\underline{A}$'.

I have arranged the argument so that premise (8) plays a key role, and consequently the brunt of my criticism will be directed at (8). However, (8) needs some explanation. Stevenson focuses on defining 'sensation' in his statement of the argument; but we can and should generalize the argument to cover not only 'sensation', but also the names of particular kinds of sensations and other mental states, e.g., 'pain', 'the feeling of nausea', 'anger', etc.

We can, more or less, distinguish from other words the words of our language which constitute the mental vocabulary. I mean by 'mental vocabulary' all the words normally used to describe minds, their activities, and their contents. For example: 'pain', 'fear', 'thinking', 'feel', 'mental', 'sensation', 'psychic', 'mind', 'sensation of yellow', 'experience', 'conscious', 'believes'.

Now consider argument (A) in its most generalized form. The generalized version of (8) would assert that the ordinary names of mental states, which are formed from the mental vocabulary, are not synonymous with any word or words in the materialist's preferred vocabulary. Thus (8) would be rewritten as:

(8a) The ordinary names of mental states are not themselves in the materialist's preferred vocabulary

and

(8b) The ordinary names of mental states are not synonymous with any other words which are in the materialist's vocabulary.

I will postpone evaluating (8a). However, even if (8a) is true, a materialist could refute this argument by providing the appropriate definitions which would show (8b), and therefore (8), to be false.

Smart originally replied to the irreducible properties objection by trying to provide such definitions. He attempted to give translations of first person sensation reports in which the ordinary names of sensations are replaced by "topic-neutral words"[61] which are obviously not excluded from the materialist's preferred vocabulary. Smart has conceded that these early attempts at definition were inadequate; nevertheless, a refined program for providing such definitions has been proposed by D. M. Armstrong

[61]"S & BP," p. 167.

and D. K. Lewis.[62]

It is possible to interpret Stevenson as providing an argument which purports to prove that no such definitions can be successful.[63] The argument is as follows. If mental states can be defined in terms of M-properties, then the identity of mental states with brain states cannot be contingent; for mental states would then be defined by the same properties which define brain states. But the identity of mental states with brain states is contingent, and therefore mental states cannot be defined in terms of M-properties. Jerome Shaffer seems to argue in a similar fashion:

> In general, it is hopeless to expect to be able to define psychic properties in terms of physical properties, and still hold, as Identity Theorists do, that it is a factual discovery that [mental states] and [brain states] are identical. Unless there are special features that allow us independently to identify [mental states], we can never be in a position to discover their de facto identity with [brain states].[64]

A simple example should be sufficient to show that the above argument is unsound. Consider the contingently true statement 'A = B', where 'A' is the ordinary name of some mental state and 'B' is the neurophysiological name of some brain state. Suppose the defining properties of B are the M-properties expressed by the predicates F_1 and F_2. Now, if the defining properties of A were also expressed by F_1

[62]D. M. Armstrong, A Materialist Theory of the Mind (New York: Humanities Press, 1968); David K. Lewis, "An Argument for the Identity Theory," The Journal of Philosophy 63 (1966), 17-25. For criticism of these definitions, see Neil Lubow, "The Mind-Body Identity Theory," Diss. UCLA, 1974, Ch. 3.

[63]Stevenson is interpreted in this way by Judith Economos in "The Identity Thesis," Diss. UCLA, 1967, p. 127.

[64]Jerome Shaffer, "Could Mental States Be Brain Processes?," The Journal of Philosophy, 58 (1961), rpt. in Borst, p. 119.

and F_2, then the statement '$A = B$' would of course not be contingently true. But the above argument provides no reason to believe that some other M-properties, expressed by the predicates F_6, F_7, and F_8, could not be the defining properties of A. In general, it is possible both for the defining properties of mental states to be M-properties, and for the mental state-brain state identities to be contingent, as long as the M-properties which define mental states are not the same M-properties as the ones which define the relevant brain states.[65] Consequently (8b) has not been proven true.

Although the materialist can defend his position by offering definitions which would demonstrate that (8b) is false, it seems far easier to deny that (8a) is true. Neither Stevenson nor any other proponent of this objection offers any reason to believe that the mental vocabulary is not part of the physicalist's vocabulary; and I will try to show that the mental vocabulary is not so excluded.

In Section II two classes of expressions were said to be excluded from the materialist's vocabulary. The first class includes any words which are not part of the physical$_1$ language. If non-physical$_1$ words are used in a statement, that statement would not be intersubjectively confirmable. The question now is whether or not statements containing words from the mental vocabulary are in principle intersubjectively confirmable. Feigl believes they are, for he says:

> Is it not an "objective" fact of the world that Eisenhower experienced severe pain when he had his heart attack? Is it not a public item of the world's history that Churchill during a certain speech experienced intense sentiments of indignation and contempt for Hitler? Of course! What is meant here is simply that statements about facts of this sort are in principle inter-

[65]This same point is made by Economos (p. 133) in response to Stevenson; and by James Cornman in response to Shaffer; see Cornman, "The Identity of Mind and Body," The Journal of Philosophy, 59 (1962), rpt. in Borst, pp. 125-126.

subjectively confirmable ...[66]

It would be absurd to claim that no statements employing the mental vocabulary are intersubjectively confirmable. The statements, 'Dramamine often prevents sensations of nausea' and 'Morphine is an effective pain-killer' are well confirmed by modern medical science, and it is easy to produce numerous additional examples.

It is important to see just how broad the term 'physical$_1$' is. Any phenomenon is physical$_1$ if it is causally efficacious (directly or indirectly) with respect to some event that is detectable by the senses. Thus, even if two-way interactionist dualism were true, mental events would still be physical$_1$.[67] Furthermore, even if our ordinary concept of a mental state entails interactionist dualism, the ordinary mental language would still be physical$_1$ language.

Feigl does claim that "a purely phenomenal or absolutely private language"[68] is possible. Although such a language would not be intersubjectively confirmable, nevertheless it "would still enable the solitary thinker silently to label the qualities of his direct experience . . . ."[69] However, introspective reports, like 'I am in pain', made in languages like English, cannot be statements in an absolutely private language. Such languages, as Feigl describes them, cannot be spoken or written,[70] and this is obviously not true of English. Thus it is a mistake to suppose that the mental vocabulary is not part of the physical$_1$ language.

[66]"M & P," p. 398.

[67]In "Physicalism," pp. 252-53, Feigl says that interactionist dualism is compatible with mental events being physical$_c$. The term 'physical$_c$' has essentially the same meaning as 'physical$_1$'.

[68]"M & P," p. 402.

[69]"M & P," p. 403.

[70]"M & P," p. 402.

The second class of expressions excluded from the materialist's vocabulary are those expressions which imply the existence of emergent properties. However, Stevenson provides no reasons to suppose that the mental vocabulary is excluded on this ground. Neither Smart nor Feigl makes such a claim about the mental vocabulary. On the contrary, Smart's attempts at translation are intended to prove that the mental vocabulary has no such anti-materialist implications. To claim that the mental vocabulary implies emergence, without providing support for that claim, is simply to beg the question.

Whether or not the properties characterized by the mental vocabulary are emergent is in part an empirical question. In Section II it was pointed out that the materialist's denial of emergence on the biological and psychological levels is a prediction about future theories. The materialist believes that, in the future, chemistry-physics will be able to provide the required explanations, even though they cannot be provided by present day physical theories. To prove that there is no emergence, the materialist will have to wait for scientific progress. If science develops as the materialist expects, his claims will be proven; but this involves both a matter of time and a matter of fact. For the present, the materialist can use the micro-reductions already carried out at various levels as evidence for his general claim.[71] Consequently, the materialist now has some evidence for his denial of emergence, but his case has not been decisively proven. On the other hand, the critic who claims that properties characterized by the mental vocabulary are emergent (relative to future physical theories) has offered no conclusive support for his claim. It may be possible to provide an a priori argument to show that these properties are emergent, but no such argument has been provided. At the present time, the materialist and his critic are at a stand-off; neither has proven his position on emergence. Consequently, any argument containing the premise that the mental vocabulary implies emergence must be considered unconvincing, since that premise is not known to be true.

[71]See Oppenheim and Putnam, "The Unity of Science," secs. 5 & 6.

On these grounds the materialist can respond to argument (A) by denying premise (A:8a). The mental vocabulary is not excluded from the physicalist's preferred vocabulary for either of the two possible reasons. The mental vocabulary is physical$_1$ vocabulary, and the materialist can point out that there is no reason to believe that it implies emergence. As a consequence, the materialist can also deny that premise (A:9) is true, since it was purportedly proven by a sub-argument using (A:8) as a premise.

The tendency to exclude the mental vocabulary from the physicalist's preferred vocabulary is fairly widespread; consequently some speculation as to its source seems worthwhile. It might be that philosophers with this tendency are confusing contemporary materialism with an older version of materialism that was once advocated by the logical positivists. This is especially plausible, since Feigl himself was once a major proponent of logical positivism.

According to this earlier view, physical terms are "those terms which we need -- in addition to logico-mathematical terms -- for the description of processes in inorganic nature ..."[72] In this sense of the expression, ordinary mental terms are not physical terms, nor do they appear to be definable in physical terms. This concept of a physical term is of course much narrower than the concept of a physical$_1$ term.

Furthermore, the older positivist version of materialism employed a notion of reduction that is quite different from the contemporary reduction models. According to this older view, the reduction of psychology to physics requires the translation of all psychological terms into physical terms. However, as was pointed out above, the contemporary reduction models require no such translations.

It is easy to see why a philosopher who identified contemporary materialism with its earlier positivist counterpart would assume that the mental vocabulary is not part of

[72]Rudolf Carnap, "Logical Foundations of the Unity of Science," in International Encyclopedia of Unified Science, I, ed. Otto Neurath, et al. (1938; rpt. Chicago: University of Chicago Press, 1955), p. 46.

the physicalist's preferred vocabulary. I have tried to show, however, that this assumption is unwarranted.

The Argument from Introspection

Another version of the irreducible properties objection is concerned with properties of mental states which are observed or noticed in introspection. The proponents of this form of the objection contend that when we are introspectively aware of our own mental states, we are aware of certain properties of those states. For example, we are usually able to tell what mental state we are in at a particular time. We do not mistake pains for sensations of yellow, nor do we confuse itching sensations with feelings of nausea. We are able to distinguish various types of sensations from one another, and we do so by observing, in introspection, the features or properties of our current inner experiences. The properties which we are purported to notice in introspection are supposed to be what allow us to recognize and identify our own mental states. It is then claimed, for one reason or another, that these introspectible properties lie outside the physicalist framework.

To make the case seem stronger, the critic of the identity theory can relate those introspectible properties to the defining properties discussed above. It might be claimed that the introspectible properties by which we recognize our own mental states are among the defining properties of those states. But this assertion is not an essential part of either version of the irreducible properties objection.

In a generalized form, the argument from introspectible properties can be outlined as follows:

B

(1) The identity theory is true.

(2) If the identity theory is true, there are brain processes which are identical with sensations.

(3) The introspectible properties of sensations are also properties of those brain processes with which the sensations are identical.

(4) If the introspectible properties of sensations are not M-properties, then some brain processes have properties which lie outside the physicalist framework.

(5) If there are brain processes with properties outside the physicalist framework, then materialism is false.

(6) The introspectible properties of sensations are not M-properties.

(7) Therefore, materialism is false.

This argument is obviously similar to argument (A) discussed above. Premises (1) through (5) are the same, except that 'defining properties' is replaced in (B) by 'introspectible properties'. Premise (B:6) is the analogue of (A:9). This is the premise that needs support from additional arguments, and there are some suggestions in the literature. Shaffer provides one such argument in the following passage:

> Let us take the case where a person reports the having of some mental event, the having of an after-image, a thought, or a sensation of pain. Now such a person has surely noticed that something has occurred, and he has surely noticed that this something has some features (or how could he report it was an after-image rather than a sensation of pain?) Now it seems to me obvious that, in many cases at least, the person does not notice any physical features--he does not notice that his brain is in some particular state, nor does he notice any external physical stimulus, nor any physical event between the stimulus and the neurological response. Yet he does notice some feature. Hence he must notice something other than a physical feature. The noticing of some non-physical feature is the only way to explain how anything is noticed at all.[73]

[73]Jerome Shaffer, "Mental Events and the Brain," The Journal of Philosophy, 60 (1963), rpt. in Borst, p. 136.

For the sake of argument, let us accept the dubious explanation of introspection and of our ability to recognize our own mental states contained in the above passage. Shaffer's argument is then as follows:

(i) I notice in introspection some property, P, of my current mental state.

(ii) I do not notice in introspection any M-property.

(iii) Therefore, P is not an M-property.

The problem with this argument centers around (ii). If (ii) is taken to mean

(ii') There is no property which I notice in introspection and which is an M-property.

then the argument begs the question against the materialist. The materialist claims that nothing has anything but M-properties. Thus he will not grant that both (i) and (ii) are true, and Shaffer has offered no reason to believe that they are true. Since Shaffer denies that his argument begs the question,[74] perhaps (ii') is not what he intended. It seems likely that he had in mind instead

(ii'') I do not notice in introspection that any property is an M-property.

The materialist can grant (ii''), for it does seem to be obviously true, as Shaffer claims. But if (ii'') replaces (ii'), the argument is no longer valid. From the fact that I notice P and do not notice that P is an M-property, it does not follow that P is not an M-property. Suppose I notice a red haired man standing in a corner of a crowded room, and suppose I do not notice that any man in the room has a wooden leg. It certainly does not follow that the red haired man does not have a wooden leg. Arguing from ignorance in this fashion is simply invalid. Thus, taken one way the argument begs the question; taken the other way, it is invalid.

[74]Shaffer, "Mental Events and the Brain," p. 137.

I am not aware of any other arguments explicitly stated in the literature which purport to show that introspectible properties are not M-properties. Max Deutscher, who tries to defend the identity theory, hints at such an argument. The suggestion is that introspectible properties are "non-public,"[75] and consequently they are not M-properties. The central claim here seems to be that properties, which we are aware of via introspection, are not intersubjectively knowable. But this claim is greatly in need of support. Although introspection is not an intersubjective means for acquiring knowledge, it does not follow that what is known about via introspection cannot also be known about via intersubjective means. I know that I have a pain through introspection, but others come to know about my pain via their senses; hence my pain is intersubjectively knowable. The general principle that anything we are aware of via introspection cannot also be known about intersubjectively is false. Consequently it cannot be used to support the claim that properties known introspectively cannot also be intersubjectively known. The materialist can claim that introspectible properties are M-properties, for their being introspectible does not make them non-intersubjectively knowable.

Thus the two arguments for (B:6) have been shown to be unsound. As a result, (B:6) remains a dubious premise, and without further support, argument (B) is ineffective. The materialist can claim that (B:6) is false, and hence that (B) is unsound.

Arguments from Specific Properties

Some proponents of the irreducible properties objection avoid general arguments for the existence of such properties, and instead pick out some particular property and claim that it is outside the physicalist framework. This move is usually made by calling our attention to first person introspective reports (e.g., "I have a pain in my foot," "My skin itches") which Smart calls avowals. The statement "I have a yellowish-orange after-image"[76] is frequently discussed

[75]Max Deutscher, "Mental and Physical Properties," in Presley, p. 70.

[76]Smart, "S & BP," p. 161.

in Smart's writings on the identity theory, and critics of the theory naturally focus their attention on this and similar avowals. Cornman writes: "assuming, for example, that the sentence 'I see a yellowish-orange after-image' is a report about some brain process, that brain process would have the property of 'being a yellowish-orange after-image'. If this is a property it is certainly a psychic property, that is, a property that lies outside a materialistic framework."[77] First of all, Cornman is mistaken when he claims that the brain process would have the property of being a yellowish-orange after-image. It is true that the brain process in question will have all the properties of the sensation with which it is identical. However, as Smart has frequently pointed out, it is not the after-image, but the experience of having an after-image, which is supposed to be a brain process. The property of being a yellowish-orange after-image is not a property of the experience, and therefore not a property of the brain process. Nevertheless, we can make Cornman's point by arguing that an experience of having a yellow-orange after-image has the property of being an experience of having a yellow-orange after-image (a dog has the property of being a dog). Another critic, M. C. Bradley, uses a similar example. Bradley points out that the sensation reported in the statement 'I have a sharp pain in my hand' has the property of being a pain.[78]

It does follow from the identity theory that some brain states have the psychic properties described above. But this admission does not lead directly to the demise of materialism, for it has not yet been shown that these psychic properties are not M-properties. Nevertheless neither of these critics makes any attempt to show this. In the passage quoted above, Cornman simply asserts that "certainly" such psychic properties lie "outside a materialistic framework." Later in his article Cornman takes a more moderate position, stating that such psychic properties "at least

[77]Cornman, p. 125.

[78]M. C. Bradley, "Sensations, Brain-Processes, and Colours," Australasian Journal of Philosophy, 41 (1963), 388.

seem to lie outside a physicalistic framework."[79] Bradley also adopts this more moderate view.

Both Cornman and Bradley, along with many other philosophers, simply assume that such psychic properties (e.g. being a pain, being an experience of having a yellowish-orange after-image) are at least prima facie not M-properties. The identity theorist is then expected to prove that these properties are M-properties. Although it is never made very explicit, the underlying assumption seems to be that any property expressed or named by a word or words in the mental vocabulary is prima facie not an M-property. Consequently it is up to the identity theorist to show that these psychic properties are identical with properties expressed or named by words in the materialist's preferred vocabulary, which they also assume to contain no words from the mental vocabulary.

I have already argued that this tendency to exclude the mental vocabulary from the physicalist's vocabulary is unjustified. It apparently stems from a misunderstanding of materialism, or some other confusion. In any case, the materialist can simply deny that his principles exclude the mental vocabulary, or the properties expressed or named by that vocabulary. No good reason has been provided to show that such psychic properties lie outside the materialist framework. Consequently, this last version of the irreducible properties objection is unconvincing.

To summarize, the various versions of the irreducible properties objection follow this pattern. If the identity theory is true, then brain states have all the properties of the mental states with which they are identical. Some of these properties are considered prima facie outside of the physicalist framework because they are either: defining properties of mental states, introspectible properties, properties reported in avowals, or simply properties named or expressed by words from the mental vocabulary (or for some combination of these reasons). These prima facie irreducible properties will thwart materialism unless they are actually M-properties.

[79]Cornman, p. 127.

Some critics then provide arguments which purport to prove that certain of these *prima facie* irreducible properties cannot be M-properties. They then conclude that materialism is false. Other critics simply point to certain properties of mental states and then claim that they are not M-properties. I have tried to show how a materialist could reply to these critics. In response to those critics who provide arguments which attempt to prove that certain properties are not M-properties, the materialist-identity theorist can deny the truth of certain key premises, (A:8), (A:9), and (B:6). The materialist can at least reach a stand-off with his critic as to the truth of these premises, and consequently these arguments are unsuccessful. In response to those critics who simply assert, without giving reasons or arguments, that certain properties are not M-properties, the materialist can simply assert that they are mistaken. Thus the irreducible properties objections are all unsuccessful.

Irreducible Properties, The Identity Theory, and Materialism

I have already pointed out that the irreducible properties objection is not an objection to the identity theory by itself; instead, the objection contends that the identity theory is incompatible with materialism. Having discussed in detail several versions of this objection, it is now possible to see in just what way the identity theory is supposed to conflict with materialism.

Rogers Albritton has suggested[80] that we can construct an argument against materialism which is similar to the arguments employed in the irreducible properties objection, but which makes no mention of the identity theory or brain states. The argument is as follows:

C

(1) Mental states exist.

(2) If mental states exist, there are things (the mental states) which have properties that are outside the physicalist framework.

[80] In conversation.

(3) If there are things with properties outside the physicalist framework, then materialism is false.

(4) Therefore, materialism is false.

Premise (2) might be supported by the kind of claims used in the various versions of the irreducible properties objection. It might be asserted that the properties of mental states are outside the physicalist framework because they are either: defining properties of mental states, introspectible properties, properties reported in avowals, or properties named or expressed by words from the mental vocabulary. I have already argued that these are not good reasons for thinking that the properties of mental states are not M-properties. Consequently, the materialist can provide a reply to this argument which is similar to the replies that can be made to the various versions of the irreducible properties objection.

Nevertheless, C was not proposed because it is a good argument against materialism. Instead it was proposed because it contains the same sort of attack on materialism as is found in the irreducible properties objection, although it does not use the identity theory as a premise.

In what sense, then, does the irreducible properties objection contend that the identity theory is incompatible with materialism? In this sense: the identity theory asserts that mental states exist, and therefore the identity theory entails premise (C:1).

Of course some other mind-body theories, such as interactionist dualism, also entail (C:1); but one expects dualism to be incompatible with materialism. The identity theory, on the other hand, is supposed to be a materialist solution to the mind-body problem. Thus it is surprising to find that it also provides an essential premise in an argument against materialism. In contrast, one contemporary materialist theory of mind, (eliminative materialism) simply denies that mental states exist at all. In this way eliminative materialists avoid problems stemming from the supposed properties of mental states, albeit at the expense of having an extremely paradoxical position.

Like most identity theorists, Smart adopts the identity theory as a corollary to the more fundamental theses of materialism. At one time, Smart's inability to reply successfully to the irreducible properties objection forced him to give up the identity theory and "to go over to a more Feyerabendian position."[81] This "Feyerabendian position" is eliminative materialism. Thus, when unable to reply to the irreducible properties objection, Smart dropped the identity theory and replaced it by a theory that denies the existence of mental states. In so doing, Smart was able to retain his materialism.

The conclusion to be drawn is this. The identity theory entails that mental states exist. It is this feature, and only this feature, of the identity theory that is essentially involved in the irreducible property-type arguments against materialism. It is in this sense that the irreducible properties objection argues against the compatibility of the identity theory and materialism.

[81]Smart, "Comments on the Papers," in Presley, p. 91. Also see n. 3 above.

Neil Lubow
Department of Philosophy
University of New Hampshire
Durham, New Hampshire 03824