

INFORMATION TO USERS

This material was produced from a microfilm copy of the original document. While the most advanced technological means to photograph and reproduce this document have been used, the quality is heavily dependent upon the quality of the original submitted.

The following explanation of techniques is provided to help you understand markings or patterns which may appear on this reproduction.

1. The sign or "target" for pages apparently lacking from the document photographed is "Missing Page(s)". If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting thru an image and duplicating adjacent pages to insure you complete continuity.
2. When an image on the film is obliterated with a large round black mark, it is an indication that the photographer suspected that the copy may have moved during exposure and thus cause a blurred image. You will find a good image of the page in the adjacent frame.
3. When a map, drawing or chart, etc., was part of the material being photographed the photographer followed a definite method in "sectioning" the material. It is customary to begin photoing at the upper left hand corner of a large sheet and to continue photoing from left to right in equal sections with a small overlap. If necessary, sectioning is continued again — beginning below the first row and continuing on until complete.
4. The majority of users indicate that the textual content is of greatest value, however, a somewhat higher quality reproduction could be made from "photographs" if essential to the understanding of the dissertation. Silver prints of "photographs" may be ordered at additional charge by writing the Order Department, giving the catalog number, title, author and specific pages you wish reproduced.
5. PLEASE NOTE: Some pages may have indistinct print. Filmed as received.

Xerox University Microfilms

300 North Zeeb Road
Ann Arbor, Michigan 48106

74-25,221

HORGAN, Terence Edward, 1948-
MICROREDUCTION AND THE MIND-BODY PROBLEM.

The University of Michigan, Ph.D., 1974
Philosophy

University Microfilms, A XEROX Company, Ann Arbor, Michigan

MICROREDUCTION AND THE
MIND-BODY PROBLEM

by
Terence Edward Horgan

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Philosophy)
in The University of Michigan
1974

Doctoral Committee:

Professor Jaegwon Kim, Chairperson
Professor Richard D. Brandt
Professor Alvin I. Goldman
Professor Klaus F. Riegel
Instructor James Koski, Jr.

TABLE OF CONTENTS

CHAPTER ONE: PSYCHOPHYSICAL IDENTITY AND THE UNITY OF SCIENCE.....	1
CHAPTER TWO: THE "NOMOLOGICAL DANGLER" ARGUMENT.....	19
CHAPTER THREE: PROPERTY IDENTITY AND MICROREDUCTION...	52
CHAPTER FOUR: THE PART-WHOLE SUPERVENIENCE PRINCIPLE..	88
CHAPTER FIVE: PSYCHOLOGY AND THE PART-WHOLE SUPER- VENIENCE PRINCIPLE.....	135
CHAPTER SIX: EVENTS.....	174
APPENDIX.....	205
REFERENCES.....	213

I. PSYHCOPHYSICAL IDENTITY AND THE UNITY OF SCIENCE

Section 1

The mind-body problem is the problem of the relation between the mental and physical, in humans and in other higher animals. Recent discussions have tended to formulate the problem in terms of events: How are mental events related to physical events? A currently popular answer is the psychophysical identity thesis (henceforth "IT"), which asserts that every mental event involving a given person is identical to some physical event involving the person's body.

IT is generally thought to be closely related to the Unity of Science hypothesis. This hypothesis asserts that a unified system of microreduction is obtainable, in principle, among the various domains of science. The system would establish a reductive hierarchy among the sciences in accordance with the hierarchy of composition of the entities with which each science deals. Social groups are composed of multicellular living things, which are composed of cells, which are composed of molecules, which are composed of atoms, which are composed of elementary particles. Accordingly, the social sciences would be ranked high in the reductive hierarchy, the biological sciences would occupy a middle position, and the physics of elementary particles would be the final reductive stage. Since the reducibility relation among theories is transitive, this envisioned reductive system would microreduce all of science to elementary-particle physics, thereby establishing the Unity of Science.

Both advocates and opponents of IT are often found to agree on these two points: 1) that a successful microreduction of mentalistic psychology to natural science would be at least highly relevant to the truth or falsity of IT, and 2) that one argument for IT is that its truth would enable just such a reduction of psychology to physical theory.¹ Agreement on these contentions by both parties to the dispute establishes the putative connection between IT and the Unity of Science hypothesis, because the reduction of psychology to natural science would represent one important step toward the total reductive unification of science. Arguments in favor of the hypothesis tend to support the reducibility of psychology to physical theory, this reduction being one component of the overall pattern of reduction among the sciences. Accordingly, these arguments are thought to support IT as well. Conversely, arguments against IT are likely to be viewed as arguments against the reducibility of psychology to natural science. And the impossibility of microreduction in this particular instance would be sufficient to falsify the Unity of Science hypothesis.

In this thesis I shall question the alleged close connection between IT and the Unity of Science, and I shall explore the actual relation between mental events and physical events which would result from the microreduction of psychology to physical science. My proposed solution to the mind-body problem will be materialist, because it assumes the Unity of Science hypothesis. Within the framework of unified science, man is viewed as a complex physicochemical entity. He has no non-physical components. Furthermore, the sciences of man are considered reducible to the basic physical sciences. Such a view of man and the sciences of man surely deserves the label 'materialist'.

But my solution will also be dualist, in that it does not identify mental events with physical events. My position is quite different from Cartesian dualism, however, because I do not postulate that mind and body are distinct substances. Event dualism, as we shall see, is a far less radical proposal than substance dualism.

My approach to the mind-body problem will involve a detailed discussion of microreduction in science. Accordingly, much of what I say in this thesis will have a direct bearing on the topic of microreduction per se, a topic falling squarely within the philosophy of science. We shall see that an argument originally put forward by Herbert Feigl and J.J.C. Smart in favor of IT has profound implications concerning the nature of microreduction generally. A major part of this thesis will be devoted to a critical examination of those implications of the Feigl-Smart argument.

Section 2

At first glance, it might seem that reduction of psychology to natural science would consist simply in showing that persons are identical to their bodies and that every psychological event is identical to some physical event. (The word 'event' is used in a broad sense here, to include states and processes.) If the matter is viewed in this way, of course, IT is part and parcel of the microreduction of psychology. But if we consider the prominent accounts of reduction in the philosophy of science, we find that the relation of IT to the microreducibility of psychology is far more tenuous than it originally appeared.

This subject is examined carefully by Jaegwon Kim, in "Reduction, Correspondence, and Identity."² Much of what I shall say in this chapter is based on Kim's observations, and I refer the reader to his paper for further details to supplement the present discussion.

Let T_m be the (mentalistic) psychological theory to be reduced, and Let T_p be the physical theory serving as the reduction base. We assume that T_m and T_p are axiomatic theories with their distinctive primitive predicates and postulates. Now, there are three well-known models of reduction in the philosophy of science, one due to Kemeny and Oppenheim³ and two due to Nagel.⁴ Following Kim, we list three reduction theses, which correspond to the three models of reduction.

The Kemeny-Oppenheim Reduction Thesis:

T_m contains terms not in T_p , and any observational data explainable by T_m are explainable by T_p .

The Nagel Reduction Thesis:

T_m contains terms not in T_p , and (1) (the condition of connectability) there is a set C of established T_m - T_p connecting principles, i.e. statements including terms from both T_m and T_p such that (2) (the condition of derivability) every postulate in T_m is derivable from the postulates of T_p in conjunction with the set C .

The strong Nagel Reduction Thesis:

T_m contains terms not in T_p , and (1) (the condition of connectability) there is a set C of established T_m - T_p connecting principles, such that (2) (the condition of derivability) each T_m -postulate is derivable from T_p -postulates and C , and further (3) (the condition of translatability) for each T_m statement there is a T_p statement such that their biconditional is derivable from the set C .

The salient feature of these three theses is that no mention is made of identity in any of them. Identity of objects will indeed come into the picture, if we specify that the reduction of T_m to T_p is to count as microreduction. T_m is microreducible to T_p just in case 1) T_m reduces to T_p in one of the above three ways, and 2) the objects countenanced by T_m are structured wholes, whose parts are objects countenanced by T_p . If these conditions are met, then T_m things (persons)

are identical to certain complex T_p things: persons are complex physicochemical structures.

With respect to the alleged identity of mental events with physical events, however, the above three reduction theses are silent. T_m may well be microreducible to T_p , even if no mental event is identical to any physical event.

In order to get a firmer grip on the relation of IT to microreduction, we need a formulation of IT that is commensurate with the above reduction theses. Since reduction is a relation between theories, and the point at issue is the logical relation of IT to the reduction theses, we need a theory-relative version of IT. We construe the descriptive predicates of T_m and T_p as expressing generic events--that is, event-types. Following Kim, we state IT as follows:

The Identity Thesis:

For every concrete event x , if there is some event-type M in T_m such that x is an M -type event, then there is a concrete event y such that for some event-type P in T_p , y is a P -type event, and $x=y$.

IT thus asserts that every concrete mental event (countenanced by T_m) is identical to some physical event or other (countenanced by T_p). It does not make any claim regarding the connection between mental event-types and physical event-types.

So construed, IT turns out to be logically independent of all three of the above reduction theses. That is, IT does not imply any of the three theses, and none of the theses implies it. (The reader is again referred to Kim's paper, for a detailed argument in favor of these claims.) Thus, it is a mistake to assume that IT bears an intimate logical relationship to the hypothesis that psychology is reducible to natural science.

At this point, the identity theorist might reply that we have missed his point. As he construes IT, he

might say, the statement of a specific psychophysical identity presupposes a certain generic correlation law. That is, we identify two concrete events only when the appropriate generic events (properties) are connected by a correlation law. He might also contend that one important motivation for IT is precisely that it explains such psychophysical correlation laws. So once IT is seen to presuppose the existence of these laws, its connection to the above reduction theses will become more evident.

This line of response must be scrutinized on two counts. First, we must inquire into the metaphysical status of events themselves, to see whether there is any ground for the contention that the statement of a specific psychophysical identity always presupposes a generic correlation law. Secondly, even if there is ground for this contention, we must still ask whether reduction really requires event-identities.

If events are conceived metaphysically as structureless concrete particulars, then there is no reason to think that IT as formulated above entails that there are psychophysical laws. Every concrete mental event might be identical to some concrete physical event, and yet there might be no systematic connection whatever between mental event-types and physical event-types. I.e., there might be no generic correlation laws connecting the predicates of T_m to those of T_p .

There is an alternative conception of events, however, according to which the statement of a specific event-identity does entail the existence of a certain generic correlation law. This analysis of events, which has been proposed explicitly by Kim,⁵ treats an event as the exemplifying of a property by an object (or an ordered series of objects) at a time. Thus, an event can be represented by an ordered triplet (P, x, t) where 'P', 'x', and 't' represent the property, the object, and the time. Property P is said to be the constitutive

property of the event (P, x, t) .⁶ Existence and identity conditions for events are forthcoming: (P, x, t) exists if and only if the object x has the property P at the time t . $(P, x, t) = (Q, y, t')$ if and only if $P = Q$, $x = y$, and $t = t'$.

According to this construal, event (P, x, t) counts as a mental event if and only if property P is a mental property; in terms of the above theory-relative formulation of IT, (P, x, t) is a mental event if and only if P is a property countenanced by T_m . Similarly, (Q, y, t') counts as a physical event if and only if Q is a physical property; in theory-relative terms, (Q, y, t') is a physical event if and only if property Q is countenanced by theory T_p .

This approach treats events as structured entities, rather than concrete particulars. And because of the identity conditions for events, a specific event identity of the form " $(P, x, t) = (Q, y, t')$ " does entail a property-correlation statement of the form " $(x)(Px \text{ iff } Qx)$." The reason is simple: if $(P, x, t) = (Q, y, t')$, then the constitutive properties P and Q are identical. Identical properties are co-extensive.

Returning to IT, Kim's construal of events guarantees that every specific mental-physical event identity of the form " $(U_m, x, t) = (U_p, y, t')$ " is founded upon a psychophysical law of the form " $(x)(U_mx \text{ iff } U_px)$." So IT entails the following thesis.⁷

The Correspondence Thesis:

For each primitive predicate M in T_m there is a configuration P of T_p -predicates such that
 $(x)(Mx \text{ iff } Px)$.

The correspondence thesis, in turn, clearly is quite relevant to the microreducibility of psychology to natural science. Specifically, it is logically equivalent to the strong Nagel reduction thesis, and it entails (but is not entailed by) the weak Nagel reduction thesis. (For a detailed discussion of these various logical interconnections, I again refer the reader to Kim's paper,

"Reduction, Correspondence, and Identity.") So according to Kim's analysis of events, IT entails all three reduction theses.

I shall adopt Kim's analysis in this thesis, because it does have these logical consequences. Entailment of some such statement as the correspondence thesis is a minimal condition for any viable formulation of the identity thesis, since otherwise there is no logical link at all between IT and the various reduction theses.

Another reason why I am adopting Kim's account of events is that it is the most natural and intuitively plausible account that I know of. Intuitively, individual events do not seem to be concrete particulars like objects, nor abstract particulars like numbers, nor universals (properties). Rather, they seem to be ontologically "mixed" entities, with an element of the universal and an element of the particular. An event is the having of a property by an object (or an ordered series of objects) at a time.

We come now to the second question raised earlier: granted that events are to be conceived in such a way that IT entails the correspondence thesis, does this fact compel us to adopt IT as a part of the hypothesis that psychology is reducible to natural science? The answer is clearly negative, if we confine ourselves to the accounts of reduction reflected in the above three reduction theses. The Kemeny-Oppenheim thesis and the weak Nagel thesis are weaker than even the correspondence thesis, and the latter is itself weaker than IT. The strong Nagel thesis is logically equivalent to the correspondence thesis, and therefore it too is weaker than IT. So even under Kim's construal of events, it continues to be the case that any of the reduction theses could be true even if IT were false. We can adhere to the reducibility of psychology, and to the Unity of Science hypothesis in general, without adopting the identity thesis.

Section 3

Some writers on the subject of intertheoretic reduction have contended that none of the three standard accounts of reduction are adequate as they stand. It has been argued that fully adequate reduction would indeed involve identities, and that the standard accounts are therefore too weak.

We will begin with an uncontroversial fact, already noted above. It is agreed on all sides that micro-reduction requires certain identity statements. If theory T_2 is microreducible to theory T_1 , then each concrete entity countenanced by T_2 is composed of concrete entities which are countenanced by T_1 . (People are composed of cells, cells of molecules, molecules of atoms, etc.) Thus, each object in T_2 's domain of discourse is identical to some complex entity whose components are objects from T_1 's domain of discourse. Thing-identity statements are clearly needed in microreduction, although the questions of property-identity and event-identity are more controversial.

Turning now to a stronger claim regarding identity, Lawrence Sklar⁸ has argued that systematic thing-identity statements are needed in all types of intertheoretic reduction, not only in microreduction. His claim is supported by the examples of reduction that are usually discussed in the literature, and it is intuitively quite appealing. Take the reduction of optics to electromagnetic theory--a case of reduction which is not a microreduction. Scientists have identified light with electromagnetic radiation of certain wavelengths, and this identification seems crucial. On the hypothesis that light and electromagnetic radiation are always copresent but are not identical, we are inclined to deny that optics is reducible to electromagnetic theory, even though light obeys the laws which govern

electromagnetic radiation. Our denial is based on the firm feeling that a theory T_2 does not reduce to a theory T_1 unless the concrete objects and substances postulated by T_2 are identical either 1) to concrete objects and substances postulated by T_1 , or 2) to structured wholes whose parts are concrete objects and substances postulated by T_1 . If this condition is not met then T_1 and T_2 are "about" different kinds of things, and this disparity in their respective domains of discourse seems to render the theories irreducible to one another.

So Sklar's claim about the need for thing-identity statements in all cases of intertheoretic reduction seems quite reasonable, and it is compatible with all I shall say below. Let me emphasize once again, however, that the need for thing-identity statements leaves open the question of whether reduction requires either property identities or event identities.

It might be thought that the presence of thing identities does at least imply the identity of the properties expressed by sortal predicates. Sortal predicates classify things into kinds; examples are 'is a gene' and 'is a DNA molecule'. But thing-identity statements do not imply the identity of the relevant sortal properties. To see why, we should consider the logical form of the relevant thing-identity statements. They are general identities, which tell us that all light is electromagnetic radiation of such-and-such wavelength, that all genes are DNA molecules, that all salt is sodium chloride, and so on. Because of this generality, they have the logical form of universally quantified conditionals or biconditionals. The predicates involved in the assertions are sortal predicates.

Now, there is no a priori reason why the two sortal predicates involved in such an assertion should express the same property. In order for genes to be identical to DNA molecules, it is sufficient that anything which has

the property 'is a gene' also has the property 'is a DNA molecule'. The question of the identity or nonidentity of sortal properties therefore is left open.

We come now to the view that reduction involves the identity of properties, as well as the identity of things. This position has been explicitly advocated by several writers, including Herbert Feigl,⁹ Hilary Putnam,¹⁰ Kenneth Shaffner,¹¹ and Robert Causey.¹² Shaffner's work deserves explicit consideration here, because it also involves an attempt to unify several different accounts of reduction, some of which are not mentioned above. In particular, he takes into account the studies of Popper,¹³ Feyerabend,¹⁴ and Kuhn¹⁵ on the relation of later scientific theories to earlier ones, and the difficulties involved in obtaining an older theory as an exact derivational consequence of a newer theory cum connecting principles. He allows for these considerations by introducing the notion of a corrected theory. He adopts a version of Nagel's derivability requirement, but he allows a theory T_2 to be reducible to a theory T_1 even if it's not possible to deduce T_2 itself from the conjunction of T_1 and appropriate connecting principles. Rather, T_1 and the connecting principles need only imply a corrected version of the original theory T_2 . Thus, provision is made for the familiar fact that a reducing theory will often show that the reduced theory is only approximately correct.

Shaffner's point is well taken. I shall not mention it again explicitly, and I shall often speak of reduction as though it involved the exact derivation of a higher-level theory from a lower-level theory plus connecting principles. But the reader should keep in mind that this way of speaking implicitly allows for the fact that an actual higher-level theory might need correction before it conforms to strict derivational reduction. I shall also occasionally describe the goal of unified science as

"the reduction of all of science to microphysics." It should be understood that this ideal allows for future corrections and alterations not only in the higher-level sciences, but also in microphysics itself.

Shaffner's approach to reduction involves three key ingredients: 1) acceptance of Nagel's derivability requirement, but allowance for it to apply to a corrected version of the reduced theory rather than to the reduced theory itself; 2) acceptance of the requirement--familiar from strong Nagel reduction--that every primitive term of T_2 be connected via a biconditional connecting principle to some term of T_1 ; and 3) stipulation that the connecting principles linking T_2 -terms to T_1 -terms be "interpretable as representing referential identities."¹⁶ Thus, his view is a modified version of strong Nagel-reducibility. The modifications consist in a weakening of Nagel's strong model to allow approximate derivation of T_2 from T_1 plus connecting principles, and in a strengthening of the strong model through the requirement that all biconditional connecting principles must express synthetic identities.

Shaffner's adoption of the identity requirement is our main concern, of course. Notice that if we adopt the identity requirement, then we must disallow putative reductions which conform to the weak Nagel model but not the strong Nagel model--that is, putative reductions involving connecting principles some of which are not biconditional in form. The reason, of course, is that connecting principles must express identities; in order to designate one and the same property, two predicates must be fully coextensive.

Adoption of the identity requirement also disallows putative reductions which conform only to the K-O model. But one can still admit, as Shaffner does, that there may be cases where we have to fall back upon the K-O paradigm. These would be cases of "replacement reduction,"

in which the reduced theory is not true or even approximately true. In such cases, the reduced theory is literally replaced by the reducing theory, rather than being explained by it. Hence there is no need to derive the reduced theory T_2 from the reducing theory T_1 plus a set of connecting principles; it is sufficient that T_1 should explain the observational data originally explained by T_2 . (An example of this type of reduction might be the explanation, by oxidation theory, of all the observable facts which phlogiston theory explained.)

Independently of the question whether intertheoretic connecting principles must be identities, it seems clear that K-O reduction is best suited to cases where the reduced theory is radically false. Reduction in other cases should provide a deductive-nomological explanation of the reduced theory (or of a corrected version of the reduced theory). Deductive explanation, in turn, requires the satisfaction of Nagel's derivability requirement. I will have little to say about K-O reduction in this thesis, therefore, since I will be concerned with the reduction of true (or approximately true) theories. Weak Nagel reduction will be discussed further in chapter V. Until we come to chapter V, however, our discussion will assume the strong Nagel model, with the implicit understanding that reduction may involve a "corrected version" of the reduced theory.

Section 4

What arguments can be given in favor of the claim that reduction requires property-identities? One putative consideration is theoretical simplicity. It sometimes seems to be suggested that an intertheoretic property-identity statement amounts to a definition of a T_2 -predicate by means of a T_1 -predicate.¹⁷ For example, 'temperature' is defined as 'mean molecular kinetic energy'.

Definitional introduction of T_2 -predicates is clearly simpler than non-definitional introduction, since the former method allows us to obtain T_1/T_2 connecting principles without employing any primitive terms beyond those is T_1 itself. Hence, property identity appears to be theoretically simpler than mere correlation.

This argument is clearly in error. A synthetic identity statement is not a definition. Definitions are the result of meaning analysis or direct linguistic stipulation. Microreductive property-correlations fall into neither category. Rather, they are synthetic, empirically disconfirmable hypotheses. Treating these hypotheses as identity statements does not turn them into definitions. Hence, such treatment does not reduce the number of primitive theoretical properties or laws.

A second argument sometimes cited in favor of property identity in reduction invokes considerations of ontological economy.¹⁸ The systematic identification of T_2 -properties with their correlated T_1 -properties results in our countenancing fewer distinct kinds of property. So it is suggested that we ought to identify them, by Occam's razor.

We need not deny that this argument provides some prima facie support for the identification of microreductively correlated properties. But it is far from decisive, for several reasons. First, certain microreductively correlated properties are intuitively quite different: pain and its neural correlate, for example. So there is a prima facie case against their identification, given our intuitions.

Second, the ontological economy which would result from a limitation in the number of distinct property-kinds is somewhat less impressive than the economy which would result if we were to eliminate properties-altogether. Ontological economy of the kind in question is not an overwhelmingly compelling goal.

Third, the issue must be viewed in the context of constructing a theory of properties, a task which includes the specification of identity criteria. I shall discuss criteria of property-identity in a later chapter. According to the most natural and intuitively plausible criteria of property identity, I shall contend, microreductively correlated properties are not identical.¹⁹

A third argument for property identity in microreduction is based on the writings of Herbert Feigl²⁰ and J.J.C. Smart²¹ on the mind-body problem. Their argument for psychophysical identities is often rather vague with regard to the distinction between events and properties, but I think it is best viewed as an argument primarily in favor of the identity of microreductively correlated properties, and secondarily in favor of event-identities on the basis of these property identities. Their argument has quite general implications, and it can be expanded into a general argument for the need for inter-theoretic identities at all levels of microreduction. The argument is based on considerations of theoretical simplicity, but it is considerably more complex and sophisticated than the argument which claims that identity allows for the definitional introduction of higher-level theoretical terminology.

We shall examine this argument in detail in chapter II. The argument falls short of establishing the need for property identity in microreduction, as I will argue. Nevertheless, it contains an important element of truth; it is based upon assumptions which are likely to seem highly intuitive and congenial to an advocate of the Unity of Science hypothesis. I shall formulate a principle which appears to constitute the correct explication of these assumptions. I call it the part-whole supervenience principle. According to this principle, a certain special relation obtains between microreductively correlated properties. The relation is not identity,

but it is very intimate. In particular, it is more intimate than mere nomological correlation. The supervenience principle will become the focal point of this thesis.

In chapter III I discuss criteria for property identity, and I offer arguments against the view that microreductively correlated properties are identical. In chapter IV I argue in favor of the part-whole supervenience principle. This principle is the basis for my ultimate proposed solution to the mind-body problem. In chapter V I consider some ways in which the microreduction of psychology might be expected to differ from standard microreductions in natural science, and I claim that these differences do not affect the truth of the supervenience principle.

In chapter VI I return to the subject of events, in light of my discussion of properties. Simultaneous events involving certain microreductively correlated constitutive properties are not identical, on my account, since the constitutive properties are not identical. But the relation between such events is very intimate, because of the close relation between microreductively correlated properties which is asserted by the part-whole supervenience principle. My proposed solution to the mind-body problem emerges as a special case of the event-relation I discuss in this chapter.

¹Cf. Jaegwon Kim, "Reduction, Correspondence, and Identity," The Monist 3 (1968): 424.

²Ibid., pp. 424-38.

³John Kemeny and Paul Oppenheim, "On Reduction," Philosophical Studies 7 (1956): 6-19.

⁴Ernest Nagel, "The Meaning of Reduction in the Natural Sciences," Philosophy of Science, ed. A. Danto and S. Morgenbesser (New York: Meridian Books, 1960), pp. 288-312 and E. Nagel, The Structure of Science (New York: Harcourt, Brace, and World, 1961), XI.

⁵Kim, "On the Psychophysical Identity Theory," American Philosophical Quarterly 3 (1966): 227-35. A more recent and more fully-developed analysis is presented in Kim's "Causation, Nomic Subsumption, and the Concept of Event," JP 70 (1973): 217-36.

⁶It should be noted that Kim (1973) does not represent an event by an ordered triple in the usual set-theoretical sense, because of certain complications which arise when stating identity conditions for events whose constitutive properties are relational properties. This matter can be safely ignored here.

⁷In saying that IT entails the Correspondence Thesis, I am assuming that each primitive predicate of T_m is instantiated at least once. IT refers to concrete mental events--viz., to instancings of mentalistic predicates. Hence, the thesis applies only to predicates that have at least one instance. If it turned out that T_m contained primitive predicates that have no instances, then we could state IT initially as a thesis about event-types (properties) as follows. "For every event-type M in T_m there is some event-type P in T_p , such that $M=P$."

⁸Lawrence Sklar, "Types of Inter-theoretic Reduction," BJPS 18 (1967): 109-24.

⁹Herbert Feigl, The "Mental" and the "Physical": The Essay and a Postscript (Minneapolis: University of Minnesota Press, 1967).

¹⁰Hilary Putnam, "On Properties," Essays in Honor of Carl G. Hempel, ed. N. Rescher (New York: Humanities Press, 1970), pp. 235-54.

¹¹Kenneth Schaffner, "Approaches to Reduction," PS 34 (1967): 137-47.

¹²Robert Causey, "Attribute-Identities in Microreductions," JP 69 (1972): 407-22.

¹³Karl Popper, "The Aim of Science," Ratio 1 (1957): 24-35; "Truth, Rationality, and the Growth of Knowledge," in his Conjectures and Refutations (New York: Basic Books, 1962), pp. 215-50.

¹⁴Paul Feyerabend, "Explanation, Reduction, and Empiricism," Minnesota Studies in the Philosophy of Science 3, ed. H. Feigl and G. Maxwell (Minneapolis: University of Minnesota Press, 1962): 28-97; "On the 'Meaning' of Scientific Terms," JP 72 (1965): 266-74.

¹⁵Thomas Kuhn, The Structure of Scientific Revolutions (Chicago: University of Chicago Press, 1962).

¹⁶Shaffner, p. 144.

¹⁷This view may underlie Wilfred Sellars and Paul Meehl's well-known notion of "physical₂" events and entities: "An event or entity is physical₂ if it is definable in terms of theoretical primitives adequate to describe completely the actual states though not necessarily the potentialities of the universe before the appearance of life." (Emphasis mine.) Quoted from "The Concept of Emergence," Minnesota Studies in the Philosophy of Science, 1, ed. H. Feigl and M. Scriven (Minneapolis: University of Minnesota Press, 1956), p. 252. Carl Hempel also speaks of biconditional microreductive connecting principles as "definitions," although he places scare-quotes around the word "definition." See "Reduction: Ontological and Linguistic Facets," Philosophy, Science, and Method: Essays in Honor of Ernest Nagel, ed. S. Morgenbesser, P. Suppes, and M. White (New York: St. Martin's Press, 1969), pp. 179-99.

¹⁸Shaffner, pp. 145-46.

¹⁹Feigl, pp. 78-98.

²⁰J.J.C. Smart, "Sensations and Brain Processes," PR 68 (1959): 141-56; J.J.C. Smart, Philosophy and Scientific Realism (New York: Humanities Press, 1963).

II. THE "NOMOLOGICAL DANGLER" ARGUMENT

Section 1

Feigl and Smart have claimed that identifying the mental with the physical enables us to eliminate what they call "nomological danglers," irreducible and unexplainable psychophysical laws. They appear to be arguing as follows: We adopt the Unity of Science hypothesis. In particular, we assume that mentalistic psychology is reducible to natural science (to neurophysiology, perhaps, and ultimately to microphysics). Now, suppose that we discover a set of systematic psychophysical correlation laws. The laws of psychology could then perhaps be explained on the basis of the laws of neurophysiology in conjunction with these psychophysical laws. But how are the correlations themselves to be explained? They cannot be taken as irreducible and unexplainable, because then it would not be strictly true that psychology is reducible to neurophysiology; it would be more accurate to say that psychology is reducible to neurophysiology cum correlation laws. Psychophysical laws would "dangle" outside the domain of the reducing science, since they contain psychological concepts. On the other hand, it isn't clear how these correlation laws could be explained on the basis of the reducing theory, since the laws of neurophysiology and the other natural sciences make no reference to mental or psychological concepts at all. So psychophysical correlations pose a dilemma: they are neither explainable nor acceptable as part of the reducing theory.

Feigl and Smart contend that the way out of this dilemma is to identify the entities and properties which are connected by psychophysical correlation laws. If

we discover a lawlike correlation between experiences of pain and occurrences of a particular pattern of neural stimulation, then we should identify pain with this type of neural stimulation. A correlation that is based on an identity requires no further explanation, since the identity tells us that the "correlated" entities or properties are really one and the same. Moreover, the identity itself is simply not the sort of fact which requires explanation; we don't ask why the morning star is the evening star, or why water is H_2O . Similarly, if the property pain is the associated brain-property, we don't need to ask why this is so.

Identities are different from laws of nature, it is suggested, in that it is reasonable to ask for an explanation of a law. An explanation will not always be forthcoming, of course, since the law might be a fundamental one within a particular scientific theory, and the theory itself might not be reducible to any other theory. But in these cases, the demand for explanation comes to an end because there happens to be no further explanation to be given, and not because of the nature of the law itself. In principle, it is possible that there is some true theory which microreduces the theory containing the law we are interested in. Hence, it is possible that the law is explainable on the basis of more basic laws. Identity statements, on the other hand, supposedly describe facts which don't even need explanation. It makes no sense to ask why two differently named things are in fact one and the same thing. Thus, identities differ fundamentally from standard laws of nature.

Because the question "Why?" allegedly never arises with respect to them, identity statements are held to be the only acceptable basis for connecting the concepts of psychology to those of the reducing science. Identities don't have the same status as the laws of the reducing

science, since they aren't really laws at all. Yet they don't need to be explained on the basis of the reducing science, since they don't need explanation at all. Thus, if we can derive the laws of psychology from the laws of natural science in conjunction with psychophysical identity statements, then the reduction to natural science will be complete. Identity statements, unlike mere psychophysical correlation laws, are not "nomological danglers" which themselves require explanation.

As I mentioned above, this argument of Feigl and Smart's can be applied to cases of microreduction in general, and not merely to the reduction of psychology to natural science. The generalized form of the "nomological dangler" argument (henceforth the N-D argument) concerns the status of intertheoretic connecting principles. As we saw in Chapter I, such principles are necessary if we are to accomplish a microreduction of one theory to another. The generalized N-D argument runs as follows: consider two theories, T_1 and T_2 , where T_2 is reducible to T_1 , via connecting principles C . This reducibility implies that all the laws of T_2 can be explained on the basis of the laws of T_1 . The explanation will conform to the deductive-nomological model, because of Nagel's derivability requirement. I.e., the laws of T_2 can be derived by logic alone from $T_1 \cup C$. But what is the status of the connecting principles themselves? It won't be possible to give a deductive-nomological explanation of them on the basis of the laws of T_1 , because they contain concepts which fall outside of T_1 -- concepts peculiar to T_2 . On the other hand, these connecting principles cannot be irreducible unexplained laws of nature. If they were it would not be strictly true that T_2 is reducible to T_1 alone. Strictly, T_2 would only be reducible to $T_1 \cup C$. So in order to uphold the claim that T_2 is reducible to T_1 alone, we must grant a privileged metaphysical status to the connecting principles C . Connecting principles

must be statements whose truth simply doesn't require explanation. Only identity statements can meet this description. Therefore, identity statements are the only bridge principles allowable in microreduction. Any weaker kind of principle, such as a correlation law which is not based on an underlying identity of the correlated entities or properties, will be a "nomological dangler", unacceptable as a part of the reducing theory T_1 , yet unexplainable on the basis of the laws of T_1 .¹

So the Feigl-Smart scenario for unified science is as follows. We suppose that all branches of science are reducible to microphysics via a complex system of microreductive connecting principles linking each branch in the scientific hierarchy to the branch immediately below it. Two sorts of laws will count as logically primitive within unified science: 1) the laws of microphysics, and 2) the various connecting principles. According to the Feigl-Smart argument, only the first of these two sorts of logically-primitive laws will count as fundamental laws. The second component consists entirely of identity statements, which are non-fundamental. Their nonfundamentality is closely related to the fact that identity statements never need explanation. So higher-level theories are reducible to microphysics alone, precisely because microreductive connecting principles do not count as fundamental laws within unified science.

Section 2

I will ultimately argue for acceptance of the key claims made by advocates of the N-D argument, although I will question their conclusion. That is, I shall argue that we should accept the contention that microreductive connecting principles are "non-fundamental" laws whose truth "requires no explanation." When these claims are understood, however, it will become clear that they do not support the conclusion that microreductive connecting

principles must be identity statements.

The immediate goal of the present chapter is to provide clarification of the notions of fundamentality and immunity from the need of explanation. The assertion which will emerge as explication of these concepts (the "part-whole supervenience principle" mentioned at the close of chapter I) will become the focal point in subsequent chapters. Although I will reject the thesis that microreductively correlated properties are identical, I will argue in favor of the supervenience principle.

We turn first to the idea of non-fundamentality, which has the following intuitive content. The microreductive connecting principles of unified science are non-fundamental, relative to the laws of microphysics, if the correspondence principles are "fixed" or "determined" by the laws of microphysics alone. I.e., the truth of the connecting principles is somehow inevitable, given the truth of the laws of microphysics, despite the fact that connecting principles have the status of logically primitive postulates within unified science. (Consequently the laws of microphysics also "determine," in and of themselves, the truth of all higher-level theoretical laws within unified science.)

The intuitive notion of "determination," which I will shortly discuss in some detail, also helps to express the content of the claim that microreductive connecting principles should be immune from the need of explanation. Suppose, once again, that there exists a complete system of intertheoretic connecting principles which reduces all of science to microphysics. If the laws of microphysics are to count as the only fundamental laws, then all other laws within the body of unified science, including the connecting principles, are "determined" by the laws of microphysics. On the other hand, the connecting principles will be incapable of deductive-nomological explanation on the basis of the fundamental laws, since they mention

objects and properties which are not countenanced by the laws of microphysics. So, the assertion that connecting principles "need no explanation" expresses the fact that their truth is determined by the truth of the fundamental laws, even though they are not deductive-nomologically explainable by means of the fundamental laws.²

Adoption of the Unity of Science hypothesis does not automatically commit us to the principle that the laws of the final microreducing science determine all the other laws, including the connecting principles. But this additional principle seems very reasonable. One reason for its naturalness, as noted in the above presentation of the N-D argument, is that otherwise it does not appear strictly correct to say that the higher-level branches of science are reducible to microphysics alone. Another reason for the naturalness of the principle is our feeling that the only truly fundamental laws within unified science would be those which deal with simple microphysical entities--entities from which the entities countenanced by higher-level theories are composed. Connecting principles, which mention higher-level entities and properties, seem too complex to count as basic laws of nature. Reasoning of this sort is central to Smart's development of the N-D argument, as evidenced in the following passage.

Certainly we are pretty sure in the future to come across new ultimate laws of a novel type, but I expect them to relate simple constituents: for example, whatever ultimate particles are then in vogue. I cannot believe that ultimate laws of nature could relate simple constituents to configurations consisting of perhaps billions of neurons (and goodness knows how many billion billions of ultimate particles) all put together for all the world as though their main purpose in life was to be a negative feedback mechanism of a complicated sort. Such ultimate laws would be like nothing so far known in science. They have a queer "smell" to them.³

The queer "smell" of these principles reflects our feeling that they are too complex to count as fundamental laws, since they deal with complex higher-level entities such as neurons and neuron-configurations. It seems that laws involving such entities should themselves be determined by the fundamental microphysical laws. (This consideration will be taken up again in chapter IV.)

The thesis that microreductive connecting principles are determined by the laws of microphysics, as we have seen, expresses the intuitive content of the claim that these principles are non-fundamental, and it provides a basis for their immunity from the need of explanation. We now face the question of explicating the notion of determination itself. To begin with we should consider the identity thesis as a possible explication, in accordance with advocates of the N-D argument.

The identity thesis initially seems to be the simplest and most natural way of explicating the assertion that the fundamental laws of microphysics determine the truth of microreductive connecting principles. Suppose, for example, that theory T_2 reduces to T_1 and that T_1 reduces to T_0 --where T_0 is microphysics. Suppose further that the property expressed by predicate P_2 of theory T_2 is identical to the property expressed by predicate P_1 of theory T_1 , and that this property is also identical to the property expressed by predicate P_0 of theory T_0 . Then the laws of T_0 which pertain to the property expressed by P_0 are also, ipso facto, the laws which pertain to the property expressed by P_1 and the property expressed by P_2 . Thus the connecting principles linking P_2 to P_1 and P_1 to P_0 are determined by the laws of T_0 in the simplest way possible: the laws pertain to the property expressed by P_0 , and this is the property expressed by P_1 and P_2 .

But the matter is more complex than it seems initially. The remainder of this chapter will be devoted

to investigating the complications that arise when we look more closely into the "determination" of connecting principles by laws of the final reducing science. I will argue in support of the following contentions: 1) The idea that connecting principles are determined by fundamental laws should be explicated in terms of a principle which differs from property identity (the "part-whole supervenience principle"). 2) Property identity is itself not sufficient to express the determination of connecting principles unless the criteria of identity are formulated in such a way that property identity entails the part-whole supervenience principle. 3) The supervenience principle does not imply that microreductively correlated properties are identical; hence we are not compelled by the thesis that fundamental laws determine connecting principles to conclude that the properties connected by these principles are identical. In subsequent chapters the part-whole supervenience principle, rather than the identity thesis, will emerge as the key metaphysical requirement which must be met in a complete reductive unification of science.

Section 3

It will be convenient for subsequent discussion to set forth together the various claims which underlie the N-D argument. These are best viewed as conditions of adequacy for the complete reduction of higher-level branches of science to microphysics. Let T_0 be the final reducing theory--some form of elementary-particle physics, presumably. Suppose also that we have available a set R of biconditional intertheoretic connecting principles, which reduce all true higher-level theories to theories immediately below them in the reductive hierarchy, and which thereby ultimately reduce all of these theories to T_0 . (Thus, we are assuming the truth of the Unity of

Science hypothesis. R is the complete set of connecting principles within unified science.)

Now consider two true higher-level theories, T_2 and T_1 , such that T_2 is reducible to T_1 . Let $R_{1,2}$, a subset of R , be the set of connecting principles which reduces T_2 to T_1 . It is claimed by advocates of the N-D argument that a fully adequate reduction of T_2 to T_1 requires more than the deducibility of T_2 from $T_1 \cup R_{1,2}$. In addition, the connecting principles must be immune from the need of explanation. And with regard to unified science as a whole, advocates of the N-D argument claim that a fully adequate reduction of the various higher-level theories to T_0 requires more than the deducibility of these theories from $T_0 \cup R$. The connecting principles R must themselves meet the closely related conditions of immunity from the need of explanation and non-fundamentality. We state these conditions as (E) and (F), for future reference.

- (E) The connecting principles R need no explanation.
- (F) The connecting principles R are non-fundamental laws.

The content of these two theses, in turn, is expressed intuitively by the determination thesis:

- (D) The connecting principles R are determined by the laws of T_0 alone.

As we saw, the thesis that connecting principles are based on underlying identities represents a seemingly natural explication of the intuitive claim (D).

- (I) Each connecting principle in R must be such that the individuals or properties it correlates are identical.

Statement (I), in asserting that the hierarchy of micro-reductive connecting principles all express identities, has the consequence that each higher-level theoretical

entity or property is ultimately identical to some complex entity or property countenanced by T_0 .

In evaluating statement (I), a key preliminary question which must be faced is whether true identity statements are necessarily true or only contingently true. Until recently, most philosophers have held that identity statements which are not knowable a priori--whose truth or falsity must be discovered empirically--are contingently true, or contingently false. Often the word 'contingent' is used as essentially equivalent to "not knowable a priori." Employing the word in this way, Smart maintains that statements asserting the identity of physical event-types with mental event-types are contingent.⁴ Causey, who formulates the generalized version of the N-D argument, maintains that all microreductive attribute-identity statements are contingent.⁵

Contrary to these claims, Saul Kripke has recently argued quite convincingly that identity statements containing two names are necessarily true if true at all. This is not to deny their empirical status, however. Kripke maintains that it is wrong to assert a direct connection between empirical knowledge and contingent truth, or between a priori knowledge and necessary truth. A statement like "Hesperus is Phosphorus" is empirical, since 'Hesperus' was originally used (presumably) to refer to the planet which appears on the western horizon just after dusk, whereas 'Phosphorus' was used to refer to the planet which appears on the eastern horizon just before morning. We certainly could not have known a priori that one and the same planet meets both descriptions. Still, the identity of Phosphorus and Hesperus is necessary, Kripke maintains.

Are there really circumstances under which Hesperus wouldn't have been Phosphorus? Supposing that Hesperus is Phosphorus, let's try to describe a possible situation in which it would not have been. Well it's

easy. Someone goes by and calls two different stars Hesperus and Phosphorus. It may be even under the same conditions as prevailed when we introduced the names 'Hesperus' and 'Phosphorus'. But are those circumstances in which Hesperus is not Phosphorus or would not have been Phosphorus? It seems to me that they are not....How would we describe this situation? He can't have pointed to Venus twice, and in the one case called it Hesperus and in the other Phosphorus, as we did. If he did so, then 'Hesperus is Phosphorus' would have been true in that situation too. He pointed maybe neither time to the planet Venus--at least one time he didn't point to the planet Venus, let's say when he pointed to Phosphorus. Then in that case we can certainly say that the name 'Phosphorus' might not have referred to Phosphorus. We can even say that in the very position when viewed in the morning that Phosphorus was not there--that something else was there, and that even under certain circumstances it would have been called 'Phosphorus'. But that still is not a case in which Phosphorus was not Hesperus.... For there couldn't have been such a case, given that Hesperus is Phosphorus.⁶

Two points in this argument call for comment. First, the alleged necessity of "Hesperus is Phosphorus" perhaps seems to violate our inclination to say that this statement "could have turned out false." Doesn't this inclination mean that there are possible worlds in which Hesperus is not Phosphorus? No, says Kripke. What underlies our assertion that things could have turned out differently is the following fact: "...given the evidence that someone has antecedent to his empirical investigation, he can be placed in a sense in exactly the same situation, that is a qualitatively identical epistemic situation, and call two bodies 'Hesperus' and 'Phosphorus', without their being identical. So in that sense we can say that it might have turned out either way."7 The possibility of someone's being in such a position and using 'Hesperus' and 'Phosphorus' in such a way does not affect the fact that we use the names in such

a way that Hesperus and Phosphorus are identical in all possible worlds.

Secondly, it should be noted that the argument is based on an appeal to our actual intuitions about counterfactual situations--an appeal to the way in which we would actually describe such situations. The systematic marshalling of intuitions constitutes a centrally important guideline for theory construction in philosophy. Moreover, Kripke seems to be right about our intuitions: when we imagine a possible world in which the planet over the eastern horizon in the morning is different from the planet over the western horizon in the evening, we are not inclined to describe the situation by saying "Hesperus is not identical to Phosphorus in that possible world." Rather, there is a strong inclination to say that at least one of the two planets in this possible world is not Phosphorus--and hence, not Hesperus either. Since Kripke's claim concerning our intuitions seems right, and since the systematic capturing of such intuitions is a major goal of theory construction in philosophy, his argument for the necessity of identity statements is a powerful one. So unless and until evidence to the contrary is presented, we are justified in accepting the conclusion that identity statements containing two names are necessarily true.

Kripke points out that this thesis is equivalent to the thesis that names are rigid designators. A rigid designator is a referring expression which designates the same entity relative to any possible world in which the entity exists. (In worlds where it doesn't exist, the expression either doesn't designate or else it arbitrarily designates some standard object, such as the empty set.) Not all designators are rigid. Hence, not all identities are necessary. Standard definite descriptions, like 'the president of the U.S. in 1973', are non-rigid designators: in various possible worlds, persons

other than Nixon are designated by this expression. Accordingly, it is only contingently true that Nixon is the President of the U.S. in 1973.

On the strength of Kripke's argument, we shall now adopt the thesis that identity statements involving names are necessarily true if true at all. This brings us back to the consideration of statement (I) above. In light of the necessity of identity generally, (I) entails the following claim concerning the objects and properties that are involved in microreductive connecting principles:

- (NI) Each connecting principle in R must be such that the individuals or properties which it correlates are identical in all possible worlds.

In chapter IV I shall question the truth of (NI), and hence the truth of (I) as well. But first I will consider a principle that is entailed by (NI), though weaker than (NI), and which itself seems to constitute a reasonable explication of (E), (F), and (D). I shall call it the part-whole supervenience principle. First, let me introduce the preliminary notion of a K -world: a possible world in which the only kinds of concrete indecomposable entities are those which exist in the actual world--viz., those countenanced by theory T_0 . As before, let R be the complete set of microreductive connecting principles in unified science. We can now state the part-whole supervenience principle, as follows:

- (S) In all K -worlds in which T_0 obtains, the connecting principles R also obtain.

This principle suggests that higher-level theoretical properties of structured wholes "supervene" upon the lower-level theoretical properties of the component parts. I.e., it's not possible for two structured wholes in two different K -worlds where T_0 obtains to have exactly the same structural decomposition into basic component

parts, and exactly the same T_1 properties, and yet differ in their T_2 properties. For example, it's not possible for two objects in two K-worlds where T_0 is true to be such that both have exactly the same micro-physical structure and exactly the same mean molecular kinetic energy, and yet differ in temperature. Similarly for property-correlation connecting principles generally.

Statement (S) is clearly entailed by (NI). (NI) asserts that the properties involved in connecting principles are identical in all possible worlds. A fortiori, the properties are lawfully correlated in all K-worlds where T_0 is true. So (NI) entails (S). On the other hand, (S) does not entail (NI), for the following reasons. (S) is consistent with statement 1:

- 1) There are worlds in which the connecting principles R do not obtain.

Now, properties that are not even co-extensive are not identical. So statement 1, in asserting that there are worlds in which the properties mentioned in the principles R are not even correlated, clearly implies statement 2.

- 2) There are worlds in which the properties mentioned in the principles R are not identical.

But statement 2 implies the denial of (NI). So (S) is consistent with 1, 1 implies 2, and 2 implies the denial of (NI). Thus (S) is consistent with the denial of (NI). Therefore, (S) does not imply (NI).

Section 4

We come now to the question of the adequacy of (S) as an explication of statements (E), (F), and (D). Does (S) serve to capture the intuitive force of the claim that the microreductive connecting principles R are "determined" by the laws of the final reducing theory T_0 ? And does it thereby capture the force of the asser-

tions 1) that the connecting principles are non-fundamental laws, and 2) that they need no explanation? I suggest that (S) captures the intuitive force of these claims quite clearly.

We'll first consider the adequacy of (S) as an explication of (D), the determination thesis. The notion of possible worlds arises quite naturally in connection with the idea of one set of laws L "determining" another set L' , especially when it is explicitly specified that L does not entail L' . (Recall that the laws of a reducing theory in general do not entail the microreductive connecting principles.) Intuitively, the truth of the set L' is determined by the set L just in case it's not possible for L' to be false if L is true.⁸ Entailment of L' by L is a paradigm case of determination in this sense, but perhaps not the only case. Perhaps there are cases of determination without entailment.

Statement (S) does not employ this possible-world model of determination in a full-fledged way, however. (S) is somewhat weaker than the assertion

(S') In all possible worlds in which T_0 obtains, the connecting principles R also obtain.

(S) is weaker because of its reference to K -worlds, which restricts the applicability of (S) to possible worlds in which the only kinds of indecomposable concrete entities are those which are instanced in the actual world. The point of this restriction is to rule out possible worlds in which the entities countenanced by higher-level theories have basic components over and above the kinds of basic entities countenanced by T_0 . For example, possible worlds are excluded in which persons are composed of immaterial Cartesian souls, in addition to the material components countenanced by the natural sciences. (T_0 might continue to be true in such worlds.) Similarly, possible worlds are ruled out in

which multicellular organisms contain entelechies, or "vegetative souls," in addition to the usual microphysical components. The restriction to K-worlds also excludes possible worlds in which things have material components which differ from the kinds of ultimate indecomposable entities which actually exist.

I suggest that the determination thesis (D) itself is weaker than (S'). It seems that (S), rather than (S'), represents the intended import of (D), for the following reason. The effect of the reference to K-worlds in (S) is to relativize the determination thesis to a fixed class of basic kinds of entities out of which structured entities may be composed--viz., the class of thing-kinds recognized by T_0 . When the advocate of the N-D argument asserts that the final reducing theory T_0 "determines" the connecting principles R, he is surely assuming that structured entities countenanced by higher-level theories are to be composed out of indecomposable entities of the sort which actually exist. His claim is not meant to apply to possible worlds in which there are basic entities of a radically different sort than in the actual world, even if these possible worlds also contain the same basic entities and the same fundamental physical laws as our world. For example, perhaps there are possible worlds in which 1) the entities postulated by natural science exist and the laws governing them are the laws which apply in our world, but 2) Cartesian souls also exist, and are components of persons. In such worlds, the mental states of persons might involve the soul-component. (For a Cartesian dualist, mental states are states of the mind, or soul, rather than states of the body.) And in these worlds the mental states of persons would not be "determined" by the underlying physiochemical states of their bodies, even if their bodies are physiochemically indistinguishable from ours. This is because minds are separate substances in these

worlds. Thus, the laws connecting Cartesian soul to body can differ from one world to another, even if the basic physical laws T_0 remain fixed. Consequently, psychophysical laws are fundamental laws in worlds containing mental substances. (Admittedly, the notion of a law correlating properties of a material substance with properties of a "co-present" immaterial substance is itself rather obscure. Nonetheless, the idea that various different mind-body correlation laws are possible relative to the same physical theory seems no more obscure than the overall Cartesian-dualist picture.)

The advocate of the N-D argument may or may not admit that there are possible worlds in which persons are unities of body and soul. But whether he does so or not, his thesis that psychophysical correlation laws are "determined" by the laws of natural science is not intended to apply to the Cartesian dualist case. It's intended to apply to cases in which persons are composed entirely out of the basic building blocks of the actual world--indecomposable physical entities of the sort countenanced by T_0 . And in general, the thesis that microreductive connecting principles R are determined by the final reducing theory T_0 is intended to apply only relative to the basic thing-kinds which are instanced in the actual world.

So the truth of (S) is sufficient for the truth of (D). I contend that (S) is also a necessary condition for (D). I can see no other way to capture the intuitive content of (D) except by means of a statement like (S). The notion of determination seems closely bound up with that of possibility, and (S) seems to express the way in which these notions are connected. What could it mean to assert that the truth of the principles R is determined by the truth of the laws of T_0 , if there were possible K -worlds in which the connecting principles R did not

obtain? If there are such worlds then surely the principles R are not determined by the laws of T_0 .

(S) also expresses the intuitive force of the claims (F) and (E)--that the connecting principles B are non-fundamental and that they are not in need of explanation. This is to be expected, since the notion of determination lies behind these assertions. It was suggested above that the principles R are non-fundamental just in case their truth is determined by the laws T_0 . So according to the explication of determination given by (S), the laws D are non-fundamental just in case they're true in all possible K-worlds where T_0 is true. (S) thus expresses the central intuitive element in the claim that the principles R are non-fundamental--viz., that the truth of these principles is itself guaranteed by the truth of the laws of T_0 .

To see how (S) captures the claim that the principles R need no explanation, we must briefly consider the notion of explanation itself. When we ask whether or not intertheoretic connecting principles require explanation, we are bringing into focus a somewhat broader notion of explanation than the conception which the deductive-nomological model is designed to capture. Intertheoretic connecting principles are problematic precisely because we can't obtain a deductive-nomological explanation of them on the basis of the reducing theory alone. What do we mean, therefore, when we ask whether or not they require explanation?

To answer this question, consider the deductive-nomological model in terms of possible worlds, rather than in terms of deducibility. A deductive-nomological explanation of an event shows why the event had to happen, given the relevant laws and antecedent conditions. I.e., it shows us that the event will occur in any possible world where the antecedent conditions are the same in all

relevant respects, and the laws are the same. Similarly, a deductive-nomological explanation of a law--in terms of deduction from more basic laws (and statements describing antecedent conditions, perhaps)--shows us that the nonbasic law is true in all possible worlds where the basic laws are true (and the relevant antecedent conditions, if any, are true). This conception of explanation is somewhat broader than the deductive-nomological model, but it seems to remain faithful to the spirit of the model.

Now consider microreductive connecting principles. In terms of the possible-worlds approach to explanation, these principles will need explanation just in case there is some need to show that they are true in all K-worlds where T_0 is true. They won't need explanation if we know a priori that they will be true in all such worlds. Here is where the supervenience principle, (S), comes in: given this principle, we do know a priori that microreductive connecting principles which are true in this world will be true in all possible K-worlds. So these connecting principles don't require explanation.⁹

Section 5

So far I have argued for the first and second of the three contentions I mentioned at the end of section 2. These are (1) that statement (D), and hence statements (E) and (F) as well, should be explicated in terms of the part-whole supervenience principle (S); and (2) that (S) does not imply that microreductively correlated properties are identical. I turn now to the third contention: that thesis (I) (the thesis of the identity of microreductively correlated properties) is itself not sufficient to guarantee (D), (F), and (E) unless the criteria for identity are such that this identity thesis entails the part-whole supervenience principle.

This contention brings us back to the question whether identity statements containing two names are necessarily true or only contingently true. We opted above for necessity, on the basis of Kripke's argument. And we saw that on this construal, thesis (I) entails the supervenience thesis (S). What I now wish to claim is this: if we do not opt for the necessity of identity statements that contain two names--if we subscribe to the view that such identities can be contingently true--then thesis (I) simply fails to capture the intuitions which underlie the N-D argument.

At first sight, the argument for contention 3 seems perfectly straightforward. We saw above that the supervenience principle expresses a condition that is necessary, as well as sufficient, for the truth of (D), (F), and (E). Hence, (S) must be entailed by an identity thesis which purports to reflect the intuitions expressed in these three theses. But if property-identity statements are taken to be only contingently true then (I) does not entail (S), and it thus does not guarantee the truth of (D), (F), and (E). (If statement (I) only expresses the claim that microreductively correlated properties are contingently identical, then (I) is consistent with the assertion that there are K-worlds, other than the actual world, in which T_0 obtains but the connecting principles R do not obtain. Thus (I), when so understood, does not entail (S).)

At this point the following objection might be raised, however. The argument just presented against construing contingent property identity as the basis for the truth of (D), (E), and (F), depends on the premise that (S) is a necessary condition for the truth of these statements. Yet this latter premise was advanced earlier on the basis that there seems to be no other plausible way to understand determination, non-fundamentality, and immunity

from the need of explanation. But there is another way to understand these notions, it might be claimed--viz., in terms of contingent identity.

This reply is a fair one. We must examine more carefully the question whether contingent property identity would provide an adequate explication of (D), (F), and (E). (We are waiving Kripke's arguments for the necessity of identity, for the moment.) Consider the notion of immunity from explanation. Advocates of the N-D argument have contended that we can account for the fact that microreductive connecting principles need no explanation by treating them as (contingent) identities, since identities never need explanation. But it's not true that identities never need explanation. Certain facts expressed by identity statements, like the fact that blue is the color of the sky and the fact that Nixon is the president of the U.S. in 1973, can and do receive explanation.

Treating microreductive connecting principles as contingent identities, therefore, will not by itself guarantee that they need no explanation. Furthermore, suppose that we were to deny explicitly that the connecting principles R are supervenient upon T_0 , while we also asserted that the higher-level properties mentioned in the statements R are contingently identical to their T_0 correlates. Under these circumstances, could we legitimately claim that the principles R "need no explanation"? It seems not. After all, if it were explicitly stated that the identities do not hold in other K-worlds where T_0 obtains, then the question would naturally arise as to why these identities obtain in our world, given that they do not obtain in these other worlds which are similar to ours in point of microphysical laws and basic indecomposable entities. If the identities are indeed peculiar to our world, then it seems just as legitimate to ask why they obtain as it is to ask why blue is the color of the sky. An explanation would not

be forthcoming, of course, since the contingent identities R would be logically primitive postulates within unified science. But their lack of explanation would not rest on immunity-in-principle from the need of explanation, any more than the unexplainability of the laws of T_0 rests on such a basis. Rather, they would have the status of fundamental laws within unified science.

So it turns out that the mere contingent identity of microreductive connecting principles cannot sustain the statements (E) and (F). Thus the conception of identity as a contingent relation, which we have already found reason to reject, is inadequate to reflect the intuitions underlying the N-D argument.¹⁰

Section 6

It has already been explained why (S) provides an adequate explication of (D), (E) and (F). This account does not require that microreductively correlated properties be identical, since (S) does not entail (NI) or (I). But perhaps we should briefly digress to consider identities specifically, with respect to the matter of explanation. Why are certain identities immune from the need of explanation, while others are not? Kripke's account of identity appears to give a satisfactory account of the matter. On his view, "Hesperus is Phosphorus" is necessarily true, since it is an identity statement containing two names. The statement "Hesperus is Kripke's favorite planet," on the other hand, is not a necessary truth, since it contains a definite description which designates different entities in different possible worlds. (Such terms are "non-rigid" designators.)

The following thesis suggests itself in light of Kripke's theory. An identity statement needs no explanation just in case it is true in all possible worlds.

We can explain why Hesperus is Kripke's favorite planet, because there are certain facts about this world which explain why he prefers Hesperus to all other planets. But it makes no sense to "explain why Hesperus is Phosphorus," because there are no facts about our world in virtue of which this assertion is true; the statement is true in any world. (Of course, it does make sense to explain why Hesperus is called "Phosphorus" in our world.) The truth of the identity statement in all possible worlds is no mystery, of course. It's just a reflection of the fact that " $(x)(x=x)$ " is necessarily true, together with Kripke's principle that names are rigid designators.

So Kripke's approach allows us to plausibly account for the fact that certain identities need explanation and others do not. This account constitutes an additional argument in favor of his theory, which can be added to his own argument cited earlier.

Our above discussion of (I) provides a third argument in favor of the necessity of identity statements containing two names. Notice that it is quite counterintuitive to allow the possibility that the connecting principles R of unified science are not determined by T_0 , if we assert that the correlated properties are identical. And it is equally counterintuitive to deny that (S) is a necessary condition for the truth of the determination thesis. The conclusion to be drawn is clear: we should reject as untenable the view that micro-reductively correlated properties are contingently identical, since contingent identity does not guarantee the truth of (S). If the properties are identical at all, the identity obtains in all possible worlds.

Section 7

In this final section we shall consider several frequently made claims about the nature of reduction which

are quite close in their intuitive content to the principles (E), (F), and (D). I will suggest that these additional claims can and should be understood in terms of the supervenience principle (S).

Virtually all the claims I have in mind are discussed by Carl Hempel, in a paper entitled "Reduction: Ontological and Linguistic Facets." Hempel begins by pointing out that one reason for the fascination the subject of microreduction has held for philosophers lies in the "ontological roots" of many questions concerning reduction--question such as these:

Are mental states nothing else than brain states? Are social phenomena simply compounds of individual modes of behavior? Are living organisms no more than complex physiochemical systems? Are the objects of our everyday experience nothing else than swarms of electrons and other subatomic particles? Or is it the case, as the doctrine of emergence would have it, that as we move from subatomic particles to atoms and molecules, to macroscopic objects, to living organisms, to individual human minds, and to social and cultural phenomena, we encounter at each stage various novel phenomena which are irreducible, which cannot be accounted for in terms of anything that is to be found on the preceding levels? ¹¹ (Emphasis mine)

It is certainly true that microreduction is often thought to establish that objects and processes described by higher-level theories are "nothing else than," or "no more than," or "simply compounds of," the objects and processes described by the lower-level reducing theory. In terms of the Unity of Science hypothesis, this position would amount to the view that all objects and processes described by true scientific theories are ultimately "nothing else than" complexes of objects and processes of the kind countenanced by microphysics.

Hempel goes on to consider what these theses might mean, in light of Nagel's characterization of interthe-

Filmed as received

without page(s) 43-45.

UNIVERSITY MICROFILMS.

oretic reduction in science. He contends that the theses imply at least that higher-level theoretical concepts can be extensionally characterized in terms of the concepts of the reducing theory. I.e., microreductive connecting principles must be biconditional in form, in accordance with strong Nagel-reducibility. He discusses the matter explicitly with respect to the thesis of mechanism: the view that biological characteristics "are nothing else than" physiochemical characteristics.

Even a set of connective principles powerful enough to permit a reduction of all the principles of a biological theory B to those of a physiochemical theory P need not permit the corresponding reduction of concepts; i.e., it need not provide a characterization of every concept of B by means of concepts of P. And only if this, too, had been accomplished could the biological theory be said to have been reduced to a chapter of the physiochemical one. For short of this, the reduction of B to P would amount to the establishment of a set of connecting principles which...would constitute additional biological laws and would thus simply expand the theory B, but would not make its conceptual apparatus dispensable.

Full reduction of concepts in this strict sense would require, for every term of B, a connective law of biconditional form, specifying a necessary and sufficient condition for its applicability in terms of concepts of P alone. Such a law could then be used to "define" the biological term and thus, theoretically, to avoid it. The "definition" of the temperature of a gas as the mean kinetic energy of its molecules illustrates this kind of connective principle...These principles do not, of course, express synonymies ascertained by meaning analysis, but empirical connections ascertained by chemical analysis.¹²

Several closely intercorrelated notions are introduced in this passage. A systematic set of biological/physiochemical connecting principles would permit a "reduction of concepts"--i.e., "a characterization of every concept of B by means of P." It would reduce the

biological theory B "to a chapter of" the physiochemical theory P. It would "make dispensable" the conceptual apparatus of B. And finally, it would allow us to "'define'" (note Hempel's scare-quotes) biological terms by means of physiochemical terms, thus "theoretically avoiding" the biological terms.

There does seem to be a strong inclination to use such expressions as these when we are discussing the status of microreductive connecting principles. Moreover, there also seems to be a close connection between these various claims about connecting principles and the ontological thesis that higher-level characteristics are "nothing else than" their microreductive correlates. But we must ask about the import of these various expressions. Is the existence of a pervasive set of biconditional connecting principles enough to guarantee their applicability? I suggest not.

Suppose that the Unity of Science hypothesis is true, but that the supervenience principle (S) is false. Now consider the various expressions Hempel uses. Would it be legitimate to say that the connecting principles R of unified science effect a "reduction of the concepts" of higher-level theories to those of lower-level ones, and ultimately to those of microphysics? Clearly not: the higher-level concepts would be separate and irreducible, since there would be some K-worlds where T_0 is true but these concepts are not even correlated with their lower-level counterparts from the actual world.

Would it be legitimate to say that the higher-level theories had been "reduced to chapters of T_0 " by means of the principles R? No: the principles must themselves be counted among the fundamental laws of our world, since we are hypothesizing that there are other K-worlds, similar to ours in point of microphysical laws T_0 , in which these principles do not obtain.

Could we assert that the principles R "make the conceptual apparatus of higher-level theories indispensable"? Here we must be careful. The concepts unique to higher-level theories will be dispensable in a certain pragmatic sense, since it will always be possible to employ their T_0 correlates in their place. But can they be said to be theoretically dispensable if (S) is false? It seems not: the reason why higher-level concepts are pragmatically dispensable is that the biconditional laws R obtain in the actual world. And these laws must themselves count as fundamental laws, if (S) is false. Thus, higher-level concepts would not be dispensable in a theoretically important sense.

Finally, would it be legitimate to assert that the principles R could be used to "'define'" the terms of higher-level theories by means of the terms of T_0 , and thus theoretically to avoid them? The answer is clearly negative, in light of the remarks just made concerning dispensability. We cannot "theoretically avoid" the concepts of higher-level theories if the connecting principles R themselves count as fundamental laws of nature.

These considerations indicate that the truth of (S) is a necessary condition for the truth of the various assertions put forward by Hempel concerning connecting principles. I suggest that (S) also expresses a sufficient condition for the truth of these assertions. Suppose now that the Unity of Science hypothesis is true, and that (S) is also true. The key fact to note is that the principles R are non-fundamental, in that their truth is determined by the truth of T_0 . So it's true that each higher-level theory "reduces to a chapter of T_0 ," since the laws of T_0 are the only fundamental laws.

As for the "dispensability" and "theoretical avoidability" of the concepts unique to higher-level theories, it must be admitted that the principles R are logically-primitive postulates within unified science. However,

since their truth is determined by the truth of T_0 , the applicability of higher-level predicates is itself determined by the applicability of their counterparts from microphysical theory. I.e., their theoretical avoidability consists in the fact that they do not figure in the fundamental laws of nature. This fact perhaps also justifies referring to the principles as 'definitions', if we emphasize the scare-quotes.

Concept reduction also receives a plausible explication in terms of the notion of fundamentality. As I just noted, the applicability of higher-level predicates is itself determined by the applicability of their T_0 correlates. Such determination, as explicated by (S), seems to correspond closely to the intuitive notion of concept reduction. So it appears that (S) (together with the Unity of Science hypothesis, which is presupposed by (S)) expresses a condition which is sufficient, as well as necessary, for the truth of Hempel's various claims about reduction.

This brings us back to the intuitive principle which we considered initially: the idea that higher-level qualities are "nothing else than" their T_0 correlates. This idea is captured by the closely-related theses just discussed. These theses, in turn, are captured by (S). So we can add the following statement to the list of intuitive principles which receive their explication in statement (S):

- (N) The higher-level properties mentioned in the connecting principles R are "nothing else than" the T_0 properties with which they are correlated.

It might be thought that the identity of micro-reductively correlated properties is a prerequisite for the truth of (N). But to see that this is not so, consider a phenomenal property such as having-the-sensation-of-red. Suppose this property has a precise neural (and

ultimately a precise microphysical) correlate. If this phenomenal property is identical to its microphysical correlate, then the two properties must have the same extension in all possible worlds. Now, it does seem true that if the sensation of red is "nothing else than" its microphysical correlate in the brain of an organism, then this phenomenal property and the related microphysical property will have the same extension in all K-worlds where T_0 is true. However, it seems clear that we can imagine other possible worlds where there are conscious beings who have the sensation of red without simultaneously instantiating the relevant microphysical property. Perhaps the beings have different indecomposable parts than we do, such as souls, or perhaps theory T_0 does not obtain in their world. We need not deny such possibilities when we assert that the sensation of red is "nothing else than" its microphysical correlate. So the truth of (N) does not require the truth of the identity thesis for microreductively correlated properties.

¹The generalized form of the N-D argument is presented in Causey. Sklar expresses sympathy with the argument insofar as it applies to sortal predicates.

²It will become clearer below why determination of the connecting principles eliminates the need to explain them. A full account of this fact requires prior explication of the notion of determination.

³Smart, p. 143.

⁴See Smart's reply to "Objection 2," *ibid.*, p. 147.

⁵Causey, p. 412.

⁶Saul Kripke, "Naming and Necessity," Semantics of Natural Language, ed. D. Davidson and G. Harmon (New York: Humanities Press, 1972), pp. 306-307.

⁷*Ibid.*, p. 308.

⁸Statements true in all possible worlds are thus "determined" by every statement. Similarly, logically-true statements are entailed by every statement.

⁹Jaegwon Kim has suggested to me that it may be equally plausible to say that (S) allows us to explain the connecting principles R on the basis of the fundamental laws T_0 . That is, since (S) asserts that the principles R are true in all K-worlds where T_0 is true, we can treat the conjunction of T_0 and (S) as an explanation of the connecting principles R. To me it seems more natural to say that (S) shows why the principles R need no explanation. But for my purposes it makes no real difference which way we phrase the matter. The important point is that the connecting principles R do not count as fundamental laws within unified science. (Non-fundamentality thus plays a more important role in my explication of the N-D argument than immunity from the need of explanation.)

¹⁰It has perhaps occurred to the reader that we can regard microreductively correlated properties as contingent identities, even if we accept Kripke's account of identity, by treating expressions referring to higher-level theoretical properties as non-rigid designators. This possibility is discussed in the appendix. For the present, note that the above observations carry over to such an approach. I.e., the proposed contingent identities would count as fundamental and in need of explanation, unless the supervenience principle obtains.

¹¹Hempel, p. 177.

¹²Ibid., p. 189.

III. PROPERTY IDENTITY AND MICROREDUCTION

Section 1

Although the intuitive principles underlying the N-D argument are best explicated in terms of the supervenience principle rather than in terms of property identity, the question of the putative identity of microreductively correlated properties still has considerable philosophical interest. We shall examine this question in the present chapter. I first propose a criterion for property identity, and I defend its plausibility. I then consider several ways in which the putative identity of microreductively correlated properties might be defended, relative to the proposed criterion. I argue that these approaches are inadequate, and I ultimately reject the identity thesis for microreductively correlated properties.¹

The difference between property correlations and property identities is not a straightforward empirical one. In an important sense, the scientist comes to the end of empirical discovery or hypothesis when he finds that pain and a certain neural property are strictly correlated, or when he hypothesizes that temperature is directly proportional to mean molecular kinetic energy. The question whether two such correlated properties are identical is not one which he can answer simply by further empirical investigation, or even by further abstract theory-construction within the scientific domains he deals with. This is because it is not antecedently clear what a property identity asserts, above and beyond the associated correlation statement. The criteria of property identity must be specified, and this task is

largely a matter for philosophical theory construction.²

When do two predicates attribute the same property to things? Synonymy of predicates seems too strong a criterion, since there are true property identities involving nonsynonymous expressions ("Blue is the color of the sky.") Co-extensiveness of predicates, on the other hand, seems much too weak: having a liver is not the same property as having a heart. Even nomological co-extensiveness is too weak: the Wiedemann-Franz law correlates electrical conductivity with thermal conductivity, but it has never been suggested that the two properties are identical.³

Our intuitions tell us that two predicates P_1 and P_2 express the same property if and only if there is no possible world in which there are objects to which P_1 applies but not P_2 , or vice-versa. P_1 and P_2 must have identical extensions not only in the actual world but in all possible worlds. I.e., they must have the same intension, where an intension of a predicate is understood as a function from possible worlds into sets of individuals. This is the sense of 'intension' that's employed in intensional logic. (We can, if we wish, go on to identify properties with intensions; properties themselves are then treated as functions from possible worlds to sets of individuals.⁴ But this additional step won't be required for purposes of my argument.)

The sameness-of-intension criterion (henceforth the SI criterion) conforms with our previously adopted principle that identity statements containing two names are necessarily true. This principle implies that sameness of intension is at least a necessary condition for property identity, for the following reason. Let ' p_1 ' and ' p_2 ' represent names for the properties expressed by the predicates P_1 and P_2 , respectively. In any possible world where the identity statement " $p_1 = p_2$ " is true, P_1

and P_2 will have the same extensions. (Clearly, two predicates do not express the same property in a given world if they don't even apply to the same individuals in that world.) But according to the necessity-of-identity principle, " $p_1=p_2$ " is necessarily true if true at all. Hence, " $p_1=p_2$ " is true only if in each possible world, P_1 and P_2 have the same extension. Thus, sameness of intension is a necessary condition for property identity. According to the SI criterion, it is a sufficient condition as well.

The SI criterion is met by synonymous predicates. It generally is not met by predicates which are merely co-extensive, since we usually can imagine a possible world in which two such predicates have different extensions. But what about contingent identity statements, like "Blue is the color of the sky"? It might be thought that the SI criterion is essentially equivalent to the view that property identity is a matter of predicate synonymy, and hence that contingent identities do not meet the criterion. Not so: the SI criterion results in a completely natural account of contingent property identities.

It will be worthwhile to examine the account in some detail, since the adequacy of the SI criterion depends upon it. The account makes use of Kripke's distinction between rigid and non-rigid designators, mentioned above. Recall that a rigid designator is a referring expression which designates the same thing relative to each possible world--provided, at least, that the thing exists in the world. A non-rigid designator is a referring expression which designates different things relative to different possible worlds. Contingent identity statements contain at least one non-rigid designator.

Consider the following two contingent identity statements:

- 1) Nixon is the president of the United States.
 - 2) Blue is the color of the sky.
- 'Blue' and 'the color of the sky' are rigid and non-rigid designators respectively, paralleling 'Nixon' and 'the president of the United States'. 1 and 2 are both contingently true because their non-rigid designators refer to different things in different possible worlds. They are true in this world, but not in all possible worlds.

Next, consider these two contingent identities, in which 'a' names the actual world:

- 3) The president of the United States is the president of the United States (at a).
 - 4) The color of the sky is the color of the sky (at a).
- These two statements are trivially true relative to world a, the world we happen to live in. But notice that by appending the phrase 'at a' to the definite descriptions in 1 and 2, we have converted these terms into rigid designators. 'The president of the United States (at a)' designates Nixon relative to any possible world in which Nixon exists, whether or not Nixon is president in that world. Likewise, 'the color of the sky (at a)' designates the property blue relative to any possible world, regardless of the color of that world's sky. Consequently, the following two statements, which follow from 1 and 3 and from 2 and 4 respectively, are necessarily true--i.e., true in all possible worlds.
- 5) Nixon is the president of the United States (at a).
 - 6) Blue is the color of the sky (at a).

We are thus led to the following important conclusion: contingent identity statements of any sort, including contingent property identities, correspond to certain necessary identity statements in an important way. This correspondence results from three facts:

- 1) it is always possible via trivially true statements like 3 and 4, to construct rigid designators from non-

rigid designators; 2) contingent identities always contain at least one non-rigid designator; and 3) it is always possible to obtain a necessarily-true identity statement from a contingently-true identity statement, by replacing all non-rigid designators by the appropriately constructed rigid designators.

We now consider property identities specifically. Earlier, I stated the SI criterion only for predicates: two predicates express the same property if and only if they have the same intension. But we also need to apply the criterion to referring expressions which designate properties, like the referring expressions in 2, 4, and 6. So we need a way of constructing predicates from referring expressions; the constructed predicates can then be compared with respect to the SI criterion. For rigid designators, it's clear how to proceed. The designator 'blue' refers to the property expressed by the predicate 'is blue', and the designator 'the color of the sky (at a)' refers to the property expressed by 'is the color of the sky (at a)'. So the method of constructing predicates from rigid property-designators is completely straightforward.

In the cases of non-rigid designators, however, the situation is more complicated. What property does 'the color of the sky' designate, when this term is used relative to the actual world, *a*? Statement 4 provides an a priori answer: the term designates the same property that 'the color of the sky (at a)' designates, viz., the property expressed by the predicate 'is the color of the sky (at a)'. So the latter predicate is the one which we construct from the term 'the color of the sky', for purposes of applying the SI criterion.

Why doesn't the 'the color of the sky' designate the property expressed by the predicate 'is the color of the sky'? First, the term is a non-rigid designator. So we would expect it to refer to different properties

relative to different possible worlds. But if it refers, relative to any possible world, to the property expressed by the predicate 'is the color of the sky', then it ceases to be a non-rigid designator at all. It refers to the same property relative to each possible world: the property possessed by blue things in worlds where the sky is blue, by green things in worlds where the sky is green, etc. Secondly, the property expressed by the predicate 'is the color of the sky' isn't even a color; it's a property possessed by different-colored things, depending on the color of the sky in the possible world we happen to consider.⁵ Clearly, we are not identifying this property with the color blue when we say "Blue is the color of the sky." Rather, we're identifying the color blue with the color of the sky in the actual world--i.e., with the property expressed by 'is the color of the sky (at a)'.

So the referring expressions in "Blue is the color of the sky" are understood to designate, respectively, the properties expressed by the predicates 'is blue' and 'is the color of the sky (at a)'. These two predicates are not synonymous, but they do have the same intension. In any possible world, something is blue if and only if it is the color of our world's sky. So it turns out that the SI criterion does indeed underlie the contingent identity "Blue is the color of the sky," since the criterion is met by the predicates corresponding to the terms 'blue' and 'the color of the sky'.

This same line of reasoning applies to any contingent property identity, since the argument doesn't depend upon any peculiarities of the statement "Blue is the color of the sky." Consider another example:

7) Truth is Plato's favorite property.

An argument exactly analogous to the one above will show that the predicate 'has Plato's property (at a)' is

understood as expressing the property designated by 'Plato's favorite property'. Of course, the predicate 'is true' is understood as expressing the property designated by 'truth'. As in the earlier example, the SI criterion is met by these two predicates: in any possible world, something is true if and only if it has Plato's favorite property (at a). Statement 7, and contingent property identities in general, meet the SI criterion.

We can understand why the SI criterion is met when we notice that non-rigid property-designators typically make mention of certain specific objects: the sky, Plato, etc. This feature of a non-rigid designator is precisely what makes it non-rigid; it depends for its reference on the nature of the objects it mentions. So it designates different properties relative to different worlds, since the nature of the objects it mentions varies from one world to another. This object-dependence is held constant, so to speak, when we construct a rigid designator from a non-rigid designator by appending the phrase 'at a'. Accordingly, the predicates 'is blue' and 'is the color of the sky (at a)' do not diverge in their intension. Relative to any possible world, the extension of the latter predicate still depends on the nature of our world's sky.

We can, as I mentioned above, go on to identify properties with intensions, thereby treating properties as functions from possible worlds to sets of individuals. This approach is attractive for several reasons. To mention two prominent ones, 1) it allow us to adopt an ontologically economical approach to properties (of course we may wish ultimately to adopt a similar ontological-reductive approach to possible worlds as well), and 2) it conforms with, and gives a theoretical basis for, the SI criterion. But whether or not we identify properties with intensions, my present claim is that the SI criterion does capture our fundamental intuitions about property

identity, including contingent property identities.

Finally, if the above discussion of contingent property-identities is correct, then there should be two possible formal renderings of the statement "Blue is the color of the sky." One formal version would reflect our actual normal usage of this statement. The other would represent the assertion that the property blue is identical to the property being sky-colored--i.e., the false assertion that blue is identical to the property possessed by green things in possible worlds where the sky is green, by orange things in worlds with an orange sky, etc. I shall now briefly indicate the formal representations of these two alternative interpretations.

For simplicity we will treat properties as functions from possible worlds to sets of individuals. (This treatment can be avoided, by introducing notation to represent the property which corresponds to a given function from possible worlds to sets of individuals.) F is the set of such functions. Let ' f ' be a variable function-letter ranging over the functions in F , ' w ' a variable ranging over possible worlds, ' s ' a name for the sky, and ' a ' a name for the actual world. ' I ' is the definite-description operator. T_w is the truth-operator; it operates on a statement to form the assertion that the given statement is true in the possible world w . ' Mem ' is the predicate for set-theoretical membership, ' Blu ' the predicate for 'is blue', and ' $Colorsim$ ' the predicate for 'is color-similar to'.

We designate the property blue, or blueness, as follows.

(If) ($(w)(x)(Mem(x, f(w)) \text{ iff } T_w(Blu(x))$))

Now consider the property expressed by 'is the color of the sky', in the sense that's not specific to our world's sky--viz., the property which a thing possesses at a world just in case it has the color of that

world's sky. We designate this property this way:

(If)((w)(x)(Mem(x, f(w)) iff $T_w(\text{Colorsim}(x,s))$))

Now for the second property expressed by 'is the color of the sky'--the property which a thing possesses at a world just in case it's blue in that world. Let C be the class of color properties--i.e., the subset of F which contains the properties blueness, redness, yellowness, etc. We represent the second property expressed by 'is the color of the sky' as follows:

(If)(Mem(f,C) & (x)(Mem(x,f(a)) iff $T_a(\text{Colorsim}(x,s))$))

The identity statement containing the first and third of these expressions is true, and the identity containing the first and second is false.

Section 2

We return now to microreductive identity statements of the sort which are advocated by proponents of the N-D argument. If we discover a correlation law connecting pain and some physiochemical property P, may we then go on to identify these two properties? According to the SI criterion, we may do so only if the predicates 'is in pain' and 'has P' have the same intension--viz., the same extension in all possible worlds.⁶ But it is clear that they do not have the same intension, since we can imagine possible worlds in which there are conscious beings which experience pain, with all its phenomenal features and behavioral symptoms, but which differ radically from us in their "neurophysiology." Perhaps they are biological organisms which experience pain in a physiochemically different way than we do (such organisms may well exist in the actual world), or perhaps they are beings which are physically so different from us that we wouldn't call them organisms at all.

The same is true for temperature and mean kinetic energy. We can imagine possible worlds in which macro-

scopic objects obey all the laws of thermodynamics, but in which the objects are composed of different entities, governed by different laws, than the entities and laws postulated by current chemistry and physics.

These considerations seem applicable to virtually all cases of microreduction. Consider any two theories T_1 and T_2 , where T_2 is microreduced to T_1 . Relative to T_1 , the entities and properties of T_2 are "macroscopic." So it seems that it will always be possible to imagine possible worlds in which 1) there are entities "macroscopically" similar to those of T_2 --entities which obey all the laws of T_2 , but 2) the micro-theory underlying T_2 is radically different from T_1 . So according to the SI criterion, the connecting principles used in microreduction will not be based on identity of the correlated properties.

Three potential lines of reply against this argument come to mind. The first is to deny that it's always possible for entities to be "macroscopically" similar to those of T_2 and yet to be composed of micro-entities whose governing laws differ from T_1 . The second is to deny that terms designating T_2 properties are rigid designators, and to claim that T_2 properties are contingently identical to the correlated T_1 properties, just as the color of the sky is contingently identical to blue. For example, it might be held that "pain" designates neural property P in our world, but that it designates different properties in other possible worlds. The third line of reply admits that names of T_2 properties are rigid designators, and it admits that entities can be macroscopically similar to those of T_2 while having a radically different micro-structure, but it asserts that such entities don't really possess T_2 properties; the appropriate T_1 property is held to be a criterion for the actual possession of the correlated T_2 property. For example, it might be held that entities in other possible

worlds don't really have a certain temperature unless they are composed of molecules, and unless the molecules have the appropriate mean kinetic energy.

I know of no argument in favor of the first line of reply, so I will not discuss it further. The second line has been advocated explicitly by David Lewis, with respect to mental properties such as pain. "The definitive characteristic of any (sort of) experience is its causal role, its syndrome of most typical causes and effects," he says.⁷ So if the materialist is right in maintaining that the relevant causal roles are occupied by certain physical states, then we should identify an experience-type with the correlated physical state-type, according to Lewis. The identity will be contingent, of course, because the causal role definitive of the experience will be occupied by different state-types in different possible worlds. Thus, experience-predicates like 'pain' must be viewed as non-rigid designators. They designate different properties in different possible worlds.

There are several problems with Lewis' argument. First, it seems clear that causal roles are not the definitive characteristic of many sorts of experiences. Pain is an excellent example. The definitive characteristic of pain is no more nor less than the feel of pain--the phenomenal event with which we are acquainted non-inferentially in our case, and which we inferentially attribute to others. To be sure, we learn to use the concept "pain" by learning its typical causes and effects, but the fact remains that the "raw feel" associated with pain is the key criterion which determines whether pain is really present in an organism or not. We can imagine beings which undergo states that have the causal role of pain, but which don't feel pain--perhaps they don't feel anything. Such beings could perhaps be said to "exhibit pain behavior," but it seems quite clear that they

wouldn't feel pain.

But let's set aside the problems of the phenomenal aspect of our mental life. It does seem true that certain mental states, particularly some that involve intentionality, have definitive causal roles. Wants and beliefs, for example, perhaps can be viewed as states which tend to cause certain appropriate actions. (Thus, if a person wants something and he believes that a certain action A will bring about the want, then this want-belief complex is likely to result in his doing A.) We come now to the second objection to Lewis' argument: it doesn't follow from the fact that the definitive characteristic of a certain mental property is its causal role, that this mental property is contingently identical with a physical property which happens to occupy the same causal role.⁸ On the contrary; why can't we say that the typical causes and effects of a particular mental property are also the typical causes and effects of the distinct nomologically correlated physical property? (A particular action, for example, might be caused by a particular want and by a neural event simultaneous with the occurrence of the want.)

Lewis himself considers a dual causation view of this sort, which he labels a brand of epiphenomenalism. He contends that it is unsatisfactory, for two reasons: 1) "At best, this position yields nonphysical experiences alongside the physical experiences, duplicating them, which is not what its advocates intend"; 2) "The position exploits a flaw in the standard regularity theory of cause. We know on other grounds that the theory must be corrected to discriminate between genuine causes and the spurious causes which are their epiphenomenal correlates. (The "power on" light does not cause the motor to go, even if it is a lawfully perfect correlate of the electric current that really causes the motor to go.)"⁹

The dual causation view I'm suggesting, and the relation of this view to traditional epiphenomenalism, will be discussed at length in chapter VI. For the present, I'll only sketch a brief reply to Lewis' two points. His first point is that the position yields non-physical experiences alongside physical experiences, which is not what's intended. But why should the dual cause theorist admit that the physical event-types in question are experiences? The theorist admits that a physical event-type can occupy the causal role that's occupied by an experience-type, but he denies that the physical event-type is itself an experience. So Lewis is simply asserting what the dual causation view denies.¹⁰

Lewis' second point invokes the distinction between genuine and spurious causes. He asserts that mental causes, on the proposed dual causation view, are really spurious causes. I.e., they are themselves caused by the relevant physical events (just as the electric current causes the "power on" light to switch on), and they are not the genuine causes of any subsequent actions or events (just as the "power on" light does not cause the motor to go.) This point would perhaps have some validity against a dual causation view which held that mental events are themselves caused by the simultaneous, nomologically correlated physical events. (The "power on" light is a spurious cause of the motor's running because it is itself caused by the real cause of the motor's running--the electric current.) But the dual causation theorist can deny that physical events cause the nomologically correlated mental events. In support of his denial, he can invoke the part-whole supervenience principle, which in effect asserts a necessary connection between the physical and the mental of the sort which does not obtain between cause and effect. The "spurious cause" argument could thus be avoided. (More on this point in chapter VI.)

Since Lewis has failed to rule out the dual causation view, he has not shown that the alleged contingent identity of mental event-types with physical event-types is a consequence of his premise that mental event-types have definitive causal roles. And in any case, I have already argued that we have independent reason to reject the premise as it applies to many mental event-types. I come now to a third difficulty with Lewis' argument. He is forced to deny the contention that pain is identical to the attribute of having pain. This is because he takes 'pain' to be a non-rigid designator of whatever physical property happens to occupy the causal role that's (allegedly) definitive of pain, while he admits that 'the attribute of having pain' is a rigid designator of that property which belongs, in any world, to whatever things have pain in that world. This approach is counterintuitive. Intuitively, it seems clear that 'pain' designates the same property as 'the attribute of having pain'--viz., that unique property which is expressed by the predicate 'is in pain'. Pain is the property that's possessed, in any given possible world, by the individuals to which the predicate 'is in pain' applies.

Compare 'pain' to 'the color of the sky'. The latter is a non-rigid property designator, as we saw. But the reason is clear: the expression is not a name but a definite description, and it mentions a specific entity (the sky) whose characteristics vary from one world to another. Similarly for 'Plato's favorite property'. 'Pain', on the other hand, is a name, and it makes no mention of any specific objects whose characteristics change from one world to another, such as the sky or Plato. So it seems reasonable to treat 'pain' as a rigid designator. This approach not only conforms to our intuitions, but it also reflects the sharp dissimilarity between 'pain' and 'the color of the sky'.¹¹

Although I have rejected Lewis' position, I wish to emphasize a certain basic intuition that is reflected in his view, which I too seek to capture. The intuition, which was discussed at the end of chapter II, is this: although it's possible for mental properties like pain to correspond to different physical properties than those to which they correspond in the actual world, it's true nonetheless that in the actual world, and in possible worlds that are "nomologically accessible" from the actual world, pain is "nothing more than" the correlated physical property. Lewis captures this intuition in the contention that 'pain' is a non-rigid designator, whose designatum in the actual world is neural property P. I contend, on the other hand, that the intuition is successfully captured by the part-whole supervenience principle, as applied to psychophysical correlation laws. If pain correlates with neural property P in all worlds in which the laws of natural science obtain and in which the only indecomposable entities are those postulated by natural science, then it seems reasonable to assert that pain is "nothing more than" property P.

Section 3

I turn now to a second potential reply to my argument that microreductively correlated properties are not identical. This reply has not been formulated as an explicit argument, as far as I know, but it is at least suggested by Kripke,¹² and perhaps also by Putnam.¹³ The reply is this: It's true that the SI criterion must be met if microreductively correlated properties are to count as identical, and it's true that names of properties are rigid designators. Thus, identities like "Pain is property P" and "Temperature is mean kinetic energy" are true in all possible worlds if they're true at all. But in fact, these assertions are true in all

possible worlds. An object doesn't really have a particular temperature unless it's composed of molecules, and these molecules have the proper mean kinetic energy. Likewise, an entity doesn't really have pain unless it has the neural property P. Now, perhaps there are possible worlds in which certain entities are macroscopically similar to those in the actual world, but microscopically quite different. Thus, there might be objects with properties that are macroscopically indistinguishable from temperature or from pain, and which don't correlate respectively with mean molecular kinetic energy and with property P. But these macroscopic properties would not actually be temperature and pain, since genuine temperature is mean kinetic energy and genuine pain is property P.

This position can be formulated in general terms, as follows. Suppose theory T_2 reduces to theory T_1 via connecting principles $R_{1,2}$. Let P_1 and P_2 be two properties whose correlation is asserted by the principles $R_{1,2}$, such that P_1 is a property recognized by T_1 and P_2 is a property recognized by T_2 . The position we are considering asserts that the presence or absence of P_1 is the criterion, in all possible worlds, for the presence or absence of P_2 . No entity in any possible world will really possess property P_2 unless it possesses property P_1 , no matter how much it might appear that the entity possesses P_2 in the absence of P_1 or lacks P_2 in the presence of P_1 . I shall call this view the "criterial identity thesis," since it treats instantiation of P_1 as the definitive criterion, in all possible worlds, for the genuine instantiation of P_2 .

This approach is quite implausible when applied to mental event-types. Take the case of phenomenal mental event-types, like pain or the sensation of red. Suppose there is a possible world in which certain entities have a radically different microstructure than living organ-

isms in the actual world, but these entities have conscious experiences qualitatively indistinguishable from our experiences of pain. Then surely we would wish to say that the entities do have pain, even though their pain does not correspond to physical property P. Now consider mental event-types of the more "cognitive" sort, like beliefs. Suppose our imagined entities have the same intelligence level and the same behavior patterns as humans. Then presumably it would be no more difficult to ascribe beliefs to them that it is to ascribe beliefs to one another. And we would surely say that the entities really possess the beliefs we ascribe to them, even if such beliefs in ourselves are nomologically correlated with certain physical states which these alien entities never undergo.

But although the viewpoint under consideration is implausible when applied to mental-physical property correlations, it does have a certain prima facie appeal when applied to microreductively correlated properties like temperature and mean kinetic energy. Kripke, for example, seems willing to say that the identity of temperature with mean kinetic energy holds in all possible worlds, since an object in a possible world doesn't really have a particular temperature unless it has the relevant mean kinetic energy. Being identical to mean kinetic energy is alleged to be an essential property of temperature.

What arguments are there in favor of this view? One argument might be that scientists themselves are inclined to say that temperature is "identical to" mean kinetic energy, that surface tension is identical to molecular attraction, and so on. It has been suggested that this usage, and the intuition that underlies it, constitute reasons for maintaining that such properties really are identical.¹⁴ It must be granted, I think, that they constitute prima facie reasons. But they are

not conclusive in themselves, especially if we can offer 1) some alternative way of accounting for the intuition underlying this view, and 2) a linguistic expression, other than 'is identical to', which reflects the underlying intuition. I submit that both tasks have been accomplished above. The key intuition involved here is captured in the assertion that temperature is "nothing but," or "nothing else than," mean kinetic energy. I have argued that this assertion is explicated by the part-whole supervenience principle; the additional supposition of property identity is therefore superfluous.

A second argument which might be advanced in support of the view in question is the N-D argument itself. The argument could be used as follows: If we stipulate that objects in other possible worlds don't really have a particular temperature unless they have the appropriate mean molecular kinetic energy, then we can thereby account for the fact that the correlation law involving temperature and mean kinetic energy needs no explanation. Otherwise, the law has the same status as fundamental laws of nature.

The reply to this argument is by now familiar: the intuitive premises underlying the N-D argument are effectively reflected in the part-whole supervenience principle, and this principle does not entail property identity.

A third potential argument involves considerations of ontological parsimony. It might be claimed that the stipulative identification of microreductively correlated properties effects a systematic reduction in the kinds of properties we must countenance, thereby achieving significant simplification of theory. We need not deny that this argument provides a degree of prima facie support for the contention that we should stipulatively identify higher-level properties with their microphysical counter-

parts. However, the putative gain in simplicity is not a conclusive consideration in itself. It must be weighed against any potential theoretical disadvantages of stipulative property-identification. Some important disadvantages will emerge in the next section.

A fourth argument for the criterial identity thesis, some might contend, can be developed on the basis of Kripke's discussion of alleged theoretical identities in "Naming and Necessity."¹⁵ Although Kripke specifically rejects the identity thesis with respect to psychophysical correlation laws involving properties such as pain, much of his discussion suggests that he looks favorably on the thesis as applied to microreductive connecting principles at the level of natural science. But he never specifically argues for the thesis, even in the case of connecting principles which reduce one natural science to another. It will be worthwhile to briefly examine his discussion, in order to see why it does not really apply to microreductive connecting principles.

He first considers the nature of the substance gold. He asks whether yellowness is part of the concept of gold.

Could we discover that gold was not in fact yellow? Suppose an optical illusion were prevalent, due to peculiar properties of the atmosphere in South Africa and Russia and certain other areas where gold mines are common. Suppose there were an optical illusion which made the substance appear to be yellow; but, in fact, once the peculiar properties of the atmosphere were removed, we would see that it is actually blue...Would there on this basis be an announcement in the newspapers: "It has turned out that there is no gold. Gold does not exist. What we took to be gold is not in fact gold."?... It seems to me that there would be no such announcement. On the contrary, what would be announced would be that though it appeared that gold was yellow, in fact gold has turned out not to be yellow, but blue. The reason is,

I think, that we use 'gold' as a term for a certain kind of thing...The kind of thing is thought to have certain identifying marks. Some of these marks may not really be true of gold. Further, there might be a substance which has all the identifying marks we commonly attributed to identify the substance of gold in the first place, but which is not the same kind of thing, which is not the same substance...Such a thing is, for example, as we all know, iron pyrites or fool's gold.¹⁶

Yellowness, then, is not a part of the concept of gold. We could discover that gold is not really yellow. The term 'gold' is used to name a certain natural kind, and it might turn out that the natural kind does not (or sometimes does not) possess the characteristics we thought it had when we first began referring to it as a thing-kind.

Kripke next turns to the question of delimiting natural kinds. On what basis do we demarcate the natural kind we call 'gold'? The answer, as we might expect, is provided by scientific theory: gold is the element with atomic number 79. Having atomic number 79 is thus an essential property of gold--i.e., the statement "Gold has the atomic number 79" is necessarily true.

Given that gold does have atomic number 79, could something be gold without having the atomic number 79? Let us suppose the scientists have investigated the nature of gold and have found that it is part of the very nature of this substance, so to speak, that it have the atomic number 79...Consider a possible world. Consider a counterfactual situation in which, let us say, fool's gold or iron pyrites was actually found in various mountains in the United States, or in areas of South Africa and the Soviet Union. Suppose that all the areas which actually contain gold now, contained pyrites instead, or some other substance which counterfeited the superficial properties of gold but lacked its atomic structure. Would we say, in this counterfactual situation,

that in that situation gold would not even be an element (because pyrites is not an element)? It seems to me that we would not...It would not be gold; it would be something else.¹⁷

Kripke discusses several other examples, including that of heat. Heat is in a class somewhat apart from most of his examples, in that it is a process-kind rather a thing-kind or substance-kind. In terms of the conception of events and processes we adopted in chapter I, heat is a property; specific instances of this process consist in the having of this property by certain things or substances. Kripke's discussion of heat parallels what he said about gold: heat is a natural kind, and empirical scientific discovery has demarcated the natural kind as molecular motion. Hence the statement "heat is molecular motion: is necessarily true. Science has discovered that being molecular motion is an essential property of heat.

Can we then imagine a possible world in which heat was not molecular motion? We can imagine, of course, having discovered that it was not. It seems to me that any case which someone will think of...would actually be a case in which some creatures with different nerve endings from ours inhabit this planet (maybe even we, if it's a contingent fact about us that we have this particular neural structure); and these creatures were sensitive to something else, say light, in such a way that they felt the same thing that we feel when we feel heat. But this is not a situation in which, say, light would have been heat, or even in which a stream of photons would have been heat, but a situation in which a stream of photons would produce the characteristic sensations which we call sensations of heat...What characteristically goes on in these cases of, let's say, 'heat is molecular motion'? There is a certain referent which we have fixed, for the real world and for all possible worlds, by a contingent property of it, namely the property that

it's able to produce such and such sensations in us...One doesn't know a priori what physical phenomenon, described in other terms--in basic terms of physical theory--is the phenomenon which produces these sensations..We've discovered eventually that this phenomenon is in fact molecular motion. When we have discovered this we've discovered an identification which gives us an essential property of this phenomenon. We have discovered a phenomenon which in all possible worlds will be molecular motion--which could not have failed to be molecular motion, because that's what the phenomenon is.¹⁸

In sum, Kripke has argued that "natural kind" terms have scientifically-established criteria of applicability, which may differ substantially from the macroscopic qualities by virtue of which we initially "picked out" or identified the natural kind. The essential properties of gold are delineated scientifically, in terms of molecular structure, rather than in terms of its macroscopic qualities. The essential properties of heat are delineated scientifically in terms of the physical processes which actually occur when people feel heat-sensations, rather than in terms of the presence (or potential presence) of heat sensations in a sentient organism.

Having cited the key passages in Kripke's argument, we can now assess its relevance to the criterial identity thesis. The thrust of his argument is that natural kind terms--whether they be sortal terms like 'gold' or property terms like 'heat'--have scientific criteria of applicability. The most obvious application of this contention is to terms which do not (at least initially) have a place within scientific theory itself. After all, terms within scientific theory itself do have scientific criteria of applicability: they are scientific terms, and the theory in which they occur will itself provide criteria of applicability. (Otherwise, the terms would

play no useful role in the theory.)

We are now in a position to see why Kripke's discussion has no clear application to the criterial identity thesis. This thesis concerns the status of inter-theoretic connecting principles. The higher-level predicates in these principles are already a part of scientific theory, albeit higher-level theory. Hence, these predicates already have scientific criteria of applicability, by virtue of their theoretical status. Kripke's own example of heat will serve to illustrate the point I am making. His basic claim is that heat is a natural kind: phenomenal hotness is used to pick out the phenomenon in our world, but it is not an essential property of heat. The essential properties of heat, he says, must be specified at the theoretical level. But the crucial question is this: which theoretical level? Kripke turns directly to the molecular level, in his assertion that heat is molecular motion. But the concept of heat occurs at a higher theoretical level than this--viz., the level of macroscopic thermodynamics. Now, Kripke's argument does seem to require us at least to say that the criteria for proper applicability of the term 'heat' are those provided by thermodynamic theory. But his argument does not require us to move below the level of thermodynamics, as far as I can tell. It would seem that we can regard heat as a natural kind on the basis of its status as an explanatory concept within thermodynamics.

Earlier in this chapter I denied that microreductively correlated properties are identical. I based my denial on the claim that we can imagine possible worlds in which a higher-level theory is true even if the fundamental reducing theory is false. I.e., various of our higher-level theories can obtain in worlds where T_0 does not obtain. This contention amounts to the position that there are possible worlds where the higher-level theories

have a different "reduction basis" than theory T_0 . Evidently, Kripke's argument regarding natural kinds does not rule out the view that there are possible worlds which contain an alternative micro-level reduction basis for higher-level theory. In footnote 68 of "Naming and Necessity," Kripke seems to acknowledge this point:

Some people have been inclined to argue that although certainly we cannot say that sound waves 'would have been heat' if they had been felt by the sensation which we feel when we feel heat, the situation is different with respect to a possible phenomenon, not present in the actual world, and distinct from molecular motion. Perhaps, it is suggested, there might be another form of heat other than 'our heat', which was not molecular motion; though no actual phenomenon other than molecular motion, such as sound, would qualify. Similar claims have been made for gold and for light. Although I am disinclined to accept these views, they would make relatively little difference to the substance of the present lectures.¹⁹

In sum, Kripke's basic point is that terms like 'gold', 'cat', and 'heat' designate natural kinds. Consequently, definitive criteria for their applicability are provided by scientific theory, rather than by a cluster of concepts which might seem to typify the natural kind prior to the discovery of its key theoretical properties. But this point has no apparent relevance to the metaphysical status of microreductive connecting principles. We can accept Kripke's basic claim, and still maintain that higher-level scientific theories are true in possible worlds where theory T_0 is not true. There might be other forms of heat besides "our heat," since thermodynamic theory might have an alternative micro-level reduction basis in some world where T_0 is false. So Kripke's discussion does not support the criterial identity thesis.

Section 4

In this section we continue our discussion of the criterial identity thesis. Of the various arguments we have examined in relation to this position, only the argument from ontological parsimony lends any clear support to the thesis. I turn now to an argument against the thesis, which I think will outweigh the prima facie positive support which comes from considerations of parsimony.

We first consider the case of mental properties like pain. As noted earlier, the proposal that a property isn't really pain unless it's neural property P doesn't even seem initially plausible, because the definitive characteristic of pain is quite clearly its phenomenal "raw feel," rather than the presence of the neural property which happens to correlate with this feel. When we consider possible worlds in which entities have phenomenally similar experiences, we are strongly inclined to say that these entities really do feel pain, although it is differently realized than in the actual world. But we will ignore this problem for the present, while focusing on the second argument against the proposal to treat property P as the criterion of genuine pain. This argument will be rather similar structurally to our earlier argument against the proposal to treat property identities involving two names as contingent rather than necessary. It was claimed that the latter proposal is defective because property identity so construed cannot guarantee the truth of the part-whole supervenience principle. For example, saying that pain is contingently identical to property P does not imply that pain is identical to property P in all worlds which resemble ours in point of fundamental physical laws and fundamental

indecomposable entities. And since the supervenience principle itself embodies the intuitions which are supposed to be captured by an identity thesis (viz., the intuitions expressed as (N), (D), (E), and (F)), failure to imply this principle constitutes good reason for rejecting the view that property identities involving two names are sometimes only contingently true.

The proposal for property identity that we're presently considering, on the other hand, is based on the SI criterion, and it therefore involves the necessity of identity. Moreover, the supervenience principle is implied by the present proposal: if property P is the criterion of genuine pain, then no property in any possible world will count as pain unless it's property P. Still, this proposal runs into trouble with respect to part-whole supervenience, because of the phenomenal aspect of pain. The trouble arises as follows: Even though we are granting (for the sake of argument) that it's not unreasonable to say that an event isn't an instance of genuine pain unless it's an instance of P, it seems true nevertheless that an instance of P shouldn't count as an instance of pain unless it is experienced as painful--unless it hurts. I.e., even if not all occurrences of the phenomenal "raw feel" associated with pain are instances of genuine pain, it does seem that all instances of genuine pain should include this raw feel. Now, in this world they certainly will include the raw feel, because we would not assert that genuine pain is identical to property P unless we believed that the phenomenal feel we associate with pain is present if and only if property P is present. But now consider possible worlds, other than the actual world, which resemble ours in point of fundamental indecomposable entities. There is no guarantee that in these worlds, instances of P are accompanied by the phenomenal feel of pain. On the contrary; since property P--and not the phenomenal feel--

is supposed to be the criterion of genuine pain, property P might be associated with the phenomenal feel of pleasure in some of these worlds, or with no feel at all, or with radically different feels at different times. Conversely, the feel of pain might be associated in these worlds with physical occurrences other than the occurrence of P. In short, taking property P to be a criterion of genuine pain does not guarantee that the feeling of pain is supervenient upon occurrences of P.

This result is highly counterintuitive. The assertion that pain is identical to property P seems to imply that the phenomenal sensation associated with pain is determined by P--that there's nothing more to this phenomenal occurrence than the physical property with which pain is identified. And this apparent implication is crucial, if the identity thesis is to serve as a full explication of the assertion that pain itself is 'nothing more than' property P. But on the view of property identity we are now considering, the result is not implied at all. The supervenience principle is implied, to be sure, but only in a purely trivial way: property P correlates with pain in all possible worlds whose laws and basic entities are like ours, but only because an event allegedly doesn't count as an instance of pain unless it is an instance of P. The 'raw feel' of pain is left out of this version of the principle completely.

Two conclusions can be drawn here. First, it is clear that the criterial identity of pain and property P is inadequate to capture an important belief which lies at the basis of materialism--viz., the intuition that the phenomenal aspect of pain is itself 'nothing more than' property P. Principles (E), (F), (D), and (N) are all meant to apply to pain qua phenomenal experience, since the phenomenal aspect of pain is the peculiarly mental aspect. Thus, the criterial identity

thesis is inadequate to express the materialist conception of the mental. We must add to it an assertion of the supervenience of pain-feelings upon instances of P. This principle involves a property other than P, which can be expressed roughly as follows: "has a property which involves the same 'raw feel' which we associate with the term 'pain'."

The second conclusion to be drawn is the one we have been working toward: we ought not to treat property P as the criterion for genuine pain; consequently, this purported criterion does not establish the identity of pain with P. This conclusion emerges from the considerations just cited, which show 1) that there is no point in treating property P as the criterion of genuine pain, and 2) that adopting the proposed criterion would involve unnecessary complication of theory, and 3) that the criterion would be counterintuitive. There is no point in adopting the criterion, because the resulting identity thesis fails to express the materialist conception of the mental that it's meant to express; the phenomenal aspect of pain is not guaranteed to be determined by property P. The criterion would introduce needless complexity into unified science; it would be necessary to distinguish pain qua physical property from the phenomenal raw feel we normally associate with pain, and we would have to employ the latter notion in stating the supervenience principle that's needed to express the materialist conception. Finally, the criterion would be counterintuitive, because it allows the following two possibilities: 1) that there are beings in worlds other than ours who have experiences qualitatively indistinguishable from pain, but which nevertheless are not pain; and 2) that there are beings in worlds other than ours who have pain (property P) when their phenomenal experiences differ radically from the raw feel we associate with pain, or even when they have no phenomenal experience

at all. Thus, there is no advantage to the view that property P is the criterion of genuine pain, and there are clear disadvantages.

It should perhaps be noted here that the argument just presented does not in itself rule out the possibility that pain and property P are identical properties. It is directed only against the view that P is the criterion of genuine pain, and that the properties are identical because of this criterion. It's possible to hold that the properties are identical, according to the SI criterion, without adopting the criterial view. This position is held by someone who maintains that pain (qua phenomenal property) can only be instantiated in one way in any possible world--viz., by property P. But as I said some time ago, I know of no argument in support of this extremely strong thesis. We can imagine possible worlds in which there are beings who are composed of fundamental components of an entirely different kind than those in the actual world, and which experience pain. (Imagine a world in which angels exist and feel pain, for example.)

I turn now to the case of the reduction of one physical theory to another, where the criterial identity thesis appears more plausible initially than in the case of reduction involving physical-phenomenal connecting principles. We will again consider the case of temperature and mean molecular kinetic energy. I shall contend that we ought not to view mean kinetic energy as the criterion for temperature, for reasons that are importantly similar to the reasons why we ought not to view property P as the criterion of pain.

Suppose we were to adopt the position that mean kinetic energy is the criterion for genuine temperature--a position whose corollary is that temperature is identical to mean kinetic energy, at all possible worlds.

This position runs into trouble for reasons analogous to those that emerged against the corresponding position regarding pain and property P. In the case of pain, the problem arose for the following reason: although the "raw feel" of pain correlates with property P in the actual world, we have no guarantee that this correlation will obtain in other possible worlds. In the case of temperature, the analogous problem concerns the "correspondence rules" which associate the theoretical notion of temperature with various observational statements.²⁰ (A typical correspondence rule is the principle that a particular temperature corresponds to a certain reading of a mercury thermometer.) The problem is this: although the correspondence rules associated with temperature apply to mean kinetic energy in this world, we have no guarantee that these rules will apply to mean kinetic energy in other possible worlds.

This lack of a guarantee is problematic because it leaves open the possibility that there are possible worlds in which mean kinetic energy displays none of the usual observable effects of temperature, and in which there is no explanation for this fact. (The lack of explanation is crucial, since observational correspondence rules presumably don't establish as strong a link between theoretical properties as the link between pain and the "raw feel" of pain. Thus, there certainly are cases in which a particular temperature is not reflected by the appropriate thermometer reading--for example, cases where the thermometer is defective. But in instances such as these, the failure of the usual correspondence rules is capable of explanation.)

So just as before we were faced with possible worlds in which pain doesn't hurt, we are now confronted with possible worlds in which temperature doesn't behave observationally the way it does in the actual world. The

supervenience principle is met in a trivial sense, but it fails in the important sense. It's met trivially because no property in any possible world counts as temperature unless it's mean kinetic energy; hence, temperature and mean kinetic energy are co-extensive in all possible worlds. But the principle fails in the important sense that there is no guarantee that the usual observational characteristics of temperature will remain fixed, ceteris paribus, in all possible worlds. Thus, the criterial identity of temperature with mean kinetic energy leaves open the possibility that there are possible worlds which are exactly like the actual world in composition, except that certain theoretical-observational correspondence rules are radically different than in the actual world. (Perhaps there are worlds where the mercury in thermometers falls, rather than rising, in accordance with an increase of mean molecular kinetic energy.)

This result is clearly counterintuitive. It seems that the identity of temperature with mean kinetic energy implies that the observational correspondence rules associated with temperature are true of mean kinetic energy in all possible worlds. We surely don't wish to say that a property can be temperature without displaying the usual observational characteristics of temperature (unless, of course, the lack of the usual characteristics can be explained). The reason we feel this reluctance is that 'temperature', as a theoretical concept in thermodynamics, is a useful notion largely because it helps to explain, predict, and unify certain observable phenomena. This explanatory/predictive role is crucial to the concept, as it is to thermodynamics as a whole. Moreover, if the usual connections to observation are not maintained, then the observable phenomena themselves are no longer necessarily governed

by the laws of thermodynamics. This leaves open the possibility that they don't obey any laws at all in certain possible worlds, even though these worlds have the same basic entities and the same micro-laws as the actual world.

The advocates of the N-D argument surely don't envision this possibility. When they assert that temperature is determined by mean kinetic energy, or that it's nothing more than mean kinetic energy, they mean to imply that mean kinetic energy will display, in all possible worlds, the observational characteristics we associate with temperature. This leads us to two conclusions, parallel to those we drew earlier concerning pain and property P. First, the criterial identity of temperature and mean kinetic energy is inadequate to capture an important belief which lies at the basis of the N-D argument--viz., the belief that correspondence rules we associate with temperature are true, ceteris paribus, in all possible worlds, of mean kinetic energy. This belief is central to the intuition that temperature is "nothing but" mean kinetic energy. Thus, we must add to the criterial identity thesis an assertion of the supervenience of these correspondence rules upon mean kinetic energy. This supervenience involves the following property: "satisfies the constraints imposed by the laws and correspondence rules associated with the term 'temperature'." The required supervenience asserts that this property is co-extensive, in all possible worlds, with mean molecular kinetic energy.

The second conclusion to be drawn, as in the earlier example of pain, is that we ought not to treat mean kinetic energy as the criterion of genuine temperature; consequently, this purported criterion does not establish the identity of the two properties. The reasons are the same as in the earlier case. First, there is no point in adopting the criterion, because the re-

sulting identity thesis fails to express the intuitive content of the assertion that temperature is "nothing but" mean kinetic energy. Second, the criterion would introduce needless theoretical complexity, because it would be necessary to distinguish temperature qua mean kinetic energy from the property which entities have just in case they satisfy the constraints imposed by the laws and correspondence rules associated with 'temperature'. It's simpler to identify temperature with this latter property. Third, the criterion is counterintuitive, because it fails to rule out the existence of possible worlds in which mean kinetic energy lacks--without explanation--the usual observational traits of temperature. So there is no theoretical advantage in the view that temperature is criterially identical to mean kinetic energy, and there are clear disadvantages.

In sum, the proposal to treat mean kinetic energy as the criterion for genuine temperature has the effect of divorcing the concept of temperature from the usual macroscopic, publicly observable phenomena with which this concept is associated in the actual world. If mean kinetic energy were the criterion for genuine temperature, then the role of theoretical-observational correspondence rules would be quite secondary. But correspondence rules are actually linked to theoretical concepts in a vitally important way, since one key purpose of theoretical notions is to facilitate explanation and prediction of the observable phenomena with which they are associated. Consequently, the applicability of the appropriate correspondence rules becomes at least a partial criterion for the applicability of the concept of temperature in other possible worlds. And if we accept this latter criterion, we must reject the criterial identity thesis.

In conclusion, let me briefly emphasize the difference between microreductively correlated properties,

according to the position I advocate. I contend that we should reject the identity thesis for microreductively correlated properties, but that we should adopt the part-whole supervenience principle. How exactly do temperature and mean kinetic energy differ, on this construal? The difference is based on the fact that temperature is ascribable to whatever entities satisfy the constraints imposed on the concept 'temperature' by the laws and correspondence rules of thermodynamics. We can imagine possible worlds where T_0 does not obtain, such that the relevant constraints are satisfied even though molecular theory does not obtain. In such worlds, an object would possess a certain temperature without possessing a mean molecular kinetic energy. But the fact that temperature can be instanced this way in some possible world does not affect the contention that in our world, and in all K-worlds where T_0 obtains, temperature is co-extensive with mean molecular kinetic energy. And the truth of the latter contention is sufficient to establish that temperature is "nothing else than" mean molecular energy.

¹I reconsider one possible form of the property identity thesis in the appendix. There, I continue to reject the thesis with respect to psychophysical connecting principles, but I leave open the possibility of a certain kind of identity thesis in natural science.

²Cf. Kim, "On the Psycho-Physical Identity Theory," pp. 227-28.

³Cf. Sklar, p. 119.

⁴Cf. Richard Montague, "On the Nature of Certain Philosophical Entities," Mon 53 (1969): 159-164.

⁵Jaegwon Kim has asked whether I am prepared to say the same about 'the teacher of Plato': that this expression doesn't denote an individual because it applies to different things in different possible worlds. In reply, let me point out that I do maintain that the expression 'the color of the sky' denotes a color--it non-rigidly

designates the color of the sky in whatever world is under discussion. Similarly, 'the teacher of Plato' designates (non-rigidly) an individual. But these facts are consistent with my claim in the text--viz., that the property expressed by 'is the color of the sky' is not a color. This claim is based on the fact that different-colored objects have this property in different possible worlds. One further comment: As we actually use the predicate 'is the color of the sky' in ordinary parlance, it perhaps has the meaning of 'is the color of the sky (at a)'. But this usage would help, rather than hinder, the argument in the text, since the latter predicate has the same intension as 'is blue'.

⁶I am assuming that 'pain' and the expression designating the neurological correlate of pain are rigid property designators. They make no mention of any specific objects whose characteristics change from one world to another, such as the sky or Plato. So there is no initial reason to think that they designate different properties in different possible worlds. (This matter is taken up in the appendix.)

⁷David Lewis, "An Argument for the Identity Theory," JP 63 (1966): 17-25.

⁸When we speak of a property occupying a causal role, we mean of course that the concrete events which are its instances occupy the causal role.

⁹Lewis, p. 17.

¹⁰Lewis' position is actually ambiguous. He phrases his "causal role" analysis of experiences by saying that certain causal roles "belong by analytic necessity to experiences." (p. 17) This assertion can be interpreted in two ways:

- a) By analytic necessity, a property is an experience of kind E iff it has causal role C.
- b) A predicate expresses a property of kind E iff by analytic necessity, it expresses a property with causal role C.

Statement (a) has the consequence that my proposed dual-causation view yields non-physical and physical experiences. But the dual-causation theorist can reply that statement (b) expresses the essential content of the "causal role" position; statement (a) is tendentious.

¹¹I will have more to say about this approach, with respect to properties other than phenomenal properties, in the appendix.

¹²See Kripke's discussion of the alleged identity of heat with the motion of molecules, "Naming and Necessity," p. 325.

¹³Putnam, p. 236. Putnam asserts that "the problem of stating a 'principle of individuation' for physical properties, magnitudes, etc., reduces...to the problem of describing the methodological constraints on reduction."

¹⁴Putnam, *ibid.*, seems to take this view, although he is not writing within the confines of the SI criterion.

¹⁵Kripke, pp. 314-42.

¹⁶*Ibid.*, p. 315-16.

¹⁷*Ibid.*, pp. 319-20.

¹⁸*Ibid.*, pp. 325-26.

¹⁹*Ibid.*, p. 353.

²⁰The notion of a correspondence rule is employed by philosophers of science to refer to a statement which connects assertions stated in theoretical vocabulary to assertions stated in antecedently-understood, pre-theoretical vocabulary. I discuss the notion critically in chapter IV, and I adopt a modified notion in its place. But the modification will not affect the points I make in the present chapter, so I will continue to use the notion uncritically until chapter IV.

IV. THE PART-WHOLE SUPERVENIENCE PRINCIPLE

Section 1

Having rejected the view that microreductively correlated properties are identical, I now turn to a defense of the part-whole supervenience principle (S).

(S) In all K-worlds in which theory T_0 obtains, the connecting principles R also obtain.

Recall that this principle assumes the truth of the Unity of Science hypothesis. T_0 is the fundamental reducing theory of unified science. R is the set of microreductive connecting principles in unified science--the principles which reduce various branches of science to branches immediately below them in the reductive hierarchy, and which thereby ultimately reduce all of science to T_0 . And K-worlds are possible worlds in which the only kinds of concrete indecomposable entities are those kinds which exist in the actual world--viz., entity-kinds recognized by T_0 .

Statement (S) is a fundamental metaphysical thesis. If adopted, it will occupy a position very close to the "center" of our theory of the world, in Quine's sense of centrality. I.e., it will resemble the principles of logic and mathematics, in that it will be subject to empirical disconfirmation only very indirectly, if at all. Empirical questions arise when we ask whether or not a particular theory T_2 is microreducible to some other theory T_1 ; proposed biconditional correspondence principles will be subject to the test of experience, to the same degree that other aspects of the particular scientific theories are so subject. Similarly, the Unity of

Science hypothesis is a (very general) empirical thesis, despite the fact that it also serves as a guiding methodological principle in research and theory construction. But the supervenience principle is much less subject to the vicissitudes of experience. For, it seems that the best we can do empirically is to determine whether or not certain proposed microreductive connecting principles actually obtain between two theories T_2 and T_1 . We cannot tell, on the basis of empirical evidence, whether or not the connecting principles would obtain in all K-worlds in which T_0 is true. Nor can we decide the truth of this latter thesis simply by appeal to specific scientific theories. These theories inform us of lawlike regularities which obtain in the actual world, but they are silent regarding the question of whether the correspondence principles will obtain in all possible K-worlds. The supervenience principle is further from the periphery of experience than any of the individual sciences. Accordingly, empirical considerations (other than those which are involved in establishing the existence of connecting principles in the actual world) will have little bearing on the truth or falsity of the principle.

This is not to deny that arguments for the principle can be given. The present chapter will be devoted to arguments in its favor. But arguments will occur at a different level than those which occur in theory-construction at the level of the particular sciences. Considerably more weight will be placed on considerations of simplicity, economy, intuitive naturalness, conceptual coherency, and the like, while virtually no weight will be placed upon empirical considerations.

It's clear that more of an argument is needed for the truth of a principle like (S) than is needed for the truth of the principles of logic or mathematics.

The truth of the latter can be maintained on the basis of their crucial role in our conceptual scheme: we are reduced to virtual incoherency if we deny them. It does not appear that the same can be claimed for principle (S). Nonetheless, if we are strongly influenced by the various intuitions which underlie the N-D argument then we will no doubt feel an inclination to adopt the supervenience principle on the grounds of these intuitions. Thus, intuitive naturalness represents a prima facie reason for adopting the principle. But there is more to be said on the subject than this. We need to inquire more fully into the foundations of the various intuitions which underlie the supervenience principle, in order to understand why they arise and why they deserve to be heeded. And we must also confront and reply to certain arguments which purport to show the falsity of the principle.

In order to obtain a more precise understanding of (S), and also to facilitate my defense of this principle, it will be necessary to discuss first the nature of scientific theories. This is a very large topic, of course, and my discussion will be limited. My purpose is to survey some main considerations which have been offered by others in support of a particular view of scientific theories--a version of which I shall adopt myself in arguing for the supervenience principle. This view, which I shall call the "modified conception" of scientific theories, is advocated explicitly by the recent Carl Hempel, and I suspect that something like it is held by many others. The view has emerged largely as a reaction to, and modification of, what Hempel calls the "standard conception" of scientific theories. I shall briefly describe the standard conception, the major objections which Hempel and others have levelled against it, and the modified conception which has emerged as a

result of these objections.¹ I will then introduce some distinctions of my own within the framework of the modified conception, distinctions which will prove useful in defending (S). And I will indicate the relevance of the modified conception to the understanding of certain key concepts of (S), notably the concept of a connecting principle.

The standard conception of scientific theories, as Hempel notes, is represented in the writings of N.R. Campbell, F.P. Ramsey, Rudolph Carnap, R.B. Braithwaite, and Ernest Nagel. On the standard construal, a theory is characterized by two constituents. The first is an axiomatized deductive system of uninterpreted formulas, the postulates of the system corresponding to the basic principles of the theory. The postulates may be thought of as formulas obtained by axiomatizing the basic principles of the theory and then "disinterpreting" (to use Quine's expression) the primitive theoretical terms in the axioms. The second component is a set of sentences that give empirical import to the system by interpreting some of its formulas in empirical terms--in terms of the vocabulary that describes the phenomena the theory is to explain. These sentences are usually called correspondence rules. They have also been known as operational definitions, co-ordinative definitions, or interpretative principles, and they have been characterized as forming a dictionary, or an interpretative system. Writers who employ the standard conception frequently assume a sharp dichotomy of vocabulary into "theoretical" and "observational" components, and they have thought of correspondence rules as sentences containing vocabulary of both kinds. But it seems possible to adhere to the essentials of the standard conception without relying on a sharp theoretical/observational dichotomy: correspondence rules are then viewed as relating theoretical

terms to pre-theoretical ones. A pre-theoretical term is one which is understood and has empirical import independently of the particular theory under consideration; it may or may not be "observational."

I come now to a brief mention of the major criticisms that have been levelled against the standard conception. First, there are several well-known critiques of the theoretical/observational dichotomy.² These critical discussions have sometimes been presented as an argument against the standard conception itself.³ And this approach has often been justified, since certain advocates of the standard view (notably Carnap) have tended to give the distinction a central role. But the objection can perhaps be avoided, by employing the weaker theoretical/pre-theoretical distinction.

So stated, however, the conception becomes vulnerable to a second objection. The construal of a theory as an axiomatized uninterpreted calculus plus a set of interpretative correspondence rules strongly suggests that all the terms in the axiomatized calculus are "new" theoretical ones, and that antecedently understood pre-theoretical terms are all introduced via the correspondence rules. Actually, however, the basic laws of most scientific theories employ not only "new" theoretical concepts but also "old" or pre-theoretical ones. For example, the basic laws of the classical kinetic theory of gases attribute to atoms and molecules such properties as mass, volume, velocity, momentum, and kinetic energy--all of which were previously used in antecedent theorizing about physical objects; the wave theory of light employs the antecedently available notions of wavelength and wave frequency, etc. Thus, the two-tier conception of a theory as an uninterpreted calculus plus a set of interpretative correspondence rules linking theoretical to pre-theoretical terms,

cannot be maintained.

A third objection to the standard conception concerns its treatment of the role of models. Certain proponents of the standard conception, notably Nagel, have regarded a theory as having a third component, namely "a model for the abstract calculus, which supplies some flesh for the skeletal structure in terms of more or less familiar conceptual or visualizable materials."⁴ Nagel observes that a theory is usually embedded in a model, rather than being formulated simply as an abstract calculus and a set of correspondence rules. The main reason for this, he suggests, is the greater ease of understanding which ensues. The implication is that the model, although it facilitates understanding and it can sometimes be a heuristic aid in formulating new hypotheses, nevertheless is logically dispensable.

The objection is that not all models used in science seem to be logically dispensable. Certain of them are dispensable, admittedly, such as the mechanical or hydrodynamic representations of electric currents or the luminiferous ether that played a considerable role in the physics of the late nineteenth and early twentieth centuries. But consider the models implicit in such theories as the kinetic theory of gases, classical wave and particle theories of light, Bohr's theory of the hydrogen atom, or the molecular-lattice theory of crystalline structure. All these claim to offer, not picturesque analogues, and not re-interpretations of a theory's abstract calculus in concretely visualizable terms, but tentative descriptions of the actual microstructure of the objects and processes under investigation. A model is often an integral and logically indispensable component of a theory.⁵

A fourth objection concerns the notion of a correspondence rule. Their designation as rules, as opera-

tional definitions, or co-ordinative definitions suggests that they have the status of metalinguistic principles which render certain sentences true by terminological convention or stipulation. But there are several problems with this conception: 1) Scientific statements that are initially introduced by "operational definitions" or more general rules of application for scientific terms usually change their status in response to new empirical findings and theoretical developments. (Hempel gives the following example. Some periodic process, such as the axial rotation of the earth as reflected in the apparent daily motion of some fixed star, might be used initially to "define" equal intervals of time in experimental terms. But ancient astronomical reports have been used to establish a very slow deceleration of the earth's axial rotation, with a consequent slow lengthening of the mean solar day.) 2) A theory usually links a given theoretical concept to several distinct kinds of phenomena that are characterized in terms of the antecedently available vocabulary. Hence, not all the true interpretive sentences thus provided for a given theoretical term can be true by convention, since they jointly imply certain empirical statements to the effect that if one of the specified methods yields a certain numerical value for the quantity in question, then the other methods will yield the same value. 3) Quine's critique of the analytic/synthetic distinction suggests that no sentence, with the possible exception of the truths of logic and mathematics, qualifies as immune to revision should the theory encounter certain recalcitrant empirical findings. Hence, it's not clear what claim is being made when certain sentences of a theory are characterized as "true by convention."

A fifth objection concerns the initial motivation

of the standard construal. The presumptive problem which the standard conception sought to solve was the question of how the meanings of the "new" terms in a theory are specified. But writers like Putnam and the recent Hempel have argued that this presumptive problem does not really exist.⁶ Those who posed the problem assumed that we cannot come to understand new theoretical terms except by way of sentences specifying their meanings with the help of previously understood terms. This idea seems untenable, in light of the ways in which new theoretical terms are introduced into a language, and the ways in which we learn to use them properly: from instances of their use in particular contexts, from loose paraphrases that employ "broad spectrum" terms (like 'thing' and 'physical magnitude') or certain imprecise everyday terms (e.g., 'harder to accelerate', 'heavier than'), from reading the theory or text in which the term is employed, and so on.

A sixth objection⁷ involves the notion that the correspondence rules specify the interpretation of the abstract calculus. The traditional view suggests that the cognitive significance of theoretical terms and statements is solely a function of their relation to assertions framed in the pre-theoretical vocabulary. The significance is drawn up to the theoretical terms--by osmosis, as it were--through the theoretic/pre-theoretic connections asserted by the correspondence rules, and then secondarily by means of the intratheoretic connections established by the postulates of the calculus. If this view is taken literally, it seems to imply a fairly drastic instrumentalism. No provision has yet been made for treating theoretical referring expressions ('electron', 'DNA molecule') as having any actual referents, and no provision has been made for treating theoretical predicates ('has a unit negative

charge', 'is radioactive') as having an extension other than the null extension. We have been given no reason to think that theoretical entities exist, or that theoretical predicates are true of anything. So if the only cognitive significance possessed by theoretical assertions is that which is provided by the correspondence rules (and then, indirectly, by the postulates of the calculus), then the theory becomes a mere instrument for purposes of systematically connecting pre-theoretical assertions.

Before we can embrace even a minimal scientific realism, we must allow for more than mere empirical interpretation of the abstract calculus: we must also allow for semantic interpretation, in the sense of Tarski.⁸ That is, we must take theoretical denoting terms to refer to actual entities, and we must take theoretical predicates as having actual extensions among the theoretical entities. Suppose we so take them, in a suitable metalanguage. Then the set of theoretical postulates, viewed as an uninterpreted calculus, can be given the desired semantic interpretation by asserting the truth condition for each sentence of the calculus to be its restatement in, or its translation into, our metalanguage. Now, it seem obvious that we can and should so interpret the postulates of a theory, precisely because the alternative is to embrace an implausible instrumentalism. And it may be the case that some advocates of the standard conception would not wish to dispute the need for semantic interpretation. But the fact remains that provision is not made in the standard conception itself for semantic interpretation. According to the received view, a theory is an ordered couple of sets (T,R), where T is a set of statement forms--a semantically uninterpreted calculus--and R is the set of theoretical/empirical correspondence

rules. In order to allow for semantic interpretation, the standard view would have to be modified, by construing T as a semantically interpreted calculus. (In such a modified construal, correspondence rules would continue to function as statements of empirical truth conditions, thus providing an "empirical interpretation" for T. They would delimit a set of empirical constraints which must be satisfied by the actual, intended semantic interpretation.⁹)

Before leaving this objection, a historical observation is in order. The distinction between semantic and empirical interpretation of theoretical terms and statements has been blurred in the past largely because of the legacy of the logical empiricist conception of cognitive significance. For a good logical empiricist, there is simply nothing more to the cognitive content of a theoretical assertion than its empirical import. Hence, theoretical statements receive as much interpretation as they're capable of by means of correspondence rules. Carnap has been bravely consistent on this point. He accepts the consequence that the realism/instrumentalism controversy is really a pseudo-issue--that questions as to the reality or non-reality of theoretical entities are meaningless.¹⁰ But the fact that he has been forced to such a heroic stand seems to constitute a good reason, in addition to the many others that have been adduced by critics, for rejecting the logical empiricist conception of meaning.

On the basis of objections of the sort just outlined, Hempel has recently been led to adopt a conception of scientific theories which differs somewhat from the standard construal. I will briefly describe his conception, since it will play an important role in our subsequent discussion of the supervenience principle. He suggests that a theory contains statements of two

kinds, which he calls internal principles and bridge principles.

The internal principles serve to characterize the "theoretical scenario": they serve to characterize the basic entities and processes posited by the theory, as well as the theoretical laws that are assumed to govern them. The bridge principles, on the other hand, indicate the ways in which the scenario is linked to the previously examined phenomena which the theory is intended to explain...If I and B are the sets of internal and bridge principles by which a theory T is characterized then T may be represented as the ordered couple of those sets: $T=(I,B)$. Or alternatively, and with greater intuitive appeal, T may be construed as the set of logical consequences of the sum of the two sets: $T=c(I \cup B)$.¹¹

This division of a theory into two components has certain parallels to the standard division into an uninterpreted axiomatized calculus and a set of correspondence rules, but it differs in important respects. The internal principles of the theory are not conceived as an uninterpreted logical skeleton; rather, they include the full-fledged semantically interpreted laws of the theory. They can include other statements, too--specifically, statements specifying models which purport to further describe the theoretical entities mentioned in the basic laws of the theory. Internal principles may also contain pre-theoretical terms, if such terms are required in the formulation of basic laws of the theory. In all these respects, internal principles differ from the semantically uninterpreted statement-forms which comprise the first component of a theory, according to the standard conception.

Hempel's bridge principles are somewhat analogous to the correspondence rules of the received view. Their role is, essentially, to connect internal principles to the body of phenomena they are intended to

explain. But again there are important differences between Hempel's view and the standard conception. Not all bridge principles involve observational language: some will link the "theoretical scenario" to other antecedently understood theoretical phenomena. Thus, a sharp theoretical/observational dichotomy is not presupposed by the notion of a bridge principle. Rather, bridge principles are pictured as relating "new" theoretical vocabulary to pre-theoretical vocabulary, and the latter will often include vocabulary from other theories.

This broad conception of bridge principles may or may not conflict with the notion of a correspondence rule, depending upon whether or not the theoretical/observational distinction is taken to be a part of the standard construal. (I suggested above that it need not be.) In any case, we come now to some definite discrepancies between the two notions. First, Hempel's bridge principles are not considered expressions of linguistic convention. Rather, they are substantive assertions of the theory, subject to revisions in the face of recalcitrant experience. Secondly, they are not viewed as the sole meaning-confirming vehicles of the theory. They do specify the empirical content of the theory, of course, but they do so by making substantive assertions about the relations between the theoretical and the empirical, rather than by specifying the sole "cognitive significance" of theoretical assertions.

Section 2

The objections described above to the standard conception of scientific theories are jointly quite compelling, and they provide ample reason for adopting Hempel's modified view, or something like it. I shall adopt his basic conception, therefore, and I shall use

it as a basis for defending the part-whole supervenience principle. Certain aspects of the modified conception will be crucial--in particular, its thoroughgoing realism as regards theoretical micro-entities and micro-processes, and its recognition that models in science are often intended as actual descriptions of these entities and processes. (Actually, I will not adopt the modified conception in quite the form presented by Hempel. I will further modify Hempel's own conception, in a way that will be made clear shortly.)

We will also require several distinctions which Hempel himself does not make explicitly. First, as regards the set of bridge principles of a theory, we need to distinguish the subset of these which count as theoretical-observational bridge principles. Now, I have no wish to re-introduce the justly criticized distinction between observational vocabulary and theoretical vocabulary, nor do I wish to suggest that we can give a completely sharp and precise delimitation of the class of theoretical-observational bridge principles. Nonetheless, it is important to distinguish this subset, because it encompasses the more-or-less direct empirical import of a theory, as opposed to its interconnections with other theories. (Hempel's characterization of bridge principles is wide enough to include even microreductive connecting principles.)

A bridge principle will normally assert that the truth of a statement S_1 that's phrased in pre-theoretic vocabulary is necessary, or sufficient, or perhaps necessary and sufficient, for the truth of an assertion S_2 that involves certain "new" theoretical terms. (For example, S_1 might be a statement asserting that a spectral line is present at a specific position in the spectrum of a certain element; S_2 might be the assertion that electrons in the element are jumping from one

specified permissible orbit to another. The truth of S_2 would count as sufficient, or at least ceteris paribus sufficient, for the truth of S_1 .) The problem we face is to describe the conditions under which S_1 will count as observational. This task has been accomplished, in a way that seems adequate for our purposes, by Quine. He employs the notion of the stimulus meaning of a sentence for a native speaker of a language, and the notion of an occasion sentence. Roughly, the affirmative stimulus meaning of a sentence for a speaker is the class of all the stimulations that would prompt his assent; the negative stimulus meaning is the class of all the stimulations that would prompt his dissent; and the stimulus meaning in toto is the ordered pair of the two.¹² Occasion sentences, as against standing sentences, are sentences "which command assent or dissent only if queried after an appropriate prompting stimulation."¹³

Now for Quine's characterization of observation sentences:

Occasion sentences whose stimulus meanings vary none under the influence of collateral information may naturally be called observation sentences, and their stimulus meanings may without fear or contradiction be said to do full justice to their meanings...Or better, we may speak of degrees of observability....

In behavioral terms, an occasion sentence may be said to be more observational the more nearly its stimulus meaning for different speakers tends to coincide....

Among standing sentences that are well over toward the occasion end,...the notion of observability works quite as well as among occasion sentences....

Viewing the graded notion of observability as the primary one, we may still speak of sentences simply as observational sentences when they are high in observability.¹⁴

Quine observes that 'observation sentence' sug-

gests the datum sentences of science. And indeed, these are precisely the sentences we must capture in order to delimit the theoretical-observational bridge principles of a theory.

On this score our version is not amiss; for the observation sentences as we have identified them are just the occasion sentences on which there is pretty sure to be firm agreement on the part of well-placed observers. They are just the sentences on which a scientist will tend to fall back when pressed by doubting colleagues.¹⁵

I shall use the term 'correspondence principle' to refer to bridge principles which connect theoretical assertions directly to observation sentences. These principles may be thought of as coinciding with the theoretical-observational "correspondence rules" of the standard conception. But of course they are not rules; they are substantive assertions of the theory.

We come now to another set of distinctions which Hempel does not explicitly make, but which is important for our subsequent discussion. This involves various narrower and wider senses which might be assigned to the notion of a theory. We can accept Hempel's contention that intertheoretic bridge laws, correspondence principles, and models are all "a part of theory" in some sense, and yet still ask whether there are legitimate narrower senses of 'theory'. It seems that a narrower sense is appropriate, for example, when we speak of the reduction of one theory to another; on Hempel's construal of theories, a reduced theory is actually a part of the reducing theory, rather than a separate theory. This is because the class B of "bridge principles" includes microreductive connecting principles. Thus, if T_2 is microreducible to T_1 , and $T_1 = c(I_1 \cup B_1)$, then T_2 is contained in T_1 .

A narrower sense will likewise be needed if the supervenience principle is to count as a non-trivial assertion: a reduced theory T_i will be trivially true in all possible worlds where the final reducing theory T_0 is true, if the connecting principles linking T_i to T_0 are themselves regarded as a part of T_0 . So in order to prevent the term 'intertheoretic reduction' from becoming a misnomer, and in order to save (S) from triviality, we must distinguish at least one sense of theory that's narrower than Hempel's.

A solution to the problem is ready at hand. We should distinguish different components in the set of bridge principles B, and we should only include certain of the bridge principles as a part of theory in the narrower sense. Let Bs be the set of "secondary" bridge principles--those that connect "new" theoretical assertions to antecedently understood theoretical assertions, and to observational assertions which are associated with these antecedently understood theoretical assertions through the correspondence principles of the antecedently understood theories. ("New" theoretical assertions, as against antecedently understood theoretical assertions, are statements which make essential use of vocabulary unique to the internal principles I of the given theory.) Bs thus includes microreductive connecting principles which link T to other theories. It also includes correspondence principles which link the internal concepts of T to observation indirectly--correspondence principles which follow logically from intertheoretic connecting principles together with the correspondence principles of antecedently understood theories.

Let Bp be the set of principles which are "primary" correspondence principles--those which associate observational statements directly with the theoretical assertions employing only the vocabulary of the theory's internal principles. Bp and Bs together comprise the

full set B of bridge principles.

I suggest that the most useful sense of 'theory', both generally and for purposes of understanding intertheoretic microreduction and the supervenience principle, is this: $T=c(I \cup Bp)$. Bridge principles which involve antecedently understood theory should be treated as auxiliary to a given theory, as the phrase "intertheoretic connecting principle" itself suggests. Before we accept this narrower sense, however, we should note the reason why Hempel himself employs a broader conception of theory than this. In the above-quoted passage in which he describes his conception, he characterizes the bridge principles as statements which indicate the ways in which the "theoretical scenario" described by the internal principles is linked to the previously examined phenomena which the theory is designed to explain. The class of such principles is wider than Bp, because the phenomena explained by theory T will not necessarily be observationally described phenomena. Accordingly, Hempel's conception of theory is broader than ours.

It is certainly true that certain theoretically-described phenomena are sometimes explained by means of a theory other than the theory which provided the original description. In fact, this is just the sort of explanation which goes on when one theory has been reduced, or partially reduced, to another. So it might seem that if we wish to assert that lower-level theories do provide explanations for phenomena described in terms of higher-level theory, then we must say with Hempel that intertheoretic connecting principles are a part of the lower-level theory, on a par with the theoretical-observational correspondence principles of the lower-level theory. However, I shall argue in chapter VI that we are not obliged to adopt such an inclusive sense of 'theory' in order to allow for the fact that theoretic-

cally described phenomena are explainable by more basic theories. We can provide for the motive underlying Hempel's wide sense of 'theory' without actually adopting this wide sense. Accordingly, we will understand the notion of theory in the sense of $c(I \cup Bp)$, the sense which saves the supervenience principle from triviality and which prevents the term 'intertheoretic reduction' from becoming a misnomer. This conception of theory is itself a modification of Hempel's "modified conception" of scientific theories.

In light of our conception, we should take note of the types of statements that must be included in a class $R_{1,2}$ of connecting principles which reduce a theory T_2 to a theory T_1 . Of course, the set $R_{1,2}$ will include certain higher-level biconditional assertions which link the theoretical laws in I_1 to those in I_2 . But it must contain more than this, if Nagel's derivability requirement is to be met for correspondence principles of T_2 . In order for T_2 to be deducible from $T_1 \cup R_{1,2}$, the class $R_{1,2}$ must also contain principles which allow us to deduce T_2 's primary correspondence principles Bp_2 from $T_1 \cup R_{1,2}$. Thus, $R_{1,2}$ will contain a set of auxiliary correspondence principles, $R\text{-}Aux_{1,2}$, which result from replacing the non-observational T_2 predicates in Bp_2 by their nomologically equivalent T_1 counterparts. These auxiliary correspondence principles will not necessarily be biconditional in form, since primary correspondence principles are not necessarily biconditional in form. Non-observational concepts employed in the laws of I_2 must still have co-extensive counterparts from I_1 , however. So we continue to conform to the spirit, if not the letter, of strong Nagel reducibility.

The presence of models within some theories threatens to further complicate the nature of micro-

reductive connecting principles. At first sight, it would appear that $R_{1,2}$ must contain principles which allow us to deduce a description of the model of T_2 entities, if there is a model, from T_1 . However, I suggest that such principles will never be necessary, because of the nature of the models themselves. If theory T_2 is reducible to theory T_1 , then the model for T_2 entities will be constructable from the model for T_1 entities. I.e., it will be possible to deduce, on the basis of the model and theoretical laws of T_1 , that the T_2 model is satisfied by the T_1 thing-complexes with which T_2 things are microreductively identified.

A vivid illustration is provided by a consideration of Bohr's "solar system" model of the atom in microphysics and the "elastic sphere" model of the molecule in molecular chemistry. The latter model asserts that molecules are roughly spherical in shape, that they have varying degrees of mutual cohesiveness, that these degrees of cohesiveness are related to their kinetic energies, that groups of molecules can take the form of solid, liquid, or gas, that they are composed of varying numbers of atoms of varying kinds, and so on. The Bohr model, on the other hand, gives us more specifics about the internal structure of molecules and of the atoms of which they are composed. Atoms are solar systems with protons and neutrons at the center and electrons circling around in various orbits; the orbits of different atoms can intertwine to form interlocking systems of atoms (molecules); and so on.

The crucial point concerning these two models is this: the Bohr model treats complexes of atoms in such a way that they satisfy the "elastic sphere" model for molecules. The latter can be constructed from the former. Considerable additional detail about the internal structure of atoms is added, of course, but the Bohr

model guarantees that atom-complexes as a whole have the characteristics specified by the molecular model. In fact, a major task in constructing a model of the atom is to show how atoms so construed cluster together to form molecules, to explain the details of molecular bonding in various solids and liquids, and in general to explain how the antecedent model of the molecule is satisfied. If this can't be done, then one model or the other (perhaps both) must be revised or rejected.

In sum, our acceptance of models as a component of some theories does not mean that further types of connecting principles will be required for reduction, over and above 1) the biconditional principles linking lower-level concepts to higher-level concepts, and 2) the auxiliary correspondence principles. But the T_2 model must be constructable from the T_1 model before T_2 can be reduced to T_1 . Specifically, the model for T_2 things must be constructable from the T_1 model as applied to the T_1 thing-complexes with which T_2 things are putatively to be identified. Thus, the constructability of higher-level models is presupposed by the microreductive connecting principles which connect sortal properties--the connecting principles we call "thing identities." So constructability of higher-level models is a condition which must be satisfied by any adequate microreduction.¹⁶

(Having mentioned microreductive thing identities, we should perhaps digress briefly to discuss their metaphysical status. The reader will recall from chapter I that generalized thing identities, like "Genes are DNA molecules," have the logical form of universally-quantified biconditionals. They involve sortal predicates--predicates for thing kinds, like 'is a gene' and 'is a DNA molecule'. Such statements count intuitively as identities because they assert that two classes of entities (like the class of genes and the class of DNA molecules) are one and the same class.

Two further points deserve mention. First, as we saw in chapter I, these generalized thing identities do not imply the identity of the relevant sortal properties. Secondly, such identities may well be contingent, given Kripke's treatment of identity statements. The statement "Genes are DNA molecules" is equivalent to the statement "The class of things that are genes is identical to the class of things that are DNA molecules," and the terms flanking identity in the latter statement may well be non-rigid designators. In fact, as I argued in chapter III, we can imagine possible worlds (other than K-worlds where T_0 is true) in which a given higher-level theory obtains although the lower-level theory to which it is reducible in the actual world does not obtain. In such worlds, the microreductive thing-identities of our world do not obtain. For example, in a world where Mendelian genetics obtains but molecular chemistry does not obtain, genes are not DNA molecules.)

Section 3

We come now to the task of defending the supervenience principle. We must show that in all K-worlds where the final reducing theory obtains, the connecting principles of R also obtain. (In subsequent discussion, I shall refer to K-worlds where T_0 obtains as " T_0 K-worlds.") To begin with, we must ask what criteria should be used to decide whether or not the connecting principles R obtain in a given T_0 K-world. The question of the truth or falsity of (S) cannot be significantly raised until we have specified the criteria for truth of the connecting principles. So our first task is to set out the appropriate criteria. We will then turn to arguments to show that the criteria are met in all T_0 K-worlds.

Let w be an arbitrary T_0 K-world. What conditions must be met in w , in order for the connecting principles

R to obtain there? Two sorts of conditions are clearly provided by the connecting principles R themselves. First, the statements R include auxiliary correspondence principles. These link observable phenomena originally associated with higher-level theoretical properties to lower-level theoretical properties which are the putative micro-correlates of the higher-level ones. And ultimately in the hierarchy of microreduction, the auxiliary correspondence principles link observable phenomena originally associated with higher-level theoretical properties, to the T_0 properties which are the ultimate putative micro-correlates of these higher-level theoretical properties. So one clear necessary condition for the truth of the full set R of connecting principles is the truth of the auxiliary correspondence principles in R. Before we can assert that the connecting principles linking higher-level theoretical properties to lower-level theoretical properties are true in world w, we at least must be assured that the lower-level theoretical properties have the appropriate connections to observation. (For example, before we can assert that temperature is correlated with mean molecular kinetic energy in w, we must be assured that the latter property has the connections to observation which are specified by the theoretical-observational correspondence principles of thermodynamics.)

A second clear condition is provided by the thing-identities of R--the biconditional connecting principles linking higher-level sortal properties to complex lower-level sortal properties. (For example, the property "being a hydrogen molecule" is correlated with the complex property "being such-and-such a structure of subatomic particles," where "such-and-such" is spelled out in microphysical terms.) As we saw earlier, one condition for the truth of these microreductive thing-identities is that the model for higher-level things be con-

structable one the basis of the model for lower-level things. (For example, we must be able to construct the molecular model of a hydrogen molecule as an "elastic sphere" from the subatomic model which treats the entity as a "solar system" with one internal proton and one orbiting electron.) So, before we can assert that the thing identities of R are true in world w , we must be assured of the constructability of the models of higher-level entities on the basis of lower-level models as they apply to the complex lower-level entities with which higher-level things putatively are to be identified. Thus constructability of higher-level models is another condition, in addition to the truth of auxiliary correspondence principles, which must be met before we can claim that the full set of connecting principles, R , obtains.

Beyond these two conditions, the question of criteria for the truth of the full set R becomes somewhat hazy. What more is needed, above and beyond the satisfaction of these conditions, for the full set of principles R to obtain in the world w ? I submit that these two conditions are the basic ones, and that very little more is needed. Specifically, the only further condition which must be satisfied is the following "uniqueness" condition for the set of statements R : There is no set of statements R' , different from R , such that a) R' is comparable to R in theoretical simplicity, b) R' satisfies the above two requirements regarding model constructability and truth of auxiliary correspondence principles, and c) R would, if true, reduce all true higher-level scientific theories to T_0 . Unless this third condition is added, we face the prospect that two incompatible and equally simple sets of putative connecting principles might both satisfy the above two conditions in the possible world w ; we would then have no way to choose between them. Accordingly, R must be the unique simplest set of connecting principles which would reduce higher-level theories

to T_0 .

My claim is that the above three conditions are sufficient conditions, as well as necessary conditions, for the truth of the connecting principles R in world w . The conditions, once again, are these: 1) The auxiliary correspondence principles of R obtain in world w . 2) In world w , the models for the theoretical entities countenanced by higher-level theory are constructable on the basis of the T_0 model, when the latter model is applied to the structured complexes of T_0 entities which are putatively to be identified with higher-level entities. 3) There is no comparably simple set R' of putative connecting principles which satisfies conditions (1) and (2). On what basis can it be claimed that satisfaction of these three conditions in world w guarantees the truth of the full set R in world w ?

First, it's difficult to see what further conditions there could be. Consider, for example, the connecting principles which reduce thermodynamics to molecular statistical mechanics. Suppose that the relevant auxiliary correspondence principles are true in world w : a given level of mean molecular kinetic energy corresponds to a given volume of mercury in an appropriately constructed thermometer just as it does in the actual world, and so on. And suppose that there are no other micro-properties at the level of molecular theory which satisfy the relevant auxiliary correspondence principles. Can we specify any further conditions which must be met, in order for the connecting principle linking temperature and mean molecular kinetic energy to obtain in the world w ? It seems that no further conditions are specifiable, and that none are required. Those already specified are sufficient to guarantee that the connecting principle does obtain in w .¹⁷

Consider another example, in which the role of models becomes prominent. Take the Bohr theory of the

atom, and the theory of atomic and molecular bonding which is based upon it. Various laws of molecular chemistry are reducible to microphysical principles, on the basis of this model together with connecting principles which link the higher-level concept of atomic and molecular "cohesiveness" to detailed microphysical concepts which deal with electron-orbits and the like. In addition, the "elastic sphere" model of the molecule and the atom, as put forward in molecular chemistry, is constructable on the basis of the Bohr model of the atom. Suppose, first of all, that the "elastic sphere" model is again constructable in world *w*, on the basis of the putative identification of various molecules with the complex Bohr-structures with which they are identified in the actual world. I.e., the relevant Bohr-structures in world *w* meet the specifications of the molecular-chemistry model: they are roughly spherical in shape, they have varying degrees of mutual cohesiveness which are related to their kinetic energies, in groups they can take the form of solid, liquid, or gas, and so on. Secondly, suppose that the various auxiliary correspondence principles involved in the reduction of molecular chemistry to the Bohr theory are also true in the world *w*. Bohr-structures which we call " H_2O molecules" have the ceteris paribus observable characteristics which water possesses in the actual world; the same is true for Bohr-structures which we call "sodium chloride molecules," and so on. Finally, suppose that there is no comparably simple set of putative chemical/microphysical connecting principles which meet the above two conditions in world *w*. What more would we require, above and beyond the truth of the above three suppositions, in order for the full set of chemical/microphysical connecting principles to obtain in the world *w*? As with the case of temperature and mean molecular kinetic energy, no further conditions

seem specifiable, and none seem needed.

One reason why the above mentioned conditions are sufficient for the truth of a full set of connecting principles in a T_0 K-world, therefore, is that further conditions are hard to find and seem unnecessary. This claim is further reinforced when we consider the status of microreductive connecting principles in the actual world. Suppose T_1 and T_2 are both true theories. Let $R_{1,2}$ be a set of putative connecting principles which would, if true, reduce T_2 to T_1 . What conditions must be met, in order for the principles $R_{1,2}$ to be true? The conditions are very similar to those discussed above. First, the auxiliary correspondence principles of $R_{1,2}$ must be true. I.e., it must be true that the observable occurrences originally associated with T_2 entities and their properties are also associated with those T_1 entities and properties proposed as the microreductive counterparts of the T_2 entities and properties. Secondly, the T_2 entities must actually be identical to complex T_1 entities, in accordance with the thing-identity statements of $R_{1,2}$. Thirdly, there must be no comparably simple set of connecting principles, $R_{1,2}'$, which meets the first two conditions. I submit that if these three conditions are met, then we won't hesitate to accept the connecting principles $R_{1,2}$ as true in our world. There are no further conditions which must be established.

The three conditions just mentioned apply in the actual world. Let me digress here to note the difference between these conditions and the above conditions for the truth of connecting principles in T_0 K-worlds other than the actual world. The difference involves the second condition in the two cases. In the case of reduction in the actual world, the second condition specifies that the thing identities of $R_{1,2}$ must be true, in order for the complete set $R_{1,2}$ to be true. In the case of alter-

native T_0 K-worlds, the second condition merely asserts that the model for higher-level theoretical entities must be constructable on the basis of the lower-level model. The reason for the difference is this. In the actual world case, we were given that the higher-level theory T_2 obtains, and hence that T_2 things actually exist. So in order for the full set of statements $R_{1,2}$ to obtain, T_2 things must be composed in the way the thing-identities of $R_{1,2}$ say they are composed. In the case of a T_0 K-world other than the actual world, however, we do not know in advance that any higher-level theories obtain, or that any higher-level theoretical entities exist. Accordingly, the problem of finding conditions for the truth of the connecting principles includes the problem of determining conditions under which the higher-level sortal predicates apply. Constructability of the model for higher-level entities thus becomes a relevant condition.¹⁸

Returning to the question of criteria for the truth of connecting principles R in T_0 K-worlds, we have now seen two reasons for accepting the above-discussed conditions as sufficient. First, it is difficult to find further conditions, and it does not appear that further conditions are needed. Second, very similar conditions are adequate to establish the truth of a set of putative microreductive connecting principles in the actual world. A third reason, perhaps implicit in the first two, involves simplicity. We would introduce needless complication and obfuscation were we to assume that further conditions must be met in order for the principles R to obtain in T_0 K-worlds. This contention is well illustrated by the case of temperature and mean kinetic energy. Consider a T_0 K-world, w , in which molecular statistical mechanics obtains, in which the full set of auxiliary correspondence principles involved in reducing thermodynamics to statistical mechanics also obtains, and in

which there is no comparably simple alternative set of putative connecting principles whose auxiliary connecting principles are true. In such a world, mean molecular kinetic energy has all the observable characteristics of temperature, and no other molecular property has these characteristics. Suppose it were claimed that these facts alone do not suffice to establish that temperature is correlated with mean kinetic energy in world w , because there is some further "something"--something essential to the presence of temperature--which may or may not be present in world w at times when molecules have the appropriate mean kinetic energy.

An advocate of this position would have considerable difficulty in describing this further "something." Furthermore, his contention has no evident theoretical advantage. On the contrary, this proposal threatens to entirely obfuscate the main question we are considering--the question of criteria for the truth of connecting principles in T_0 K-worlds. These observations apply to microreductive connecting principles quite generally. There is no evident theoretical advantage in the claim that the above-mentioned conditions are inadequate to establish the truth of a given set of microreductive connecting principles in a T_0 K-world. Furthermore, such a claim could only serve to obfuscate the question of truth conditions. Considerations of simplicity and clarity thus provide a third reason, in addition to the two already mentioned, to accept the criteria discussed above.

Having established the criteria for truth of connecting principles in K-worlds where T_0 obtains, we are now in a position to argue for the supervenience principle. We must argue, with respect to the set R of microreductive connecting principles in unified science, 1) that the auxiliary correspondence principles in R are true in all T_0 K-worlds, 2) that higher-level theoretical models are constructable, in all T_0 K-worlds, on the basis

of the T_0 model, and 3) that in no T_0 K-world is there a comparably simple alternative set of connecting principles which meets the conditions of model constructability and truth of auxiliary correspondence principles.

Section 4

The argument for the supervenience principle will be presented in two stages. The first will occupy the remainder of the present chapter, and the second will be presented in chapter V. The first stage deals with standard cases of microreduction in natural science, while the second deals with the reduction of psychology (and, at least by implication, with the reduction of the higher-level social sciences). One reason for the separate treatment involves the fact that there are two types of microreductive correspondence principles, which will be handled quite differently: I'll call them theoretical-observational correspondence principles and theoretical-phenomenal correspondence principles. The former are principles connecting theoretical assertions to publicly perceivable phenomena--phenomena to which no person has 'logically privileged access'. Meter readings, cloud chamber photographs, computer print-outs, slides under microscopes, and the myriad of other phenomena observable by reserachers in the laboratory all fall within this class. Theoretical-phenomenal correspondence principles, on the other hand, connect certain theoretical assertions to phenomena of a logically privileged sort--phenomena like pain, the sensation of redness, the sensation of heat. These phenomena are logically privileged in the following sense: it is only possible for me to directly experience my own sensation of redness, or my own feeling of pain. Others can perhaps infer from my behavior that I am undergoing these experiences, but they cannot have a direct awareness of my phenomenally-present mental states.

Within psychology itself there are no theoretical-phenomenal correspondence principles, except in the degenerate sense in which the "correspondence" relation is that of identity. 'Pain', for example, is both a phenomenal term and a theoretical term: we are directly aware of our sensations of pain, and pain is also the subject of psychological theorizing. And similarly for other sensation-describing terms of theoretical and common sense psychology. But there can also be nondegenerate theoretical-phenomenal correspondence principles. Suppose, for example, we achieve one step in the envisioned overall reduction of psychology to natural science--viz., the discovery of a correlation between pain and neural property P. Such a correlation law would count as a "theoretical-theoretical" connecting principle, since pain is a subject of psychological theory. But it would also count as a microreductive correspondence principle--a theoretical-phenomenal correspondence principle.

Because theoretical-phenomenal correspondence principles would be a part of the reduction of psychology to natural science, this case of microreduction will be considered separately in our argument for (S). Correspondence principles involved in other parts of science, and in other cases of microreduction, can all be regarded as theoretical-observational. This is certainly not to deny the obvious fact that we constantly use the phenomenal contents of our awareness as a basis for affirming public observation statements. But it is to deny that phenomenal content is essential to the truth of statements about intersubjectively observable events and processes. Redness, for example, is a property that's describable in terms of electromagnetic radiation theory, and so described it's independent of how the object happens to affect us phenomenally. (If a

person has an optical device planted behind his retina which converts the neural impulses from red light into green-light impulses--so that red objects look green to him--the fact remains that such objects are really red.) A meter needle that moves to the left three notches might appear to someone to move to the right, through some neurological quirk of his; but again, it really moves left, regardless of what it appears to do.

It's no accident that correspondence principles for theoretical statements other than assertions about our own states of awareness are all theoretical-observational. If this were not the case, then the truth of various theoretical assertions about what goes on "out there" in the world, independently of the conscious observer, would be made to depend upon the truth of assertions about the observer's states of consciousness. And this is a very dubious dependence, for several reasons. First, some neurological quirk in a particular observer might cause him to fail to have the appropriate phenomenal experiences in the presence of a certain theoretically-described phenomenon. (He might not feel hotter in the presence of increased heat because of brain-damage, or a cleverly implanted electrode.) Secondly, one observer might have the appropriate phenomenal experience when the other doesn't, because of different neurological hookups or for whatever reason. These problems can only be avoided if we connect theoretical assertions about the way things are "out there" to observational assertions about the way things are "out there," rather than to phenomenal assertions about the way things are in the consciousness of various observers.

So the unique presence of theoretical-phenomenal correspondence principles in the microreduction of psychology constitutes one major reason why we will need to consider this case of microreduction separately in our argument for (S). Another reason for separate con-

sideration involves psychological properties of the cognitive and volitional variety. As we shall see in chapter V, there seem to be excellent reasons for thinking that strong Nagel reducibility is not obtainable with respect to those parts of psychology employing typical cognitive and volitional concepts. (These considerations will apply to higher-level social sciences also, insofar as they employ such concepts.) Accordingly, we must reconsider the concept of reduction itself with respect to psychology, before we can argue for the supervenience principle.

Section 5

We turn now to the argument for the supervenience principle as it applies to cases of microreduction in natural science. We shall begin the argument with a discussion of the metaphysical status of theoretical-observational correspondence principles, both primary and secondary. As we observed in chapter III, the role of correspondence principles within scientific theory is central and vital. One major purpose of scientific theory is to explain and predict observable phenomena, and correspondence principles have the role of linking the observable phenomena to the sometimes unobservable theoretical entities and processes which underlie them. But although the principles are vital to scientific theories, they are not mere rules of linguistic usage which are true by stipulation. As we saw above, they must be treated as substantive empirical components of theory, and not as rules.

Even though these principles are vital to a scientific theory, however, it seems clear that in a certain sense they are not "basic" statements of the theory. The basic statements, it seems, are those which describe the micro-level entities, processes, and laws which the

theory postulates. These basic entities and processes underlie the observable phenomena with which they are associated by means of theoretical-observational correspondence principles. It is more than mere historical accident that certain observable phenomena are initially associated with certain theoretical phenomena. The assumption that is normally operative is that the observable phenomenon is the cumulative result of the micro-phenomena. Consider, for example, the theoretical notion of electron transition in the Bohr theory of the atom. The theory associates the wave length of a light ray emitted by an atom with the jump of an electron from one of its permissible orbits to another. The theoretical notion of an electron jump is thereby linked to the experimental notion of a spectral line, via the electromagnetic theory of light.¹⁹

The spectral lines that result from specific levels of electron jumps are the cumulative effects of the latter micro-processes, in the following sense. The difference in energy-levels between the electron's original orbit and its new orbit is just enough that, whenever a jump from the first orbit to the second is made, the same wavelength of electromagnetic radiation is emitted. And the specific wavelength is the one which leaves a line in a particular position in the spectrum of a given chemical element, according to electromagnetic theory. So the observable spectral lines are the direct macro-level effects of unobservable micro-level processes. A continuity of micro-level processes with macro-level processes is thus involved in the correspondence principles which connect specific electron jumps to particular observable spectral phenomena.

Underlying this example is a very general principle of theory construction: the principle of the continuity of micro-level phenomena with macro-level phenomena. If we adopt an attitude of thoroughgoing realism with

respect to theoretical entities--if we assume that they really exist, and that they really do possess properties like size and mass--then the observational components of theoretical-observational correspondence principles must be viewed as the cumulative results of micro-processes involving the theoretical entities. There is a continuity from the micro-level to the macro-level, since publicly observable macro-processes are the "resultant" of innumerable micro-processes.

The principle of micro-macro continuity constitutes an important methodological constraint upon the adoption of theoretical-observational correspondence principles. These principles cannot simply be stipulated haphazardly, and of course they never have been in actual scientific practice. In practice, I think it's fair to say, scientists have indeed viewed the observable phenomena which they seek to explain as the cumulative result of the theoretical phenomena in terms of which they construct explanations. And it is clear that they should treat the observable phenomena in this way, because otherwise it is not at all clear how the theoretically described phenomena could provide any genuine explanation of the observable phenomena. Purely haphazard theoretical-observational correspondence principles would not permit genuine scientific explanations of observable phenomena, since the correspondence principles would themselves lack justification. A theoretical-observational correspondence principle is justifiable only insofar as it treats the observable phenomena as the cumulative outcome of the theoretically described phenomena. I know of no other adequate justification for the adoption of these principles, either in scientific theorizing or in discussions by philosophers of science.²⁰

So the micro-macro continuity principle has the status of a methodological constraint on the theoretical-observational correspondence principles of a theory, a

constraint which is grounded on the metaphysical requirements for adequate explanation of macro-level processes on the basis of micro-level processes. In fact, we can even view the correspondence principles as asserting that the observational phenomena which they describe are the cumulative result of the micro-level phenomena they describe. This makes the micro-macro-continuity perfectly explicit. (For example, the Bohr theory of the atom asserts that observable lines in the spectrum of a given chemical element are the cumulative result of millions of electron quantum-jumps from one specified energy-level to another.)

In light of the methodological constraint on correspondence principles which is imposed by the principle of micro-macro continuity, we can discern an important element of truth in earlier characterizations of these principles as rules of correspondence. When theoretical concepts are introduced in an effort to explain certain observable phenomena, one condition which we place on the postulated theoretical processes from the outset is that the given observable phenomena must be the cumulative result of these putative processes. Thus, the theoretical-observational correspondence principles have a regulative status in theory construction from its inception. These principles must continue to be held true for the most part, as we go about the business of refining and elaborating the theory, if the theory is to successfully explain the observable phenomena it was introduced to explain. Of course this regulative status does not mean that the correspondence principles are immune from revision. They are empirical propositions themselves, and they may very well be altered or replaced in the course of the ongoing effort to achieve a great "closeness of fit" between scientific theory and empirical data. Nevertheless, a theory must continue to treat the observable

phenomena as the cumulative outcome of the postulated theoretical processes, or else the link between theory and observation will be broken.

I have expressed the principle of micro-macro continuity by saying that observable phenomena are the "cumulative result" of the theoretical phenomena with which they are associated. We turn now to the task of explicating (or at least partially explicating) the principle as so characterized. When we say that an observable process or state of affairs is the cumulative result of numerous micro-level processes or states of affairs, we seem to have in mind at least facts of the following sort. First, the observable entities instantiated in the processes are composed of, or partially composed of, the theoretical entities mentioned by the relevant correspondence principles. Secondly, the observable process itself is determined by the theoretical laws which govern the micro-entities underlying the process.

The first of these two conditions, concerning the composition of the relevant observable entities, will normally be reflected in the correspondence principles themselves. I.e., it will be clear from the wording of a given correspondence principle that the kinds of observable entities it refers to are composed of, or partially composed of, the kinds of theoretical entities it refers to. So we will focus on the second condition, involving the determination of observable processes by the theoretical principles which govern the component theoretical entities. The notion of determination is by now quite familiar. I submit that it should be explicated here in the same way as in chapter II--by reference to possible worlds. Recall that we are regarding a theory T as consisting of two components: the internal principles I and the (primary) correspondence

principles Bp. The assertion we seek to explicate is this: For any true theory T, where T is identical to $c(I \cup Bp)$, the truth of the primary correspondence principles Bp is determined by the truth of the internal principles I. The natural explication, I suggest, is as follows:

- 1) For any true theory T, the primary correspondence principles Bp of T obtain in all possible worlds in which the internal principles I of T obtain.

As I noted in chapter II, it is difficult to find any alternative way to explicate the notion of "determination" in cases where entailment is not involved, and in general the internal principles of a theory do not entail the correspondence principles. We should note, however, that the notion of "determination" itself may not do full justice to the micro-macro continuity principle. This principle not only asserts that the internal principles of a theory determine the truth of the bridge principles; it also claims that the mode of determination involves "cumulative outcome" or "resultant effect." Statement 1, therefore, evidently does not state a sufficient condition for the truth of the micro-macro continuity principle. But it does state a necessary condition: we surely cannot regard a given observable process as the "cumulative result" of certain micro-processes unless the laws governing these micro-processes determine the occurrence of the observable processes. For the purpose of our subsequent argument, it will be enough that statement 1 expresses a necessary condition for the truth of the micro-macro continuity principle.

Another point deserves re-emphasis here. As I mentioned in chapter III, theoretical-observational correspondence principles often carry an understood ceteris paribus clause. For example, specific increases

in the height of a column of mercury in a thermometer will correspond to specific increases in temperature, ceteris paribus. If the thermometer is not properly constructed, or if the tip of the thermometer has recently been dipped in icewater, then the correspondence principle might fail. In general, the correspondence principles obtain only insofar as there are no such mitigating factors present. Furthermore, the presence of such factors allows a theoretical explanation of particular instances in which the usual correspondence principles fail. Since correspondence principles have a ceteris paribus aspect in the actual world, they also will have this aspect in other possible worlds in which the internal principles of a theory are true. And mitigating factors may arise in these possible worlds which do not normally arise in the actual world. This possibility creates no problems, as long as the presence of the mitigating factor allows for a theoretical explanation of the failure of a correspondence principle. On the other hand, statement 1 rules out the possibility of in-principle unexplainable failures of the correspondence principles.

We have established that the micro-macro continuity principle expresses a basic methodological constraint on the formulation of theoretical-observational correspondence principles, and we have seen that this principle implies statement 1. But statement 1 only deals with primary correspondence principles, those which are part of an isolated scientific theory. We now come to the question of auxiliary correspondence principles. These are the ones which become associated with a theory T_1 when a higher-level theory T_2 is reduced to T_1 through a set $R_{1,2}$ of connecting principles. Auxiliary correspondence principles, a subset of $R_{1,2}$, are the reductive "translations," so to speak, of the primary correspondence principles of T_2 . Observational phenomena which

originally were associated with T_2 's theoretical properties are associated, via auxiliary correspondence principles, with the correlates of these T_2 properties. So, let P_2 be a theoretical property of T_2 and P_1 a theoretical property of T_1 , such that the two properties are linked by a biconditional connecting principle. Then for any primary correspondence principle of T_2 which involves property P_2 , there will be an auxiliary correspondence principle in $R_{1,2}$ which is identical in content, except that the reference to P_2 is replaced by reference to P_1 .

Our above discussion of the micro-macro continuity principle applies, mutatis mutandis, to the case of auxiliary correspondence principles. Consider an observational correspondence principle which involves the higher-level theoretical property P_2 . If we wish to assert a parallel correspondence principle linking the relevant observable phenomena with P_1 , the putative micro-correlate of P_2 , then we must treat the observable phenomenon as the cumulative result of processes involving P_1 . The reasons are exactly the same as in the case of primary correspondence principles. Viz., the observable process cannot reasonably be regarded as explainable on the basis of the presence of property P_1 unless the process is the cumulative result of micro-processes involving P_1 . If a theoretical-observational connection is postulated without the accompanying assumption of micro-macro continuity, then the alleged connection will at best be a statement of unexplained co-incidence of theoretical and observational phenomena, and at worst a purely arbitrary postulation with no justification whatever. If the observable phenomena are to be explained on the basis of the associated theoretical phenomena, the former must be the cumulative result of the latter. Otherwise, the putative auxiliary correspondence principle will be inadequate for its basic role--the role of establishing the

explainability of observable phenomena on the basis of theoretical phenomena.

So the principle of micro-macro continuity constitutes a methodological constraint on auxiliary correspondence principles, as well as primary ones. Hence, the auxiliary correspondence principles associated with a given theory will be determined by the theory's internal principles, just as the primary correspondence principles are so determined. Once again, the appropriate determination thesis itself is evidently best explicated by reference to possible worlds; no other type of explication seems adequate. The determination thesis for auxiliary correspondence principles is as follows. (Recall that auxiliary correspondence principles involved in a particular instance of microreduction are a subset of the class of microreductive connecting principles.)

- 2) For any two true theories T_1 and T_2 , such that T_2 is microreducible to T_1 via connecting principles $R_{1,2}$, the auxiliary correspondence principles $R\text{-}Aux_{1,2}$ obtain in all possible worlds where the internal principles I_1 of T_1 obtain.

Statement 2 emerges as a consequence of the principle of the continuity of micro-level phenomena with macro-level phenomena. The latter principle expresses an indispensable methodological constraint on primary and auxiliary correspondence principles. The reason why this constraint is vital, as we saw, is that it reflects the metaphysical conditions which must be met before an observable phenomenon can properly be explained on the basis of underlying theoretical micro-phenomena.

Statements 1 and 2, whose truth I endeavored to establish, are directly relevant to the truth of the supervenience principle (S). In particular, statement 2 entails that in all T_0 K-worlds, the auxiliary correspondence principles of unified science obtain. This

result is one of the three conditions which must be established in order to demonstrate the truth of (S). It now remains to argue for the remaining two conditions: a) that higher-level theoretical models are constructable, in all T_0 K-worlds, on the basis of the T_0 model; and b) that in no T_0 K-world is there a set of putative connecting principles for unified science, other than the set R, which is comparable to T in simplicity and which meets the conditions of model constructability and truth of the correspondence principles.

Before we turn to these claims, however, we should consider the relevance of statement 1 above to the supervenience principle. Recall that the supervenience principle asserts the following: in all K-worlds where T_0 obtains, the connecting principles R of unified science also obtain. Statement 1 has no direct bearing on (S) as so stated. However, we should perhaps reconsider our formulation of (S) in light of the conception of theories we have adopted in this chapter. A theory T, on our present construal, consists of a set of internal principles I and a set of primary correspondence principles B_p . Statement 1 tells us that the truth of the latter component is determined by the truth of the former. Thus, the primary connecting principles of theory T_0 are themselves determined by T_0 's internal principles. This fact is of considerable importance, if we recall the various intuitive notions that (S) was designed to capture: the idea that theory T_0 contains the only fundamental laws of unified science, the idea that higher-level theoretical properties are "nothing else than" the microreductively correlated lower-level properties, and the idea that the truth of statements about higher-level properties is determined by the truth of statements about the entities described by fundamental microphysical laws. When we reconsider these intuitive

principles, it becomes clear that the fundamental laws they refer to are the internal principles of T_0 . These determine the truth of all the other logically-primitive principles of unified science, including the primary correspondence principles of T_0 itself. Accordingly, we should modify our formulation of the supervenience principle. Let I_0 be the set of internal principles of T_0 , and Bp_0 the set of primary correspondence principles of T_0 . As before, R is the set of microreductive connecting principles of unified science. The modified supervenience principle is this:

(S*) In all K-worlds in which I_0 obtains, the correspondence principles Bp_0 and the connecting principles R also obtain.

We have already established that the principles Bp_0 obtain in all K-worlds where T_0 obtains; this conclusion follows from statement 1 above. Hence, it will be sufficient, in defending (S*), to defend the supervenience principle as originally stated. I.e., we must show that the principles R obtain in all K-worlds where T_0 obtains.

Pursuant to this task, we have already shown that the auxiliary correspondence principles in R , R -Aux, obtain in all T_0 K-worlds. We turn now to the constructability of higher-level models on the basis of the model for T_0 entities which is specified by T_0 's internal principles. Consider what is involved in constructing a higher-level theoretical model on the basis of the internal principles of T_0 . We are given a set of putative thing-identities: statements with certain complex sortal predicates of T_0 . Suppose S_0 and S_i are two sortal predicates from theories T_0 and T_i respectively, and suppose that the statement " $(x)(S_0x \text{ iff } S_ix)$ " is under consideration as a potential microreductive thing-identity. Before the statement can be accepted, it must be shown that the T_i model for entities of kind S_i

is satisfied by entities of kind S_0 . I.e., we must construct the model for such S_1 entities by applying the model and theoretical laws of T_0 to entities of kind S_0 . This task of construction is accomplished by showing that the applicability of the higher-level model to entities of kind S_0 is entailed by the internal principles of T_0 . (Recall our earlier example of the Bohr model of the atom and the "elastic sphere" model of the molecule in molecular chemistry. The applicability of the latter model, with respect to the complex Bohr-structures with which molecules are microreductively identified, is entailed by theoretical laws and the model of the Bohr theory itself.)

Given the nature of model-construction as we have described it, little argument is needed to show that complex T_0 entities which satisfy certain higher-level models in the actual world will satisfy the same models in all T_0 K-worlds. Satisfaction of the higher-level models is entailed by the internal principles I_0 of T_0 . Therefore, model constructability will obtain in all K-worlds where the internal principles I_0 obtain, and hence in all K-worlds where T_0 obtains.

Our argument for the truth of (S) requires one further step. We must show that the connecting principles R of unified science satisfy the uniqueness requirement in all T_0 K-worlds. I.e., we must show that in no T_0 K-world is there an alternative set of putative microreductive connecting principles for unified science, R' , which is comparable to R in theoretical simplicity and which meets the requirements of model constructability and truth of auxiliary correspondence principles. This assertion too will require little argument. For suppose there were such a set of putative connecting principles in some T_0 K-world, world w . This would mean that observable phenomena which are accounted for one way by means of the auxiliary correspondence principles of R ,

could be accounted for another way by means of the putative auxiliary correspondence principles of R' . The condition of micro-macro continuity would be met by these alternative correspondence principles, as well as by those of R . Also, the putative thing-identities of R' would meet the condition of model constructability, as do the thing-identities of R . Now, supposing that such an alternative set of connecting principles were available in the world w , we could conclude that it is also available in the actual world. After all, its availability in w would not depend on any unique features of w . Rather, it would depend on certain assumed features about theory T_0 itself--viz., that observable phenomena can be systematically connected with T_0 's theoretical phenomena in a way that's incompatible with the way indicated by the correspondence principles of R , and that higher-level theoretical sortal predicates can be successfully applied to complex T_0 entities in some way other than the way specified by the principles R . If theory T_0 meets these conditions in world w , then it also meets them in our world. But in fact it does not meet them in our world, since our world has a unified science based on the uniquely simple set of microreductively connecting principles R . (The reader is reminded that the latter assumption about our world underlies the entire discussion of the supervenience principle in this chapter, and that the assumption is central to this entire thesis.) Thus, the uniqueness of the connecting principles R in our world suffices to establish their uniqueness in all T_0 K-worlds.

Let me put this argument in slightly different terms. In all T_0 K-worlds, the same sorts of micro-level entities exist as in our world, and the same micro-physical laws apply to these entities. Thus, the same kinds of cumulative macro-level outcomes of micro-level processes are possible in all T_0 K-worlds. And the same

kinds of complex T_0 entities will be possible in all such worlds. Accordingly, if an alternative set of connecting principles R' is available in one T_0 K-world, it will be available in all. So, since such an alternative set is not available in our world, it will not be available in any T_0 K-world. Therefore, we are assured that the uniqueness condition for the connecting principles R is satisfied in all T_0 K-worlds.

This completes the argument for the supervenience principle (S), with respect to cases of microreduction in natural science. I have argued, first, that three criteria must be met in order to guarantee the truth of the connecting principles R in all T_0 K-worlds: a) the truth of the auxiliary connecting principles in all such worlds, b) the constructability of higher-level theoretical models in all such worlds, and c) the uniqueness of the connecting principles R in all such worlds. Secondly, I have argued that each of these criteria is indeed satisfied, with respect to the microreduction of higher-level natural sciences to T_0 . Together, the two components of the argument establish the truth of (S), with respect to microreduction in natural science. In addition, we saw in the course of the argument that T_0 's primary correspondence principles are themselves supervenient upon the truth of the internal principles I_0 . Combining this conclusion with the truth of (S), we may conclude that (S*) is also true, with respect to natural science. This latter assertion, as we saw, is better suited than (S) to express the full content of the intuitive principles which underlie the N-D argument. The only truly fundamental principles of unified science are those which specify the nature and the behavior of the basic concrete indecomposable entities of our world--viz., the internal principles of theory T_0 . In subsequent chapters, references to "the supervenience principle"

will be understood to designate (S*), rather than (S).

¹My description will rely heavily on Hempel's paper, "On the 'Standard Conception' of Scientific Theories," Minnesota Studies in the Philosophy of Science, 4, ed. M. Radner and S. Winokur (Minneapolis: University of Minnesota Press, 1970), pp. 142-163.

²For example, G. Maxwell, "The Ontological Status of Theoretical Entities," Minnesota Studies 3, and H. Putnam, "What Theories are Not," Logic, Methodology, and Philosophy of Science, ed. E. Nagel, P. Suppes and A. Tarski (Stanford: Stanford University Press), pp. 240-51.

³Cf. Putnam, *ibid.*

⁴Ernest Nagel, The Structure of Science, p. 90. Quoted by Hempel, p. 156.

⁵For further argument along these lines, as well as a more detailed critique of the standard conception's account of the role of models, see Marshall Spector, "Models and Theories," BJPS 15 (1965): 121-42.

⁶Putnam, pp. 249-51 and Hempel, pp. 162-23.

⁷I have not found this objection formulated explicitly by others, but similar observations are made by Fredrick Suppe, "On Partial Interpretation," JP 68 (1971): 57-76. See pp. 67-70 in particular.

⁸Alfred Tarski, "The Concept of Truth in Formalized Languages," in his Logic, Semantics, Metamathematics, trans. J.H. Woodger (Oxford: Oxford University Press, 1956), pp. 152-87.

⁹For further details on such a proposed modification of the standard conception, see Suppe.

¹⁰Rudolph Carnap, "Empiricism, Semantics, and Ontology," Revue Internationale de Philosophie 11 (1950): 20-40; Carnap, "The Methodological Character of Theoretical Concepts," Minnesota Studies in the Philosophy of Science, 1, 44-45.

¹¹Hempel, pp. 142-43.

¹²Willard van Orman Quine, Word and Object (Cambridge, Mass.: MIT Press, 1960), pp. 32-33.

¹³*Ibid.*, pp. 35-36.

¹⁴Ibid., pp. 42-44.

¹⁵Ibid., p. 44.

¹⁶The requirement that higher-level models be constructable from lower-level models commits us to the position that T_0 must provide some sort of model for at least some of the entities it countenances. I am strongly inclined to accept this position, since the alternative seems to be an instrumentalist interpretation of microphysics. But because of my nearly total ignorance of quantum mechanics and related disciplines, I can't discuss this issue with any competence.

¹⁷Interconnectedness of models doesn't come into play in this example, because there really isn't any model within thermodynamics that's associated with temperature. Temperature is a property of macroscopic entities and systems. Models generally come into play when the entities described by a theory are microscopic.

¹⁸At this point, it might occur to the reader that we should opt for sortal essentialism: the view that complex T_0 entities possess, in all possible worlds, the same sortal properties from higher-level theories which they possess in the actual world. I do not assume the truth of sortal essentialism in advance, but a version of this view is entailed by the supervenience principle. I return to sortal essentialism in the appendix.

¹⁹This example is taken from Nagel, The Structure of Science, pp. 94-95.

²⁰This claim needs qualification. The claim does hold true, I think, when the theoretical entities in question are micro-level entities. When we move to the macro-level (as in the case of psychology, where the relevant theoretical entities are entire human organisms), the relation of the theoretical to the observational may be somewhat different. For example, certain inner psychological properties may be linked to observable behavior by their analytically-characterized causal role in bringing about the behavior. We consider this type of theoretical-observational connection in the latter part of chapter V.

V. PSYCHOLOGY AND THE PART-WHOLE SUPERVENIENCE PRINCIPLE

Section 1

In this chapter we consider the supervenience principle with respect to the microreduction of psychology. Phenomenal properties on the one hand, and cognitive and volitional properties on the other, each raise special difficulties which were not encountered in our above defense of (S) and (S*) as applied to natural science. With respect to the aspects of psychology that involve phenomenal properties, we will find that problems arise because the principle of micro-macro continuity evidently does not apply to microreductive theoretical-phenomenal correspondence principles. Hence, we need an independent argument to establish the supervenience of these principles. With respect to the cognitive and volitional aspects of psychology, difficulties arise because it appears that the prospect of strong Nagel-reducibility is rather bleak. I will suggest a more plausible form of reduction, and I will argue that the supervenience principle applies with respect to the connecting principles involved in this weaker type of reduction.

We turn first to the consideration of (S*) with respect to those parts of psychology involving phenomenal concepts. The uniqueness of this envisioned instance of microreduction results from two facts: 1) instances of phenomenal properties are directly experienced, and 2) instances are directly experienced only by the organism who is itself instantiating the property. As a conse-

quence of these facts, the reduction of the relevant parts of psychology will involve connecting principles which serve two roles simultaneously: a) the role of "theoretical-theoretical" connecting principles which treat phenomenal properties as subject matter for the internal principles of psychology; and b) the role of theoretical-phenomenal correspondence principles. The envisioned correlation of pain with the neural property P is an excellent example: pain is experienced by the subject as phenomenally present, and it is also the subject of psychological theorizing.

The uniqueness of this case stands out in even sharper contrast when we consider it in light of the arguments just presented in support of (S*). The fundamental notion employed was that of the continuity of micro-phenomena and macro-phenomena. For example, the mercury in the capillary tube undergoes visible expansion upon increase in temperature because the millions of invisible molecules are increasing their average distance from one another, as the result of increased mean kinetic energy. The continuity here centers around concepts like distance and volume. Distances between molecules are like visible distances, only smaller. Accordingly, increases in distance among the invisible molecules result in a visible increase in the volume of the mercury.

We saw that the principle of the continuity of micro-level phenomena with macro-level phenomena is fundamental to all observational-theoretical correspondence principles, primary as well as auxiliary. Here is where the sharp contrast with the present case of microreduction emerges. Continuity of micro-level phenomena with macro-level phenomena is not present in the case of theoretical-phenomenal correspondence principles. A characteristic of phenomenal experiences is that they are qualitatively unique: there is an un-

bridgeable qualitative gap between the third-person world of physical things and physical process, and the first-person world of phenomenal sensation. The sensuality of phenomenal experience does not emerge as the obvious resultant extension or cumulative effect of micro-processes in the brain. On the contrary, there is something qualitatively new about the visual sensation of redness, the tactual sensation pain, and the olfactory sensation of Chanel No. 5. There is "something that it is like" to experience these sensations, and the only way to know what it is like is to actually undergo them yourself, or at least to undergo similar sorts of experiences. No amount of insight into the physiological processes underlying these experiences will give us any inkling of what they are like qua experiences. There is a glaring lack of continuity between physical processes and phenomenal experiences.

The fact that phenomenal awareness occupies a wholly different qualitative dimension than the dimension of the physical has been a major source of perplexity over the mind-body problem. And this fact makes itself felt with respect to the problem of defending (S*), the principle which will be central to the solution I shall eventually propose. Because of the qualitative gap between the phenomenal and the microphysical, the micro-macro continuity principle cannot be used to defend the supervenience of phenomenal mental properties upon the correlated neurophysiological properties. A different defense must be found.

Before embarking on this defense, we should note that the lack of continuity in the present case of micro-reduction does not seem to alter the following intuitions: that pain is "nothing but" the correlated neurological property, that this correlation needs no explaining, and that only the laws of T_0 should count as fundamental laws

in unified science. The fundamental intuitions behind the N-D argument thus do not seem to conflict with the qualitative gap between the microphysical and the phenomenal. Rather, it seems that phenomenal experiences occupy the role of the "inner aspect"--the felt aspect--of neurophysiological processes. This "inner aspect" notion does not conflict with the principles (E), (F), (D), and (N). I.e., it is consistent with the supervenience of phenomenal properties upon the correlated physical properties. It does seem to involve the view that the properties are not identical--that phenomenal properties are individuated by their "raw feel," and hence that they are instanced in some possible worlds in ways other than by the instantiation of their actual neural correlates. But of course, this non-identity position is one I have already advocated in chapter III.

In facing the problem of defending the supervenience of phenomenal properties upon the correlated neurophysiological properties, we must address the question of the relation between possibility and imaginability. It has often been assumed by philosophers that whatever is imaginable or conceivable is possible. Contradictions are not supposed to be imaginable, so they're not possible. And some philosophers have held that certain statements other than outright contradictions are such that we can't imagine their being true. (For example, Kripke maintains that we can't imagine a world in which Nixon has different parents than his actual parents, or a world in which a table which is wooden in our world can be the same table and yet be composed of ice. Parentage and composition therefore are essential properties of things, according to Kripke. A particular individual has the same parents and the same composition in all possible worlds.¹)

The notion of conceivability or imaginability is relevant to our earlier argument in support of the view

that microreductive theoretical-observational correspondence principles are true in all K-worlds where the reducing theory is true. As a result of the micro-macro continuity that's involved in these correspondence principles, we can't really imagine K-worlds where the reducing theory is true but the correspondence principles are not. For example, it seems clear that we can't imagine a possible world in which 1) the individual molecules in a quantity of mercury undergo the same increase in their average distance from one another as a result of increased mean kinetic energy as they do in the actual world, but 2) the total volume of the mercury differs from the total volume in the actual world, even though the initial volume and temperature were the same in the two worlds.

But just as the notion of micro-macro continuity does not seem applicable to the case of theoretical-phenomenal correspondence principles, the question of imaginability becomes somewhat problematic in the case of these correspondence principles. Two questions must be considered: 1) Can we imagine possible K-worlds in which all of physical theory is true, but in which the theoretical phenomenal correspondence principles which hold in our world do not obtain? 2) How does this question of imaginability related to the question of whether there are such possible worlds? My defense of the supervenience principle for phenomenal properties will be predicated on the conclusions we reach concerning these two questions.

One way in which we can perhaps imagine different psychophysical correlations obtaining, relative to the truth of physical theory, is to consider possible worlds whose basic indecomposable components include Cartesian souls. If the mind is a separate entity than the body, then psychophysical correlations might vary from one world to another without any change in the truth of

physical theory. But of course, the supervenience principle does not apply to such worlds. It applies to K-worlds: worlds whose basic indecomposable entities are all of the sort which exist in the actual world. (I'm assuming, of course, that Cartesian souls don't exist in the actual world.)

There is at least one acceptable sense of 'possibility' according to which the truth of an assertion can be "possible" even though there are no possible worlds in which it is true. I'm referring to what has sometimes been called epistemic possibility. A statement is epistemically possible if, for all we know, its truth value could turn out either way. An example often cited is Goldbach's Conjecture. This is a mathematical assertion which is true in all possible worlds if true at all, and false in all possible worlds if false at all. But it is "possible" that it's true and "possible" that it's false, in that we don't know the truth value which it necessarily possesses.

Related to the epistemic sense of possibility is a certain rather loose sense of imaginability. We can imagine Goldbach's conjecture having either truth value, in that we can imagine what it would be like to read in the newspaper that a renowned mathematician had proved (or disproved) it. We can also imagine what it would be like to read and understand the proof, or even to discover the proof ourselves. This sort of imagination is a very loose affair: we simply imagine an experience that is roughly the same as reading or discovering any other proof, except that the last line of the proof is Goldbach's conjecture (or its negation). The salient feature of this sort of "imagination" is sketchiness and non-specificity. Imagining proving Goldbach's conjecture is really no different from imagining proving any other mathematical assertion.

Imaginability in this loose sense does not seem to be restricted to statements whose truth value is presently in doubt. We can imagine Godel having proved the completeness of arithmetic. Notice, however, that there is some tendency to say that one can't imagine Godel's theorem really being false -- one can only imagine what it would be like to prove the completeness of arithmetic. We'll return below to this linguistic nuance.

These observations indicate that certain assertions can be epistemically possible even if they're not metaphysically possible, and they also indicate that 'imaginability' can be understood in so loose a sense that even truths that we know to be necessary can be imagined false. We must therefore focus our discussion somewhat more narrowly. We are concerned with metaphysical necessity and possibility, and we are concerned with the concrete, specific imaginability of the truth of certain assertions, rather than the vague sort of imaginability discussed above. With these points in mind, let's address the first of our questions concerning the supervenience of phenomenal properties: Can we imagine a possible K-world in which all of physical theory is true, but in which the theoretical-phenomenal correspondence principles which hold in our world do not obtain?

Our present lack of neurophysiological knowledge concerning the correlates of phenomenal properties is not directly relevant to this question, since lack of knowledge only establishes an epistemic possibility of other correlations obtaining than those which actually do obtain. So the question is not whether we can "imagine" different alternative psychophysical correlations being discovered, in the sense that any of these alternatives might be the correlation law which actually obtains, for all we know. Rather, the question is whether we can concretely and vividly imagine two K-worlds in which the physical theory of our world obtains, but in which the

psychophysical correlation laws are different.

It seems that the answer is yes. Consider the relevant sort of experimental situation: a conscious, awake subject is reporting his phenomenal states honestly and accurately. Meanwhile, complex electronic apparatus that's been wired into his brain by the experimenter is accurately monitoring the physical events in his brain. Now, it seems that we can imagine two K-worlds, both of which are the same as ours in point of physical laws, in which the experimental results are different. Different psychophysical laws are discovered. Moreover, we can add indefinitely many specific details to the imagined cases without changing the result. The cases are therefore quite different from the case in which we "imagine" -- in a loose and nonspecific way -- proving Goldbach's conjecture. We can imagine the experimenter implanting electrodes at very specific sites in the brain, we can cite in precise detail the different experimental results which we imagine the experimenters to obtain in the two worlds, and so on. In short, we can imagine the two envisioned experiments in minute detail and specificity -- something we clearly cannot do when we "imagine" proving Goldbach's conjecture.

Our ability to so imagine is not surprising, however. It results from the very fact which forced us to give separate consideration to phenomenal psychology in the first place: the lack of continuity between the micro-physical and the phenomenal. Since the phenomenal is in a different qualitative dimension from the physical, we are free in our imaginations to vary this dimension with respect to the physical dimension. I.e., we are free to imagine K-worlds where different psycho-physical correlations obtain, even though the same physical theory obtains.

This brings us to our second question: Does the imaginability of such worlds imply that they really are

such worlds? Given our affirmative answer to the imaginability question itself, it's clear that we must answer this question negatively if we are to defend the supervenience of phenomenal properties.

To begin with, let me mention an inclination which I feel, and which the reader will perhaps share. I'm inclined to say this: although I can imagine, in the way just cited, what it would be like to discover different psycho-physical connections than those which obtain in the actual world, I cannot imagine such a discovery really being made in a K-world whose physical laws are the same as ours. This linguistic nuance was mentioned above in connection with Godel's theorem. We have an inclination to say that although we can imagine what it would be like to prove Godel's theorem false, we can't imagine it's really being false in some world. We say this because we know that Godel's theorem is necessarily true. Thus, the assertion that we "can't imagine it's really being false in some world" reflects our belief that there are no possible worlds in which it is false. Imaginability in this sense presupposes a prior belief in metaphysical possibility.

The locution functions similarly in the case of phenomenal properties. When I say that I can't imagine a discovery of different psychophysical connections really being made in a K-world where the physical laws are the same as ours, I mean that I don't believe that there are any such K-worlds. If the reader shares my inclination to say this, he no doubt shares my belief.

Of course, the inclination is based on an acceptance of the intuitions underlying the N-D argument as they apply to phenomenal properties, and on an acceptance of the supervenience principle (S*) as an explication of those intuitions. If we subscribe to the N-D principles (N), (E), (F), and (D) as applied to psychophysical correlation laws, and if we subscribe to (S*) as an

explication of these principles, then we cannot also subscribe to the view that vivid imaginability is the criterion of possibility. Something has to give: either the N-D principles, of the contention that (S*) explicates them, or the imaginability criterion.

We thus arrive at a prima facie case against the imaginability criterion. Many philosophers are inclined to accept the premises of the N-D argument as applied to phenomenal properties. (Recall that Feigl and Smart's original version of the argument applied specifically to phenomenal properties.) And we saw in chapter II that the supervenience principle constitutes an explication of the N-D intuitions. So the philosophers who accept the N-D principles are obliged to reject the view that vivid imaginability guarantees metaphysical possibility.

They need only reject this view to limited extent, however. It still seems true in most cases. Vivid imaginability will fail to guarantee metaphysical possibility only in a limited number of cases--those in which we imagine phenomenal properties altered in their relation to the physical world, even though the physical laws remain unchanged. And the failure of the imaginability criterion in these cases is not really surprising. It results from the lack of qualitative continuity between the microphysical dimension and the phenomenal dimension, and the consequent free reign which the imagination possesses to contemplate various alternative ways that the two dimensions might be interconnected. The supervenience principle asserts that only one such system of interconnections is really possible: the actual one.

Section 2

Having sidestepped the obstacle presented by the problem of imaginability, I shall now offer positive arguments in support of the supervenience of the phenomenal

upon the physical. My general strategy will be to argue that a theory which treats phenomenal properties as supervenient has the double advantage of conceptual conservatism and simplicity, as compared with alternative theories. The two closely-related notions of conceptual conservatism and simplicity are very general principles of theory construction, which deserve to be invoked wherever possible.

The theory I'm discussing includes the reductionist's ideal of Unified Science, but it is broader than this. It also includes certain philosophical components, primarily metaphysical ones. The supervenience principle, (S*), is an example of such a component: it tells us about the nature of microreductively correlated properties, above and beyond the fact of their correlation. The thesis that such properties are identical, which I have argued against, is another example. The arguments below are designed to show that the principle of the supervenience of phenomenal properties affords a significant degree of simplicity to our envisioned philosophico-scientific theory, and to show that adoption of the principle is conceptually more conservative than failure to adopt it. (The arguments have application to cases of (S*) other than the microreduction of phenomenal psychology. But phenomenal psychology is the one instance where our earlier arguments don't apply, because of the qualitative dissimilarity problem.)

The principle of conceptual conservatism comes into play when we compare the envisioned microreduction of phenomenal psychology to other cases of microreduction. In the other scientific theories, where (S*) holds true, microreductive connecting laws between these theories are non-fundamental laws. So if the Unity of Science hypothesis is true as regards the domain of natural science, then the laws of microphysics become the only fundamental laws. Thus, reduction of the higher-level natural

sciences to microphysics does not increase the number of fundamental laws.

Now suppose that the Unity of Science hypothesis extends to psychology. The conceptually conservative approach would be to adopt (S*) in this case as well: the microreduction of psychology could thereby be accomplished without an increase in the number of fundamental laws of nature. If we do not adopt (S*) in this case, then we must add psychophysical correlation laws to the body of fundamental laws, even though this step is not taken with other sorts of microreductive correlation laws. Conceptual conservatism dictates that we keep the number of basic laws in our theory as small as possible. Hence, we should adopt (S*) for all cases of microreduction, including the reduction of phenomenal psychology.

The principle of conceptual conservatism provides us with a second argument, as follows. As we saw in the original presentation of the N-D argument in chapter II, complete microreduction of higher-level scientific theories to T_0 requires that the intertheoretic connecting principles R be non-fundamental. Otherwise, higher-level theory cannot properly be said to be reducible to T_0 ; rather, it is reducible to $T_0 \cup R$. The nonfundamentality of microreductive connecting principles has already been sustained for microreduction in natural science, as a result of our earlier arguments in favor of (S*). The following alternatives face us if we fail to adopt (S*) in the present instance: 1) We can retain the principle that in complete microreduction, only the laws of the final reducing science may count as fundamental laws; on this approach, we must reject the hypothesis that phenomenal psychology is reducible to natural science. 2) We can lower our usual standards of adequate microreduction in order to accommodate phenomenal psychology; this line entails a partial rejection of the view that complete microreduction requires that the connecting laws be non-

fundamental. Each alternative requires us to drop a key thesis that we are strongly inclined to assert: either the complete microreducibility of psychology to physical theory, or the principle that complete microreduction requires non-fundamental connecting laws. Conceptual conservatism dictates that we adopt principle (S*) as applied to phenomenal psychology rather than give up either thesis.

Conceptual conservatism also provides the basis for a third argument in favor of the supervenience of theoretical-phenomenal connecting laws. Theorists who adopt the Unity of Science hypothesis usually connect fundamentality of scientific laws with indecomposability of the entities which the laws describe. Thus, the laws of microphysics are basic because they deal with the properties and interrelations of the basic entities from which other entities are composed. The nomological nexus at the level of microphysics is thought to be complete, if the Unity of Science hypothesis is true. I.e., the laws governing all higher-level processes are determined by the laws which govern the basic components from which higher-order entities are constructed. (This type of determination, as we saw in chapter II, requires that microreductive connecting principles themselves are determined by the laws of microphysics, where the determination is expressed in (S*).)

The conception of fundamental laws as those which deal with basic entities is borne out, in most cases, as the argument of 4.4 in favor of (S*) indicates. But unless (S*) is adopted with respect to the reduction of phenomenal psychology, then the conception must be given up. So once again, the principle of conceptual conservatism dictates that (S*) be adopted for microreductive property correlations generally. In this connection, it is appropriate to quote again the following passage from J.J.C. Smart, originally quoted in Chapter II.

Certainly we are pretty sure in the future to come across new ultimate laws of a novel type, but I expect them to relate simple constituents: for example, whatever ultimate particles are then in vogue. I cannot believe that ultimate laws of nature could relate simple constituents to configurations consisting of perhaps billions of neurons (and goodness knows how many billion billions of ultimate particles) all put together for all the world as though their main purpose in life was to be a negative feedback mechanism of a complicated sort. Such ultimate laws would be like nothing so far known in science. They have a queer "smell" to them.²

I turn now from conceptual conservatism to considerations of simplicity. I shall contend that the various intuitive claims which underlie the N-D argument--(N), (R), (D), (F), and (E)--involve an important element of simplicity. This simplicity element is what makes them intuitively compelling, and it constitutes good reason for accepting the principle (S*) which explicates them.

Take the principle (N). When we assert that pain is 'nothing more than' neural property P, there is an obvious suggestion that the relation between the properties is more intimate than it would be if the assertion were false. In addition, the assertion connotes that a theory which postulates this intimate relation is simpler than a theory which does not. This connotation of simplicity is re-inforced by principles (D) and (F): we don't need to treat the law connecting pain and property P as a fundamental law; its truth is determined by the laws of natural science alone. (E) connotes simplicity as well in suggesting that it's clearly out of place to ask why the two properties are correlated.

(S*), I have alleged, explicates these closely connected intuitions. It describes the intimate connection between pain and neural property P in terms of extensions of the two properties in possible worlds other than the actual world. The suggestion is that the

extensions do not diverge from one another in any K-world where the fundamental laws of natural science obtain.

In claiming that a theory which postulates this non-divergence of extensions is simpler than one which does not, I am not invoking the traditional measures of simplicity: ontological parsimony, inductive simplicity, or the type of simplicity described by Goodman.³ An ontological parsimony argument would require us to identify phenomenal properties with their physical correlates, thereby effecting a systematic reduction in the number of distinct property-kinds that are countenanced by theory. But I have already argued against such identification in this case. Inductive simplicity is concerned with "the problem of curve fitting"--the problem of "inducing" the simplest ("smoothest") curve or function that's descriptive of a body of empirical data. Inductive simplicity isn't relevant to (S*), because the question we face regarding (S*) is whether or not there are relevant possible worlds in which the empirical data changes in certain ways --specifically, whether there are relevant K-worlds where different psychophysical correlations obtain. Goodman's approach measures the simplicity of predicate bases of descriptive systems or theories, in terms of certain formal properties of the predicates. This type of simplicity isn't relevant to (S*) either, since we are not envisioning possible worlds whose description requires different predicates. Rather, the question is whether there are relevant worlds with different psychophysical connections than ours; such correlations would involve no new descriptive vocabulary.

My claim is that supervenience is itself a type of simplicity, distinct from the types just mentioned. How can this claim be justified? To begin with, (S*) passes the test of intuition. It is backed by the Intuitions (N), (R), (F), (D), and (E). And these intuitions themselves all connote simplicity, as we saw.

But there is a more direct way to bring out the fact that supervenience is a type of simplicity. Suppose, that phenomenal properties are not supervenient on their physical correlates. This means that there are possible worlds which are just like ours at the physical level, but in which the phenomenal-physical correlation laws are different. Now, as soon as we admit such possible worlds, the following question unavoidably arises: why are the psychophysical correlations the way they are in our world? If they could have been different without any accompanying differences in the physical aspects of the world, then what explanation is there for their being as they are?

We will be told, of course, that this question is unreasonable. Psychophysical correlation laws are fundamental laws of nature: they have no further explanation on the basis of prior laws, just as the laws of microphysics have no further explanation.

But this reply is not satisfying. Psychophysical laws are just too early an end to explanation. The connection between consciousness and the physical world becomes inordinately puzzling, if this connection can differ in possible worlds whose physical aspects are just like ours. We are puzzled by the fact that we have our particular psychophysical laws and not others, given that others are perfectly compatible with the physical aspects of the world. We feel the need for an explanation of this fact, but no explanation is forthcoming. If phenomenal properties are not supervenient, then we must accept psychophysical laws as fundamental and unexplainable.

Our puzzlement is removed, on the other hand, if we adopt the supervenience principle for phenomenal properties. There is no impulse to ask why the actual psychophysical correlations obtain rather than others, since no others are possible relative to the truth of the fundamental laws of natural science. The key factor

in the simplicity of supervenience is that it eliminates puzzlement about the connections of the phenomenal to the physical.

Scientific explanation must come to an end within unified science. Certain principles must be accepted as fundamental and unexplainable, as the basis for the others. But we are not compelled to conclude that we have reached the end of explanation until we come to the level of scientific principles which govern the ultimate indecomposable entities of our world. Our unified science will be simpler if we do not declare an end to scientific explanation until we are compelled to do so.

If we adopt the supervenience principle, then the instanting of phenomenal properties can be viewed as the "inner aspect" of the instanting of the correlated physical properties. I.e., it is the nature of neural property P that its instances involve the "raw feel" of pain; this raw feel is not a part of the nature of any other property countenanced by natural science. To say that pain is the "inner aspect," or part of the "nature," of property P is to say that the link between the two is stronger than mere nomological correlation. And it is stronger, according to the supervenience principle: they are co-extensive in all relevant K-worlds.

This completes the simplicity and conceptual conservatism arguments for the supervenience of phenomenal properties upon the physical properties with which they are nomologically correlated. It must be admitted that the arguments are less conclusive than those given earlier for supervenience in other instances of microreduction. But it does not appear that stronger arguments can be given in the phenomenal case. The sharp lack of continuity between the phenomenal and the microphysical precludes any stronger argument. Specifically, it precludes any argument based on limitations in what can be imagined; it is fairly easy to imagine different cor-

relation laws obtaining than the ones which actually do obtain, precisely because the phenomenal dimension is so dissimilar from the physical.

But although a stronger argument is not possible because of this gap in continuity between the phenomenal and the physical, it does not appear that a stronger argument is needed. The supervenience principle is supported by several compelling intuitive principles, it affords a significant degree of simplicity to our envisioned Unified Science, and adopting the principle is conceptually more conservative than failure to do so. Therefore we should adopt the principle, with respect to the theoretical-phenomenal connecting principles of unified science.

Section 3

We now focus on the subject of cognitive and volitional mental properties, as against phenomenal properties. 'Cognitive' applies to properties like 'is thinking of Vienna', 'knows that the earth is round', and 'believes that the moon is inhabited'. Examples of volitional properties are 'wants an ice cream cone', 'hopes to win the race', and 'intends to leave'. In general, it appears that cognitive and volitional properties fall within the class of so-called "propositional attitudes." I shall discuss, rather speculatively, the relationship that psychology might be expected to bear to the properties countenanced by natural science, in light of the pervasive presence of cognitive and volitional notions within psychological theory. I shall suggest several reasons for thinking it unlikely that the cognitive and volitional elements of psychology are strong-Nagel-reducible to natural science. I will then describe a weaker sort of reduction, akin to weak Nagel reduction, which is more likely to be obtainable in the

case of psychology. Finally, I will contend that the supervenience principle applies with respect to the connecting principles involved in this weaker type of reduction; just as it applies to cases of strong Nagel reduction. (Much of what I shall say about psychology may be applicable to the more general problem of the reduction of social science theories to theories in natural science, in light of the pervasive presence of cognitive and volitional concepts throughout social science.)

A number of reasons have been given by various authors for thinking that cognitive and volitional properties differ in fundamental ways from the properties countenanced by the natural sciences, and hence that psychology itself differs fundamentally from the natural sciences. Among the alleged differences are these: 1) There are no precise laws involving cognitive and volitional properties; 2) Many of the principles concerning these properties that come closest to being "lawlike" are virtual analytic truths; 3) There is no neat mapping of these properties onto the properties countenanced by natural science; 4) These properties possess intentionality; and 5) Statements ascribing these properties to things contain opaque contexts.

In light of such claims, four general stances are possible. One can take the view that psychology is not fundamentally different from natural science, and then either accept or reject the hypothesis that psychology is reducible to natural science. Or, one can concede that there do exist fundamental differences, and then either accept or reject the reducibility thesis. Notice that the fourth position is compatible with Unity of Science, if we are willing to dispense ultimately with cognitive and volitional notions in science. Quine, for example, adopts the view that the intentionality of these notions and their tendency to create opaque contexts place them outside of objective science.⁴ And

he evidently holds that a reduction of an intentional psychological theory to natural science would not be possible, because of his principle of the indeterminacy of translation.⁵ But he takes this result to show the "baselessness of intentional idioms and the emptiness of a science of intention," rather than the importance of an autonomous and irreducible science of intention.

Quine's position seems to me to be unnecessarily austere. It is not clear to me what the doctrine of indeterminacy of translation implies with respect to the question of reducing intentional psychology to natural science, and I am not an advocate of this doctrine in any case.⁶ Moreover, it seems that intentional notions are quite centrally entrenched in psychology as we know it today, and in many other social sciences as well.⁷ Skinner's attempt to found a non-intentional science of behavior is one notable exception, but so far the attempts to produce an austere stimulus-response science have been rather unsuccessful.⁸

On the other hand, it remains true that there are aspects of intentional psychology which render it quite different from the natural sciences, and which suggest that strong Nagel-reducibility is unlikely. We will focus on the first three of the five aspects mentioned above. First, consider the conjecture that there are few fully precise laws involving intentional properties. The suggestion is that "tendency laws" and very vague general laws are likely to be the best that we can do in this realm. Cognitive and volitional concepts, with their origins in our common-sense conceptual scheme and their relatively loose conditions of applicability, are perhaps inherently lacking in precision. If so, then we cannot expect to discover precise, exceptionless, universal laws involving these concepts. The much-discussed lack of precise laws in the social sciences is perhaps inevitable, in light of the concepts that are employed.

Given this lack of laws, and given the common-sense origins of many of our intentional notions, it seems plausible to conjecture that there indeed are inherent limitations on the degree to which the generalizations of intentional psychology can be improved.

In Davidson's terminology, the conjecture is that the generalizations of intentional psychology are "heteronomic," as against the "homonomic" generalizations we find in the basic natural sciences. The distinction is as follows:

On the one hand, there are generalizations whose positive instances give us reason to believe the generalization itself could be improved by adding further provisos and conditions stated in the same general vocabulary as the original generalization. Such a generalization points to the form and vocabulary of the finished law: we say that it is a homonomic generalization. On the other hand there are generalizations which when instantiated may give us reason to believe there is a precise law at work, but one that can be stated only by shifting to a different vocabulary. We may call such generalizations heteronomic.⁹

This passage from Davidson also points to the relation between the heteronomic character of psychological generalizations and the thesis that psychology is reducible to natural science. For, if the precise laws underlying our psychological generalizations are not within the domain of psychology itself, they must be formulable at the level of the natural sciences. This unique aspect of intentional psychology renders reducibility all the more likely. The form which reduction is likely to take, however, is not that of the strong Nagel model. No neat one-to-one isomorphism is to be expected between psychological properties and microphysical properties, since intentional psychology lacks the precision of natural science.

The second conjecture I wish to consider concerning intentional psychology is the assertion that many psychological generalizations are virtual analytic truths, rather than empirically disconfirmable principles. It is suggested that the generalizations are so central a part of ordinary conceptual scheme that anyone who denied them would likely be accused of failing to understand the relevant concepts. The relation between intentions, beliefs, and actions is often cited as an example. As Alan Donagan puts it, "No conceivable empirical evidence would count against the proposition that if you hold to your intention to bring something about, and believe that you must take certain steps to do so, then you will take those steps if you can."¹⁰ Another example of such a psychological principle is this: "If, given that x had not been expecting p but now suddenly judged that p would be the case, x would feel joy, then x wants p."¹¹

I think it must be admitted that such generalizations do have the status of virtual analytic truths. Their centrality in our conceptual scheme is no accident: we assume their truth as a basis for our ascription of the mental properties themselves. In this regard, they contrast rather sharply with phenomenal properties, which we can ascribe (to ourselves) independently of any assumed connection to other mental properties or to subsequent behavior. For example, the fact that we tend to avoid pain does not count as a virtual analytic truth, since pains are identifiable independently of their behavior-dispositional characteristics, by means of their 'raw feel'.¹²

The relevance of the virtual analyticity of certain psychological generalizations to the microreduction of psychology will become clearer below. But for the present, we should note that we may be able to arrive at a satisfactory account of reduction which does not meet Nagel's derivability requirement with regard to these

laws. If the laws are analytically true, perhaps they don't need to be derived from more basic laws.

A third important conjecture is that many cognitive and volitional properties do not map neatly onto any physicalistic properties countenanced by natural science. There are good reasons for suspecting that this is the case. First, we ascribe these properties to people under varied and diverse circumstances, and the conditions of ascription are rather loose. Just imagine, for example, the broad range of different situations in which we might say of a person "He is thinking of Vienna." Secondly, there is the closely-related fact that instances of cognitive and volitional properties can have quite different "contents" at different times. One person's thinking of Vienna might involve a vivid recollection of the opera house he saw there, a second person's might involve contemplation on the plight of Vienna in the second world war, and a third person's might involve the planning of a stop at Vienna in his upcoming trip to Europe. In light of such diverse possible "contents," it is manifestly implausible to suppose that "is thinking of Vienna" corresponds to some unique physical property of the brain and central nervous system. And similarly for a great many other cognitive and volitional properties.

The implications of this fact for the reducibility thesis are clear. If there is no unique mapping of such mental properties onto physicalistic properties of natural science, then we cannot expect to obtain the biconditional psychophysical correlation laws which are needed for strong Nagel reduction. Reduction presumably must take some weaker form.

Before turning to a description of the weaker form that seems appropriate, let me briefly mention some connections between the above observations about intentional psychology and some recent work on the mind-body problem. The causal theory of the mental, as advo-

cated by D.M. Armstrong¹³ and by David Lewis,¹⁴ has much in common with the view that many mental notions belong to a common-sense conceptual framework, many of whose truths are virtually analytic. The causal theory, which I discussed in chapter III in connection with property identity, asserts that certain mental event-types have definitive causal roles: typical causes and typical effects under various circumstances, which by analytic necessity belong to a mental event-type and are jointly distinctive of it.¹⁵ Thus, it claims not only that there are analytic principles involved, but also that these principles are strong enough to provide a full characterization of certain psychological states in terms of their typical causal connections to one another, and ultimately to sensory stimulation and behavior.

The causal theory also makes room for the other two likely characteristics of intentional psychology noted above: the heteronomic character of psychological generalizations and the lack of a neat one-to-one relationship between intentional mental properties and physical properties. The latter possibility is allowed because there may well be different physical properties which can occupy the causal role that's definitive of a given intentional property. The possibility of heteronomic generalizations is allowed for, because nothing in the causal theory rules it out.

The functionalism of Putnam,¹⁶ Fodor,¹⁷ and Dennett¹⁸ also has much in common with the view that intentional mental states are characterized largely by their role in virtually analytic generalizations of our common-sense conceptual scheme. Putnam's analogy between the mental states of persons and the logical states of Turing machines can be understood in various ways, some more literal and specific than others.¹⁹ But at a general level, the analogy suggests that mental states are characterized in the same general way that logical

states of Turing machines are characterized--viz., by their relation to one another and to (sensory) "input" and (behavioral) "output." At this level of generality, the functionalist viewpoint merges with the causal theory. Thus, my remarks about the compatibility of the causal theory with the heteronomic nature of psychology and with the lack of mental-physical isomorphism, are applicable here too.

Section 4

What sort of reducibility of psychology to natural science might we reasonably expect, in light of the unique features of psychology that we have been discussing? Strong Nagel reduction seems ruled out, as we saw. I shall suggest in this section that a form of weak Nagel reducibility seems fairly plausible. The modification results from a weakening of the derivability requirement. Suppose we drop this requirement with respect to analytic psychological principles, precisely because these principles are true on the basis of their role in our conceptual scheme. We are then left with the task of constructing a set of connecting principles between natural science and psychology, such that the many non-analytic generalizations of psychological theory are derivable from these connecting principles in conjunction with statements from physical theory.

What form shall these connecting principles take, if they are not to be biconditionals that connect primitive predicates of intentional psychology to predicates of natural science? An alternative scenario that comes to mind involves statements of two general sorts: 1) "one-way" correlation principles, of the form " $(x)(\text{If } Px \text{ then } Mx)$ ", where 'P' represents a predicate from physical theory and 'M' represents a primitive intentional predicate;²⁰ and 2) various statements about

the scope and completeness of the set of laws in the first component. Suppose, for example, that the first component contains three laws connecting various physical properties to mental property M_i : $(x)(\text{If } P_1x \text{ then } M_ix)$, $(x)(\text{If } P_2x \text{ then } M_ix)$, and $(x)(\text{If } P_3x \text{ then } M_ix)$. And suppose the second component contains a statement asserting that the property M_i is usually (though not necessarily always) accompanied by one or another of the physical properties mentioned in these three laws. Now, if the two components contained such one-way correlation laws and "typical accompaniment" statements for enough intentional properties, then it might be possible to derive the general tendency-statements of intentional psychology from some branch of natural science. If so, then a form of reduction will have been achieved.

The limiting case of a "typical accompaniment" statement would be an assertion that instances of a given mental property M_i are always accompanied by some member of the set of physical properties mentioned in the one-way correlation laws involving M_i . But such limiting-case assertions would not be necessary in order to derive psychological generalizations which have the form of non-universal tendency statements.

Certain analytic psychological principles, on the other hand, do not have the form of tendency statements. For all their vagueness, they are still stated in universal form. The reason for dropping the derivability requirement with respect to them thus becomes clear: our set of one-way connecting laws may not be complete. Thus, we may not be able to actually derive certain analytic principles. But we know that they're true anyway, because of their analyticity.

I turn now to an argument designed to show that weak Nagel reducibility of this sort does indeed obtain, however difficult it might be to exhibit an actual reduction of intentional psychology in detail. The argument

is based on several crucial assumptions: 1) The laws of microphysics remain true of the basic microphysical entities, no matter how these entities are grouped together into "higher levels" of integration. That is, the laws governing the basic indecomposable entities of the universe remain in force even when these entities are combined to form atoms, or molecules, or cells, or multicellular organisms, or living human beings. The laws are not abrogated at high levels of organizational complexity. 2) If the basic laws of microphysics are only statistical in nature, then this statistical factor is not eliminated in the behavior of complex entities by the higher-level laws governing those entities. (Thus, if a person is only 99.9 per cent likely to perform a certain act on the basis of his present physical state, the chances of his doing the act aren't increased because he is also in a particular psychological state. The laws of psychology don't "fill in" the causal gaps which result from quantum-level indeterminacy.) 3) The analytic principles involving cognitive and volitional properties are sufficiently rich to guarantee that such properties can be characterized solely in terms of their typical connections to one another, to phenomenal properties, to sensory input, and to behavior output. I.e., the analytic principles provide "functional" characterizations of cognitive and volitional psychological properties.

The first two of these assumptions are implicit in all the standard reducibility theses: strong Nagel, weak Nagel, and Kemeny-Oppenheim. A reductionist always assumes that the laws of the reducing science are universally true at all levels of integration, or at least that whatever degree of lawlike regularity that exists in the universe is a result of these laws alone. On the basis of this assumption, the Nagel-reductionist proceeds to look for connecting principles which will

allow a derivation of higher level laws from the basic theory. The Kemeny-Oppenheim reductionist, on the basis of the same assumption, attempts to explain the observable phenomena involving higher-level entities in terms of the basic theory.

The third assumption is admittedly quite tentative. Interesting attempts in the direction of providing detailed functional characterizations of various types of mental states have been given by Armstrong and by Dennett, but whether or not the program can be successfully carried through remains to be seen. Let me emphasize, in addition, that I am tentatively adopting a functionalist viewpoint only with respect to cognitive and volitional properties. As I have already argued, it seems clear that phenomenal states are not susceptible to functional characterization, since their definitive characteristics are their experienced "raw feels."²¹ Accordingly, I need not deny that many cognitive and volitional properties often are instanced in ways that involve vivid phenomenal content. For example, instances of the property "wanting an ice cream cone" might often include a vivid sensory image of Bressler's for a given individual, and a feeling of the "gnaw" of hunger. The functionalist contention is simply that the definitive features of such a want are its tendencies to produce further mental and behavioral states, rather than its phenomenal accompaniments.

Given these three assumptions, the argument proceeds as follows. Consider first the case of mental properties which can be assigned to an organism directly on the basis of its dispositional tendencies to certain forms of overt behavior. (Wanting to raise one's arm is an example.) Suppose a person has such a property, M_1 , at time t_1 . Now, given the assumption of the universality of basic physical laws, the person will also instantiate some physical property P_1 at time t_1 , which itself is likely to give rise to the relevant overt

behavior. I.e., there must be some physical basis for the relevant disposition to overt behavior. Now, since the relevant dispositions of the organism are the basis for assigning to it the mental property M_1 , and since the universality of our physical laws guarantees that these dispositions are present whenever the organism has property P_1 , it follows that the organism will have M_1 whenever it has P_1 . Thus, we are guaranteed the existence of a "one-way" law, connecting instances of P_1 to simultaneous instances of M_1 .

Now consider a mental property M_2 , which is assigned to organisms on the basis of dispositional tendencies of the following two sorts: a) tendencies to overt behavior, and b) tendencies to result in mental properties which themselves have definitive connections to characteristic overt behavior. The argument proceeds as before. Suppose M_2 applies to an organism at time t_2 . Then the universality of physical law guarantees that there is some underlying physical property P_2 which tends to cause the relevant overt behavior, and which also tends to cause certain physical states which themselves are related via one-way laws to the mental states to which M_2 is dispositionally related. Now, since P_2 will have the same dispositional tendencies whenever it is instanced, and since these tendencies are definitive for the presence of M_2 , we are guaranteed that M_2 occurs whenever P_2 occurs.

The argument proceeds by analogous steps, to include mental properties that are progressively farther removed from direct overt behavior and which have analytic interconnections with increasing numbers of other mental states. Also, the same form of argument can be used to include certain analytic dispositional connections between cognitive or volitional properties and phenomenal properties, if we assume that phenomenal properties are themselves connected to physical properties

at least by one-way laws. (The present argument cannot be used to establish such a connection, of course, since I have denied that phenomenal properties can be functionally characterized.)

The argument as just presented made the assumption that intentional psychological properties fall into a sequence, with some intentional properties directly linked to physical input and output, others linked to physical input and output or to intentional properties at the first level, still others linked to physical input and output or to intentional properties at the first and second level, etc. Actually, however, intentional properties may not fall into a sequential ordering of this sort. So we need a different version of the argument just presented, which will apply even if the assumption of sequential ordering is false.

The argument is quite similar to that just given. Consider an intentional property M_1 , characterized in terms of its typical connections to other intentional properties, to sensory input, to behavioral output, and to phenomenal properties. (We assume the existence of pervasive physical-phenomenal correlation laws.) Suppose property M_1 is instanced in a particular person at time t . Because of the presence of M_1 at t , the person has a tendency to instantiate certain other intentional mental properties, or certain phenomenal properties, or certain behavioral properties, or perhaps properties of each kind. Given the universality of physical laws at all levels of organizational complexity, we know that at time t the person's body instantiates a physical property, P_1 , which establishes certain definite physical tendencies in the person. Because of P_1 , he has a tendency to instantiate a) certain physical properties which are connected, via physical-phenomenal correlation laws, to phenomenal properties of the sort which M_1 tends to bring about, b) certain behavioral properties,

of the same sort which M_1 tends to bring about, and c) further physical properties which in turn have the same phenomenal and behavioral tendencies as the intentional properties which M_1 tends to bring about.

Category (c) involves physical properties which would underlie occurrences of intentional properties subsequent to the occurrence of M_1 . Suppose part of the analytic characterization of property M_1 is its dispositional tendency to result in intentional property M_2 , and suppose that the person who has M_1 at time t does subsequently instantiate M_2 at time t' , as a result of M_1 . Then the person will also instantiate some physical property P_2 at time t' , a property with these characteristics: 1) the person had a tendency to instantiate this property by virtue of the presence of P_1 at time t , and 2) the presence of this property renders the person likely to subsequently instantiate a) certain physical properties which are connected, via physical-phenomenal correlation laws, to phenomenal properties of the sort which M_2 tends to bring about, b) certain behavioral properties, of the same sort which M_2 tends to bring about, and c) further physical properties which in turn have the same phenomenal and behavioral tendencies as the intentional properties which M_2 tends to bring about.

Similar remarks continue to apply, with respect to any further intentional property M_i which M_2 is likely to lead to, with respect to any still-further intentional property M_j which M_i is likely to lead to, and so on. Thus the relevant dispositions toward phenomenal and behavioral properties are "carried along," by the temporal sequence of physical events which underlies any given sequence of intentional-psychological events. The conclusion to be drawn is clear. Whenever a person instantiates an intentional mental property, he also instantiates some physical property which is such that the relevant intentional property would be attributable to

him at any other time in which he instantiated this physical property. I.e., we are guaranteed of one-way laws from the physical to the intentional-psychological. Analytically-characterized regularities involving intentional properties "ride piggy back," as it were, upon the regularities of physical theory.

Thus, our argument appears to guarantee the following result. If a functionalist account can be provided for the full range of cognitive and volitional properties that are needed for psychology, then any instance of such a property in an organism will be accompanied by some underlying physical property, such that whenever this physical property is instanced the mental property is instanced as well. I.e., we are guaranteed a pervasive system of one-way physical-mental correlation laws, even if we cannot yet state such laws in any detail.

What now of the existence of "typical accompaniment" principles? The physical properties which normally underlie instances of functional psychological properties must be such as to guarantee the truth of our empirical generalizations. Given that a pervasive system of one-way correlation laws exists, we can be assured that the typical physical accompaniments of psychological properties do indeed meet this requirement, since if they did not do so then our empirically discovered generalizations simply would not obtain. Thus, it would appear that the modified weak Nagel reducibility of intentional psychology is guaranteed by our argument. If we assume that the phenomenal properties of psychology are also connectable to physical properties, either by pervasive one-way laws or by conventional microreductive biconditional laws, then we are evidently assured of the reducibility of psychology in toto.

Actually, however, a further complication must still be dealt with. We have thus far used the notion of "behavior," or "behavioral output," rather freely.

But a crucial ambiguity is thereby concealed. Behavior of an organism, as describable by the principles of physics, must be viewed essentially as simple motion of various bodily parts or of the organism as a whole. The psychologist, on the other hand, deals not only with "raw motion" but with human action. A satisfactory microreduction of psychology would have to allow for connections between raw motion and action, as well as allowing for connections between physical events and intentional mental events.

One-way laws are not in order here. For example, not all occasions in which a person's arm rises will be occasions where he performs the action of raising his arm. In order for the action to occur and to count as a "basic" action (one which isn't performed by performing some other action), the arm-raising must be intentional. It must be caused, in a certain characteristic way, by wants and beliefs of the agent in question. Causation by wants and beliefs seems to be built into the concept of action.²² Thus, a set of connecting principles that establishes a modified weak Nagel reduction would presumably have to include certain analytically true statements, to the effect that certain actions occur simultaneously with certain motions, provided that the motions are caused in the appropriate way.

A further issue that arises concerning action, and which presumably would have to be dealt with before an adequate microreduction of psychology could be complete, is the question of the relation between basic and non-basic actions. (For example, the relation between my extending my arm and my signalling for a turn, in the situation where I signal for a turn by extending my arm.) If non-basic actions are countenanced as part of the domain of psychological theory, some appropriate further connecting principles will be required before a satisfactory microreduction of psychology to natural

science can be effected. Thus, my above argument for reducibility is incomplete insofar as it ignores this topic.

To speak glibly about this matter, I'll simply note the clear presence of intuitions of the sort that underlie the N-D argument, with respect to the relation between basic and non-basic actions. Thus, we clearly feel that signalling a turn is "nothing more than" extending one's arm, provided that a suitable convention exists. And the ultimate origin of this convention is presumably explainable in principle in terms of certain neurological processes in the brains of those who established the convention, if the Unity of Science doctrine is true. Thus, it is reasonable to conjecture that the problem of the relation between basic and non-basic actions can indeed be accommodated within a microreduction of psychology, although I have said little about how to do so. The problem of action, then, does not appear to seriously threaten the above argument concerning the microreducibility of intentional psychology to natural science.

It should also be noted, at this point, that there are special problems involved in delimiting the class of "theoretical terms" in intentional psychology. From the point of view of ordinary parlance, there are indefinitely many action-types that can be described, and indefinitely many propositional attitudes as well. An intentional verb like "thinks that," or "believes that," can be followed by virtually any sentential clause at all. Actions, similarly, can be described in a myriad of ways, many of which are a clear part of the colloquial idiom and quite far removed from the sort of discourse we consider "scientific." Thus, two closely related problems must be given more attention by scientists and philosophers than they have received so far, before the general requirements for a microreduction of psycho-

logy can be specified. First, we need some delimitation of the class of "theoretical" predicates in psychology: those predicates which must be systematically associated with predicates from natural science if microreduction is to be achieved. Secondly, we need some way of connecting the theoretical vocabulary of psychology to the ordinary psychological idiom--the idiom whose range of possible "content" for intentional verbs and for action ascriptions is virtually endless. (A theory specifying the relation between basic and non-basic actions would perhaps be a part of this latter program.) My remarks concerning the microreduction of psychology have been rather inconclusive precisely because the two problems just mentioned have not yet been addressed in any systematic way.

But although I haven't proposed a solution to the problems, the above argument for the existence of one-way connecting laws between physical properties and intentional mental properties is certainly quite relevant to both. If a given mental property counts as a theoretical property, then the existence of a one-way connecting law involving this property will be crucial for modified weak Nagel reducibility of psychology to natural science, as I described it earlier. If the mental property does not count as a theoretical property, then the existence of such a law might still be crucial in an account of the relation between "scientific" discourse and "ordinary" intentional-psychological discourse. For the remainder of the present chapter it will be assumed that at least some intentional predicates constitute theoretical predicates in psychology. We shall hypothesize that psychology is capable of modified weak Nagel reduction.

Section 5

Assuming that psychology is so reducible to natural science, the question we now face is whether the supervenience principle will apply to modified Nagel reducibility. Can we uphold this principle with respect to the class of connecting principles that would establish a weak Nagel reduction? If so, then the various intuitive principles underlying the N-D argument would apply in the present case as well. If not, then we would have to add these connecting principles to the set of fundamental laws in science, thereby violating the N-D intuitions. This consequence would be undesirable, since the full reducibility of one science to another seems to require that the intuitive principles be satisfied. They are satisfied in standard cases of microreduction, according to the argument in chapter IV. I shall contend that they are satisfied in the present instance also.

Consider first the one-way connecting laws, which link properties from physical science to intentional psychological properties. Will these laws remain true in all T_0 K-worlds? I suggest that they will, for precisely the same reasons which established that such one-way laws exist in the actual world. We have adopted a functionalist view of cognitive and volitional properties, which asserts that such properties have definitive connections to one another, and ultimately to sensory input and to behavioral output. Given this assumption, it follows that such properties will be possessed by entities in other possible worlds whenever the relevant dispositional tendencies are present in the entities. But in the relevant K-worlds, these dispositional tendencies are fixed by the physical properties of the entities, since the physical theory of our own world

(which is complete for all "levels" of integration, by hypothesis) is true in all K-worlds as well. Accordingly, the same one-way laws from the physical to the intentional-psychological will apply in all T_0 K-worlds.

Any analytically-true statement in the set of connecting principles will also remain true in all relevant K-worlds, of course. (For example, the set might include a principle which asserts that an arm-raising occurs whenever an organism's arm rises and the rising is caused (in an appropriate way) ²³ by the organism's want that it rise.) Also, the analytically-true statements which may not be derivable from the reducing theory plus connecting principles will be true in all possible worlds, and hence in all T_0 K-worlds.

What about the "typical accompaniment" principles, which allow the derivation of non-analytic psychological tendency-laws? Here, we face two possibilities: 1) the actual relative frequency of occurrences of the various possible physical correlates of any given psychological property is determined by the laws of physical theory alone; or 2) this relative frequency is not so determined, but depends upon certain antecedent physical conditions, which might be different in some other T_0 K-world. In the former case, but not the latter, the "typical accompaniment" principles are supervenient. If the latter case obtains, then these principles are not even lawlike, since they depend upon specific antecedent conditions. Hence, the psychological generalizations which depend upon the principles would also be non-lawlike.

Either possibility is acceptable, as regards the N-D intuitions. If the first obtains, then we have standard supervenience. If the second obtains we do not, but we still have an acceptable form of supervenience. That is, we are still guaranteed that in all T_0 K-worlds in which the appropriate physical antecedent conditions

obtain, the same "typical accompaniment" principles obtain as in the actual world. In all such worlds, just as in the actual world, these principles will be non-lawlike. Their truth is determined by the conjunction of basic laws and antecedent physical conditions. But their truth is determined, nevertheless, at the level of physical theory: the only difference is that their truth is determined by the conjunction of physical laws and antecedent physical conditions, rather than by antecedent laws alone.

¹Kripke, p. 314.

²Smart, "Sensations and Brain Processes," p. 143.

³For a concise description of Goodman's work on simplicity, see D. Hillman, "On the Measurement of Simplicity," Probability, Confirmation, and Simplicity, ed. M. Foster and M. Martin (New York: Odyssey Press, 1966), pp. 235-79. On the various types of simplicity, see R. Rudner, "An Introduction to Simplicity," PS 23 (1961): 109-115.

⁴Quine, p. 218.

⁵Ibid., p. 220. It isn't clear, however, whether Quine really objects to empirical microreduction, or whether he only objects to the philosophical program of giving a "conceptual analysis" of intentional statements by means of behavioristic translations.

⁶For further discussion of the relation between Quine's indeterminacy thesis and the status of psychology, see D. Davidson, "Mental Events," Experience and Theory, ed. J.W. Swanson (Amherst: University of Massachusetts Press, 1970), pp. 79-101.

⁷For some support of this claim, see A. Goldman, A Theory of Human Action (Englewood Cliffs: Prentice Hall, 1970), chap. 5.

⁸Cf. Daniel Dennett, Content and Consciousness (New York: Humanities Press, 1969), pp. 33-34.

⁹Davidson, p. 94.

¹⁰Richard Brandt and Jaegwon Kim, "Wants as Explanations of Action," JP 60 (1963): 427.

¹¹Alan Donagan, The Later Philosophy of Collingwood (Oxford: Oxford University Press, 1962), p. 185.

¹²Of course, we do need to know the behavioral-dispositional characteristics of pain in order to initially come to know which of our own various phenomenal experiences to count as pain.

¹³D.M. Armstrong, A Materialist Theory of the Mind (New York: Humanities Press, 1968).

¹⁴Lewis, "An Argument for the Identity Theory."

¹⁵Ibid., p. 25.

¹⁶Hilary Putnam, "Minds and Machines," Dimensions of Mind, ed. S. Hook (New York: NYU Press, 1960), pp. 148-479; Putnam, "The Mental Life of Some Machines," Intentionality, Minds, and Perception, ed. H. Castaneda (Detroit: Wayne State University Press, 1967), pp. 177-200.

¹⁷Jerry Fodor, "Explanations in Psychology," Philosophy in America, ed. M. Black (Ithaca: Cornell University Press, 1965), pp. 161-79.

¹⁸Dennett, Content and Consciousness.

¹⁹For criticisms of a fairly literal and specific form of functionalism, see J. Fodor and N. Block, "What Psychological States Are Not," PR 81 (1972): 159-81.

²⁰These one-way principles are simultaneity laws. I suppress the time-references for simplicity.

²¹On this point, my view differs sharply from Lewis and Armstrong.

²²Cf. Goldman, chap. 3.

²³For an explanation of this hedge-clause, see Goldman, *ibid.*, pp. 55-63.

VI. EVENTS

Section 1

We began in chapter I with a discussion of the relation between two questions: 1) whether physical events are identical to mental events, and 2) whether psychology is microreducible to natural science. I adopted a conception of event which is such that the identity thesis entails, but is not entailed by, the strong Nagel reduction thesis. Events were conceived as mixed entities. An event is the having of a property by an object (or group of objects) at a time. The identity conditions for events included identity of the constitutive properties. The psychophysical identity thesis thus entails that there are biconditional correlations between physical properties and mental properties, and such biconditional connecting laws establish strong Nagel-reducibility.

Since chapter I, our attention has focused primarily on relations between properties. There has been little attention to events as such. Our major conclusions regarding the properties involved in microreduction were these: 1) The supervenience principle obtains with respect to biconditional microreductive property-correlations; 2) This principle expresses the content of the various intuitive principles which underlie the N-D argument for property identity; 3) Such properties in fact are not identical; and 4) The supervenience principle also obtains with respect to a certain form of weak Nagel reduction, which involves one-way property correlation laws.

In the present chapter, we shall take up the topic of events. We shall consider the nature of the relation between two simultaneous events both involving the same entity, where the constitutive properties of the events are connected by a microreductive connecting principle. (We are concerned, for example, with the relation between a particular gas' having a certain temperature at time t , and the component molecules' having the appropriate mean kinetic energy at time t . And with the relation between a person's having a yellow-orange after-image at time t and the person's body having the correlated physical property at time t .) Much of the discussion will focus on the question of the relation between mental events and physical events--on the mind-body problem. But it will be applicable to virtually all cases of two simultaneous events involving the same physical entity, where the constitutive properties are connected by a microreductive connecting principle

If the constitutive properties of the respective events in such event-pairs are not identical, then the events themselves are not identical. But microreductively correlated properties are not identical, according to the argument of chapter III. (And in the case of one-way correlation laws, the properties are not even co-extensive; so they clearly are not identical.) Hence, the event-identity thesis is ruled out by our earlier considerations regarding properties.

In light of the supervenience principle, on the other hand, it is clear that a very intimate relation exists between members of the event-pairs we are considering. We shall examine the relation in some detail in the present section of this chapter. In subsequent sections, we will apply the results to two issues involving events which are closely related to microreduction and to the mind-body problem: the explanation of mental events by means of physical theory, and the role of

mental events and physical events in causation. I will compare my account of the relation between mental events and physical events to traditional non-identity theories: interactionism, parallelism, and epiphenomenalism.

The most fundamental feature of microreductive connecting principles is the presence of the part/whole relation. When a theory T_2 is microreduced to a theory T_1 , the things recognized by T_2 are identified with structured wholes, whose parts are the things recognized by T_1 . Each T_2 property involved in a connecting principle applies to certain T_2 things, and hence applies to the structured sums¹ of T_1 things with which T_2 things have been identified. Each T_1 property that's microreductively correlated with a T_2 property will also apply to T_1 thing-sums with which T_2 things have been identified. In the connecting principle which links temperature to mean molecular kinetic energy, for example, particular temperatures of objects are correlated with mean-kinetic-energy values of the molecule-sums with which the objects are identified.

In the case of certain connecting principles, we initially might be inclined to say that the higher-level property applies to an entity as a whole, whereas the lower-level properties only applies to component parts of the entity. An example is the envisioned correlation of pain with neural property P. A neural property, it seems, is only instanced by certain parts of an organism, such as the brain or the components of the brain. However, a closer look at the form which the relevant connecting principle takes will reveal that this initial impression is mistaken. The connecting principle will have the form of a universally-quantified biconditional, as in the following schema: " $(x)(Ax \text{ iff } Bx)$." Accordingly, any predicate we substitute for 'A' must apply to the same kinds of entities as does the predicate we substitute for 'B'. So, if we substitute 'is in pain'

for 'A', then the relevant physicalistic predicate we substitute for 'B' must be one which applies to the whole human organism. Thus, if the predicate 'has neural property P' is substituted for 'B', this predicate is understood to apply to the entire organism. To be more specific about property P, suppose it involves the firing of certain neurons in the brain ("nociceptive" neurons). Now when we say that pain is microreductively correlated with the firing of nociceptive neurons, we are referring to two properties of the whole organism: 1) the property expressed by the predicate 'is in pain', and 2) the property expressed by the predicate 'is such that it's nociceptive neurons are firing'.

What, then, is the relation between two simultaneous instantancings of microreductively correlated properties, where the instantancings both involve the same concrete physical entity? I suggest that the term 'constitutes' is an appropriate predicate to express the relation. A gas' molecules' having a particular mean kinetic energy (at time t) constitutes the gas' having a certain temperature (at t). A human organism's having neural property P constitutes the person's (simultaneously) being in pain. An event (a, P, t) constitutes an event (b, Q, t') just in case the following conditions are met:

1. $a=b$
2. $t=t'$
3. P and Q are properties countenanced by two true theories T_1 and T_2 respectively, such that T_2 is microreducible to T_1 via a set of connecting principles $R_{1,2}$.
4. $R_{1,2}$ contains either a universal biconditional simultaneity-law connecting P and Q, or a universal conditional simultaneity-law which asserts that Q occurs whenever P occurs.
5. a is a structured sum of T_1 things (viz., things to which sortal predicates

of T_1 apply); b is a T_2 thing.

The term 'constitutes' is designed to reflect a rather deep-seated materialist intuition about event-pairs of the sort we're considering. This is the view that there is "nothing more" to any concrete mental event than the corresponding physical event. An instance of pain is held to be "nothing more than" a particular instance of property P ; someone's wanting to go to the Bahamas is "nothing over and above" the associated state of his body, etc. A similar intuition was mentioned in chapter II with respect to microreductively correlated properties, along with the other intuitive N-D principles. But in the present sense, the intuition applies to concrete, unrepeatable events, rather than to properties. I think it's fair to say that the intuition has been a strong motivating factor behind many philosophers' adoption of the thesis that mental events are identical to physical events.

But we need not postulate event-identity to account for this intuition, any more than we needed to postulate property-identity to account for the various N-D intuitions. Instead, we can account for it by observing that the "nothing but" intuition regarding individual events is closely connected to the corresponding intuitions concerning event-types (properties). Why are we inclined to say that there's nothing more to Jones' pain at ten o'clock than the occurrence of neural property P in Jones' body at ten o'clock? Clearly, because we believe that in general, there's "nothing more" to pain than the associated physical property. The intuition regarding particular events is derivative from principle (N) regarding properties. And of course, we have already seen in chapter II that (N) is explicated by the supervenience principle.

The constitution relation, as I have characterized

it, is transitive, irreflexive, and asymmetric. All three properties result from the nature of the relation of microreducibility among scientific theories. That relation too is transitive, irreflexive, and asymmetric. Transitivity transforms to the constitution relation in the following way: If (a, P, t) constitutes (b, Q, t') , and (b, Q, t') constitutes (c, R, t'') , then we are guaranteed that there are at least one-way laws connecting P to Q , and connecting Q to R . Hence, we know that there is at least a one-way law connecting P to R . So, (a, P, t) constitutes (c, R, t'') , since all other conditions are automatically met by virtue of the fact that each of these events is involved in a prior constitution relation. The irreflexivity of the constitution relation results from the fact that no "law" of the form " $(x)(Px \text{ iff } Px)$ " or " $(x)(\text{If } Px \text{ then } Px)$ " counts as a microreductive correlation law. (And hence, no scientific theory microreduces to itself.) Thus, (a, P, t) does not constitute (a, P, t) .

The asymmetry of the constitution relation is a direct consequence of the asymmetry of microreducibility. Suppose (a, P, t) constitutes (b, Q, t) . Then according to clause 3 of the above characterization, the properties P and Q belong respectively to two theories T_1 and T_2 , where T_2 is microreducible to T_1 . Now if it were to turn out that (b, Q, t) constitutes (a, P, t) , then T_1 would also be microreducible to T_2 , according to clause 3. But microreducibility is asymmetric. Hence, the constitution relation is also asymmetric.

Some further observations are in order concerning asymmetry. As I indicated above, the term "constitutes" is meant to express the "nothing more than" intuition, as it applies to the relation between mental events and their physical counterparts. Now, the asymmetry of microreducibility does not yet explain why this "nothing but" intuition is also asymmetric--why it proceeds from the

micro-level toward the macro-level, and not vice-versa. But an explanation is ready at hand, in the supervenience' principle. We have already seen that the supervenience principle is at the basis of the "nothing but" intuition. And we also know this: the principle guarantees that the truth of the microreductive connecting principles R of unified science are determined by the final reducing theory T_0 . On the other hand, the principle does not guarantee that the truth of the connecting principles is determined by any higher-level theory. On the contrary, there are possible worlds where higher-level theories have different reduction bases than T_0 . Thus, the source of the asymmetry of the "nothing but" intuition lies in the fact that micro-level theory determines macro-level theory according to the supervenience principle but not vice-versa.

This fact is important in explaining why we seek to microreduce higher-level theories, rather than to "macro-reduce" lower-level ones. We seek microreductions not only because we have been notably successful in achieving them in the past, but also because of the metaphysical primacy of the parts over the whole. This primacy is reflected in the fact that the theory governing the micro-level entities determines the truth of the theories governing complex structures composed of these entities, whereas the latter theories do not determine the truth of the former.

The constitution relation between two events is a very intimate one. In particular, it is evidently a closer and more intimate relation than causation. The Humean conception of causation, however it is to be worked out in detail, asserts roughly that two events are causally related if they are subsumable under a universal law, a law which guarantees that events of the first kind are invariably followed by events of the second kind. One reason for distinguishing constitution from

causation is the temporal factor: two constitutionally-related events are simultaneous, whereas causation is often (though not always) thought to involve a temporal precedence of cause to effect. But more important is the status of the microreductive connecting laws involved in constitution, as opposed to the subsuming laws involved in causation. The connection between properties that is involved in such laws is stronger than mere nomological correlation, because the laws are supervenient upon fundamental laws of physical theory. Their truth is thus determined by the fundamental laws, despite the fact that they are not logical consequences of these laws. They are not merely true in the actual world, or in all worlds where we explicitly specify that they obtain. Rather, they are true in all K-worlds whose fundamental laws are the same as ours.

Of course, the truth of the non-fundamental laws of the various higher-level scientific theories is also determined by more basic theory, if the Unity of Science hypothesis is correct. But this is because the non-fundamental laws are logical consequences of the conjunction of fundamental laws and connecting principles. This logical determination merely reflects the fact that the higher-level laws are derived, rather than primitive, within unified science. It doesn't involve any stronger-than-nomological connection between the properties that are connected in the laws of the higher sciences. Microreductive connecting principles, on the other hand, do count as theoretically primitive within Unified Science, since they don't follow logically from the fundamental laws. Nonetheless, their truth is determined by the fundamental laws; they are supervenient. Their determination is non-logical, and it is thus indicative of a connection that's stronger than mere nomological correlation. The additional strength of the connection between microreductively correlated properties is precisely this:

such properties do not diverge in extension in any possible K-world where fundamental physical theory is true, even though their divergence is not precluded logically. This non-divergence is not guaranteed by mere nomological correlation, and it allows us to conclude that the constitution relation between events is more intimate than the relation of causation.

Let me conclude this section by noting certain important parallels between the constitution relation and Alvin Goldman's "generation" relation for actions.² (My thesis was originally conceived on the idea of describing a relation between physical events and mental events analogous to Goldman's generation relation.) The generation relation is meant to elucidate the "by" relation, as it is expressed in ordinary discourse about human acts. (This is the relation which holds between Jones' act of shooting a gun and his act of killing Smith, in the situation where he kills Smith by shooting a gun.) Generation is transitive, irreflexive, and asymmetric. It involves simultaneous events. It is non-causal. (In the above example, Jones' shooting a gun is the cause of Smith's death; but it is not the cause of Jones' killing Smith.) Finally, a generating act is intuitively more fundamental than the generated act, and the relation between the two is intuitively very intimate. In all these respects, generation resembles constitution.

Section 2

We turn now to the topic of scientific explanation, as it is related to microreduction and the Unity of Science. It has often been held that the microreduction of theory T_2 to theory T_1 should make it possible to explain, by means of T_1 alone, events which were explained by T_2 prior to the reduction. In particular, it has been held that T_1 should be capable of explaining any obser-

ational data which T_2 can explain. The Kemeny-Oppenheim view of reduction, in fact, asserts that the reducibility of T_2 to T_1 amounts to no more than this: T_2 contains terms not in T_1 , and any observational data explainable by T_2 are explainable by T_1 .

So far, we have considered reduction in terms of the Nagel approach, with its requirement that reduction involve a set of connecting principles $R_{1,2}$ such that $T_1 \cup R_{1,2}$ entails T_2 . K-O reduction does not impose this requirement. It seems inadequate on account of this failure, since a K-O reduction of T_2 to T_1 might not be capable of explaining why the theoretical principles of T_2 are true--something which we strongly feel that a reduction of T_2 to T_1 should do. Moreover, even K-O reduction is likely to require a certain number of intertheoretic connecting principles, since the observational data explainable by T_2 will sometimes be characterized in terms of vocabulary that is not part of T_1 . (For example, a reduction of mentalistic psychology to physical theory would presumably require that physical theory be able to explain certain mentalistic data--wants, beliefs, thoughts, etc.--as well as overt behavior.³)

We shall continue to subscribe to Nagel's conception of reduction, since his derivability requirement allows for a deductive-nomological explanation of the theoretical principles of the reduced theory. Still, we are inclined to feel that an adequate account of reduction should indeed satisfy the K-O account. In fact, once we return to Nagel reducibility with its connections between theoretical predicates, it seems that the explainability requirement should be even stronger than the K-O requirement. That is, theory T_1 should be capable of explaining any event which T_2 explains, not just any observational datum which T_2 explains.

To see what is required for this desideratum to be met, we must examine the notion "explainable by a theory," as it applies to events. We will adopt the standard deductive-nomological model of explanation. The D-N model proceeds on the assumption that we have available canonical event-describing sentences, which can be assigned non-problematically to the events we are interested in. Adopting this assumption, we let ' E_i ' represent the canonical sentence describing event e_i . According to the D-N model, an ideal and fully explicit scientific explanation of an event e_i is a valid deductive argument, whose conclusion is the statement E_i (the "explanandum"), and whose premises consist of certain general laws and certain statements describing specific antecedent conditions (these premises jointly being the "explanans").⁴

The simplest notion of "explainability by a theory," in terms of the D-N model, is this: Event e_i is explainable by theory T_1 just in case there is some true antecedent-conditions statement A , such that T_1 and A constitute an explanans for E_i without A alone constituting an explanans for it.⁵ (In accordance with the D-N model, A is understood to be a statement describing an individual event or several individual events. It does not incorporate any additional theoretical laws.)

However, this notion clearly isn't strong enough for present purposes, since it doesn't allow for the explainability by T_1 of events that are described in terms of the vocabulary of the reduced theory T_2 . (Nor does it allow for the explainability by T_1 of observational occurrences which are associated with T_2 by theoretical-observational correspondence principles.) Suppose e_2 is such a "higher-level" event, (e_2 might be a mental event, while T_1 is some physical theory.) In order to be able to say that e_2 is explainable by T_1 , we

must expand the notion of explainability by a theory, to something like this: Event e_1 is explainable by theory T_1 just in case there is some true antecedent-conditions statement A (stated solely in terms of the theoretical and pre-theoretical vocabulary of T_1) and an auxiliary theory T' (which may be vacuous) such that T_1 , T' , and A constitute an explanans for E_1 without T' and A alone constituting an explanans for it.⁵

Now while this expansion of the idea of explainability does indeed allow for the possibility that the higher-level event e_2 is explainable by T_1 , the question arises whether the expansion is acceptable. Is it reasonable to say that e_2 is "explainable by T_1 ," when the construction of an explanans for E_2 requires the explicit mention of an auxiliary theory T' ? On the face of it, it seems not. The reasonable claim appears to be that e_2 is explainable by $T_1 \cup T'$, but not by T_1 alone. If T' were false, our explanation would collapse.

The force of this objection does not lie in the mere observation that a deductive-nomological explanation of e_2 cannot be constructed on the basis of theory T_1 alone. This fact has already been granted, and the proposed wider sense of "explainable by a theory" is explicitly designed to apply to such cases. Rather, the objection draws its strength from the fact that if T' were false, then the explanation of e_2 would collapse. Since the falsity of T' would vitiate the explanation of e_2 , it seems incorrect to assert that e_2 is explainable by theory T_1 alone.

The extent to which this objection can be met will depend upon the nature of T' . The sort of auxiliary theory we are interested in, of course, is a set $R_{1,2}$ of connecting principles that establish a microreduction of T_2 to T_1 , where T_2 (together with suitable antecedent-condition statements) provides a direct D-N explanation of e_2 . In this situation, a D-N explanation of e_2 can

also be obtained on the basis of $T_1 \cup R_{1,2}$, but not on the basis of T_1 alone. So we need to justify the claim that e_2 is explainable by T_1 , in the face of the fact that the explanation based on $T_1 \cup R_{1,2}$ would be vitiated if the connecting principles $R_{1,2}$ were false.

Fortunately, the requisite justification is provided by the supervenience principle. Given this principle, we know that the connecting principles $R_{1,2}$ will not be false in any possible world similar to our world in point of fundamental physical laws and basic indecomposable entities. And this fact seems fully adequate as a reply to the above objection, since the claim that event e_2 is "explainable by T_1 " is made against the background of the final reducing theory of unified science. The explainability of e_2 by T_1 obtains, modulo the truth of the fundamental reducing theory T_0 and the absence of concrete indecomposable entities not countenanced by T_0 . (So, in worlds where Cartesian souls exist but theory T_0 is true, we might deny that mental events are explainable by neurophysiology, even if systematic psychological correlation laws exist, since these laws could presumably differ from one such world to another without any change in physical theory. But this fact does not affect the claim that in our world, mental events are explainable by neurophysiology. This explainability results from the fact that psychophysical correlation laws of our world obtain in all T_0 K-worlds.)

Let me connect these considerations to the constitution relation. Consider the scenario for an explanation of e_2 on the basis of A , T_1 , and $R_{1,2}$. By hypothesis, A and T_1 do not provide a direct D-N explanation of e_2 . However, they do provide a D-N explanation of an event e_1 , which stands in the constitution-relation to e_2 . We may rephrase the claim that e_2 is explainable by T_1 as follows: the constitution relation preserves ex-

plainability. For any two events e_1 and e_2 , if e_1 is explainable by T_1 via a D-N argument, and if e_1 constitutes e_2 , then e_2 is explainable by T_1 . Consider the microreductive correlation of neural property P with pain. Such a correlation would establish that occasions of property P constitute instances of pain. So occurrences of pain would be explicable by neurophysiology, provided that the associated physical events are neurophysiologically explainable.

Section 3

In this section I shall discuss the subject of causation, in light of the constitution relation. The discussion will be somewhat tentative because the notion of causation requires careful philosophical analysis, and I shall not attempt such an analysis here.⁶ I will assume a Humean conception of causation, by which I mean the view that causation involves general laws in an essential way. Beyond this, I shall rely largely on the intuitive content of the notion; hopefully, a theory of causation can ultimately be given which captures this intuitive content.

I will focus on the question of the relation between mental causation and physical causation, where the physical events we are interested in are those which constitute the mental events that are simultaneous with them. But it should be kept in mind that the implications of the discussion are more general: they apply with respect to any two microreductive event "levels"--e.g., to the relation between microphysical causation and chemical causation, or between chemical causation and biological causation, etc.

There are two principles regarding causation which are generally advocated by proponents of the view that psychology is reducible to physical theory--principles

I think we should allow for. First, it is held that the reduction of psychology would establish a complete causal nexus at the level of physical theory. Physical events are always caused by prior physical events; or, if there are causal indeterminacies at the physical level, these indeterminacies are not "filled in," as it were, by mental causes. This principle extends to the causation of people's overt behavior: bodily motions result from prior chains of purely physical causes.⁷ Secondly, it is held that the reduction of psychology brings mental events within the closed physical-level causal nexus. Just how they are brought within this nexus is a question we shall consider below.

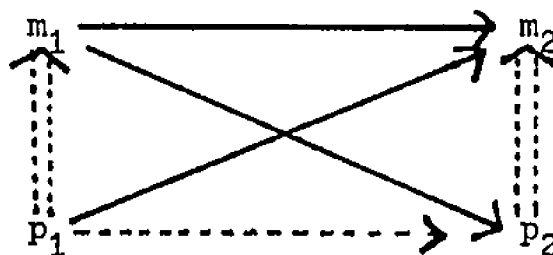
These two principles both seem quite natural, once we hypothesize that psychology is reducible to physical theory. But in addition, there are two further principles about causation which derive directly from common sense. The first is mental/physical interactionism, the view that mental events can sometimes cause physical events in the body, and vice-versa. Examples of mental-physical causation are readily at hand: pains can cause winces, thoughts can cause the heart to pound, feelings can cause trembling. Examples of physical-mental causation are even easier to find: blows that can cause dull aches, flashes of light that cause a person to have a certain after-image, pieces of music that cause a person to have certain feelings or memories, and electrical brain stimulations that cause a person to have a certain thought.⁸

The second common-sense principle is that some mental events cause other mental events. Thinking of Nixon's "dirty tricks" department might cause strong feelings of suspicion toward Nixon; thinking of John Dean's testimony might cause one to believe that John Mitchell is a liar, and so on.

It has often been held that the above four principles regarding causation cannot be simultaneously main-

tained. I shall contend that they can be, and indeed that they are all true. But before describing my own position with regard to causation, let me mention that one advantage of the identity thesis for mental and physical events is that it too allows for the truth of the above four principles. Since mental events are identical with physical events, the postulation of a closed causal nexus at the physical level automatically guarantees the truth of the second, third, and fourth principles. I think it is fair to say that this fact is one reason why the identity theory has been attractive to many philosophers.

The position I advocate can be briefly described as a dual-causation parallelist/interactionist view. I advocate parallelism regarding mental and physical events which are simultaneous, and interactionism regarding mental and physical events that are nonsimultaneous. And I hold that mental causation and physical causation operate together: events with mental causes have simultaneous physical causes, and events with physical causes may have simultaneous mental causes. The position is illustrated by the following diagram. Events p_1 and p_2 are physical events, and m_1 and m_2 are mental events.



We stipulate these facts: 1) p_1 and m_1 are simultaneous, and p_1 constitutes m_1 ; 2) p_2 and m_2 are simultaneous, and p_2 constitutes m_2 ; 3) p_2 and m_2 are later than p_1 and m_1 ; 4) p_1 causes p_2 . Given these stipulations, I make the following claims regarding causation: 1) p_1 causes m_2 ; 2) m_1 causes p_2 ; 3) m_1 causes m_2 . The single arrow represents causation, and the double arrow

represents constitution. Broken arrows represent relations which obtain by stipulation, and solid arrows represent the causal relations which obtain, I claim, when the stipulated relations are met. Dual causation is represented by the two single arrows pointing to ' p_2 ' and the two single arrows pointing to ' m_2 '. Interactionism is represented by the diagonal arrows. Parallelism is represented by the double arrow of constitution, rather than the single arrow of causation, between ' p_1 ' and ' m_1 ' and between ' p_2 ' and ' m_2 '.

As I indicated in the first section, the constitution relation differs from causation. It is more intimate than causation, since it involves the co-extensiveness of the higher-level event-types in all possible K-worlds where T_0 is true. Thus, when I apply the term 'parallelism' to the relation between simultaneous mental and physical events, I should not be taken to mean that the connection of the events is somehow fortuitous or accidental. On the contrary, the connection is stronger than a causal connection, rather than weaker. Still, the term 'parallelism' is being applied in the traditional way: to indicate that the relation between a physical event and its mental counterpart is neither identity nor causation.

Traditionally, parallelism, interactionism and epiphenomenalism have been treated as mutually exclusive views. This is because they were often stated without regard to simultaneity and non-simultaneity, as follows. "Interactionism asserts that physical events cause mental events and mental events cause physical ones; epiphenomenalism asserts that physical events cause mental events but mental events don't cause physical ones; and parallelism asserts that neither sort of event ever causes the other." Stated in this loose form, my view counts simply as interactionism. But the three views are no longer exclusive if we distinguish the case of simul-

taneous events from the nonsimultaneous case. And if we make the distinction, then the parallelist aspect of my position emerges.

My viewpoint also embodies something of an epiphenomenalist element. Epiphenomenalism asserts that the (apparent) causation of physical events by mental events, and also of mental events by prior mental events, is really grounded in physical causation. Now, I deny that mental causation is only apparent. It is perfectly real. However, I do adopt the view that any event which has a mental cause m_i also has a physical cause p_i , such that p_i constitutes m_i . Mental causation is thus grounded in physical causation.

My reasons for adopting parallelism with respect to simultaneous mental and physical events have already been made clear. As for nonsimultaneous events, I advocate an interactionist position for two reasons. First, interactionism is strongly supported by our usual causal idiom, and by our intuitions about causation. Secondly, interactionism is in accordance with the basic Humean model, which asserts that causal relations depend on underlying nomic regularities.

Several sorts of case come into consideration, as regards the applicability of the Humean approach. I will refer to the above diagram in distinguishing them. The different cases involve the nature of the constitution relation between p_1 and m_1 , and between p_2 and m_2 . The relation can be based either on a biconditional correlation law connecting the constitutive properties, or on a one-way conditional law. Thus, we have four cases to consider: 1) The constitution relation between p_1 and m_1 involves a biconditional law, and the relation between p_2 and m_2 involves a biconditional law. 2) The first involves a biconditional law, and the second a conditional law. 3) The first involves a conditional law, the second a biconditional. 4) Both involve conditional laws.

The first two form a single category, and they are nonproblematic from a Humean point of view. Let P_1 , P_2 , M_1 , and M_2 be the constitutive properties of the respective events p_1 , p_2 , m_1 , and m_2 . Since p_1 causes p_2 , we know by the Humean account that there is a lawlike regularity involving the respective constitutive properties, P_1 and P_2 . This guarantees that whenever P_1 is instanced, P_2 is instanced subsequently. Now, cases 1 and 2 involve a biconditional correlation law connecting P_1 with M_1 . So, given the law connecting P_1 to P_2 , whenever M_1 is instanced by a person, P_2 is instanced by his body. A mental/physical covering-law is therefore guaranteed. Also, cases 1 and 2 involve at least a one-way law connecting P_2 to M_2 . Hence, whenever P_1 is instanced by a person's body, M_2 is instanced by the person. A physical/mental covering law is therefore guaranteed, also. (A similar argument also establishes a covering law connecting M_1 and M_2 .)

The second two cases are less straightforward. Both involve a one-way law connecting P_1 to M_1 . The problem is that there is no guarantee, given only a one-way law from P_1 to M_1 and our basic one-way law from P_1 to P_2 , that there will be a law from M_1 to P_2 . Hence, there is also no guarantee of a law from M_1 to M_2 . So we apparently lack assurance that there are covering-laws sufficient to establish mental/physical causation, or mental/mental causation.

We can surmount this problem, however, if we are willing to modify the basic Humean conception along lines suggested by Davidson.⁹ He suggests that although singular causal statements do entail the existence of a law which subsumes the two concrete events, such singular statements do not entail that the statement of the subsuming law will employ terminology that is even remotely similar to the terminology of the singular causal statement itself. This approach conforms well with his

speculation that psychology is heteronomic--i.e., that precise laws to underlie the qualitative generalizations of psychology can only be found at the level of natural science. His approach allows us to speak of mental causation, while looking to physical theory for precise covering laws.

On Davidson's conception of events, these underlying covering laws from physical theory would directly subsume the mental events which are involved in singular causal judgments. He holds that the mental event is also a physical event, and that it can be re-described physically in such a way that the law applies. This course is not open to us, since we denied the identity of mental and physical events. However, an analogous course is open: we simply extend the concept "subsumability under a law," just as we extended the concept "explainable by a theory" in the previous section. That is, we assert that the constitution relation preserves subsumability. This gives us just the results we want, as consideration of the above diagram will show. Since p_1 constitutes m_1 and p_2 constitutes m_2 , m_1 and m_2 are subsumed under the law which subsumes p_1 and p_2 respectively. Thus, we can assert that m_1 causes p_2 and that m_1 causes m_2 .

The principle that constitution preserves subsumability is a reasonable one, I suggest, because of the extremely intimate connection between events that are related by constitution. If e_1 constitutes e_2 , then there is "nothing more" to the occurrence of e_2 than the occurrence of e_1 . The relation between the two constitutive properties is stronger than nomological, as we have seen. In short, all the reasons which spoke in favor of the view that constitution preserves explainability also allow us to conclude that constitution preserves subsumability.

The preservation of subsumability establishes that the interactionist viewpoint I advocate is indeed in

accordance with the Humean picture of causation. A glance at the earlier diagram should make this point clear: m_1 and m_2 are both subsumable under the law which underlies the causal connection between p_1 and p_2 . Specifically, m_1 falls under the constitutive property P_1 of the event p_1 , and m_2 falls under the constitutive property P_2 of the event p_2 . Thus, the causation of m_2 by p_1 , of p_2 by m_1 , and of m_2 by m_1 is supported by the underlying law connecting P_1 to P_2 .

The extension of subsumability which we have adopted does not in itself require that we assert the existence of causal relations in these three cases, of course. But it does open the door to this position, within the context of the Humean conception of causation. And I contend that we should go on to maintain an interactionist view. The reasons, as mentioned earlier, are these: 1) interactionism is strongly supported by our intuitions about causation, and by our usual causal idiom, and 2) the interactionist position I have described satisfies the Humean requirement that causally-related events must be subsumable under a general law.

Section 4

A potential difficulty arises at this point, however. The form of interactionism I am advocating requires dual causation, and the notion of dual causation has come under attack. I am referring to David Lewis' "spurious cause" argument, which was discussed in chapter III. In terms of the above diagram, the argument is this: Consider the event p_2 . If the mental event m_1 is not identical to the physical event p_1 , then m_1 is not a genuine cause of p_2 . Event m_1 occupies a role analogous to the "power on" light, which "does not cause the motor to go, even if it is a lawfully perfect correlate of the electric current that really causes the

motor to go."¹⁰ Similarly, the argument goes, m_1 does not cause p_2 , even though M_1 is a lawfully perfect correlate of P_1 , the property whose instancing really causes p_2 . So mental/physical causation is ruled out by the presence of a closed chain of physical causes.

A similar analogy can be constructed to apply to the case of mental-mental causation. Suppose our motor has a "motor on" light. This light's being on is the analogue of mental event m_2 . Now, the "power on" light does not cause the "motor on" light to come on, any more than it causes the motor to go. The electric current causes the "motor on" light to come on, just as it causes the motor to go. Similarly, Lewis might claim, m_1 does not really cause m_2 . Event p_1 is the genuine cause of both p_2 and m_2 . The "dual causation" view is thus ruled out altogether.

Presumably, Lewis' contention that the "power on" light is a spurious cause involves the following counterfactual considerations. If the light were not on when the electric current were flowing, the motor would still run and the "motor on" light would still come on. But if the current were not flowing when the light was on, the motor would not go, and the "motor on" light would stay off. Thus the current, and not the light, is the real cause of the motor's running and the light's coming on.

The analogous argument in the case of the causation of p_2 by m_1 is this. (We will assume, for simplicity, that P_1 and M_1 are connected by a biconditional correlation law and that P_2 and M_2 are connected by a biconditional correlation law.) If p_1 occurred but m_1 did not, then p_1 would still have caused p_2 and m_2 . This is because the law connecting P_1 to P_2 would still be in effect, as would the law connecting P_2 to M_2 . But if m_1 occurred but p_1 did not, then p_2 would not have

occurred subsequently to m_1 . And so m_2 would not occur either. Thus p_1 , and not m_1 , is the real cause of p_2 and of m_2 . Event m_1 is a spurious cause. The alleged dual-causation view really denies all causal efficacy to mental events.

We can add further details to this argument by referring to a recent discussion by Lewis on the truth conditions for counterfactual statements.¹¹ He takes as primitive the relation of "comparative over-all similarity" among possible worlds. Given this relation, one world is "closer to actuality" than another if the first resembles our actual world more than the second does, taking account of all the respects of similarity and difference and balancing them off one against the other. He then states the truth conditions for counterfactuals as follows. Given any two propositions A and C, the counterfactual "If A were true then C would be true" is true if and only if either 1) there are no possible A-worlds (i.e., there are no possible worlds in which A holds; in this case the counterfactual is vacuously true), or 2) some A-world where C holds is closer to actuality than is any A-world where C does not hold. That is, "a counterfactual is non-vacuously true if and only if it takes less of a departure from actuality to make the consequent true along with the antecedent than it does to make the antecedent true without the consequent."¹²

Let's apply Lewis' analysis to the cases we are considering. We must examine the following two counterfactuals, whose alleged truth is the basis for the claim that even m_1 is a spurious cause.

- 1) If p_1 had occurred but m_1 had not, then p_2 and m_2 would have occurred.
- 2) If m_1 had occurred but p_1 had not, then p_2 and m_2 would not have occurred.

We first will consider statement 1. Suppose there is a

possible world which is exactly like the actual world except that the law connecting properties P_1 and M_1 momentarily fails to obtain. The failure occurs at just the moment when p_1 occurs in this world, and the failure is limited to the one organism whose having of property P_1 is the event p_1 . Such a world would be exactly like ours up to the moment when p_1 occurs, and then would be exactly like ours from the moment after p_1 occurs. At the moment of p_1 , the only difference from our world would be that the organism instantiating P_1 does not instantiate M_1 . Clearly, this world would be the closest to actuality of all worlds in which the antecedent of statement 1 obtains. And since it would resemble the actual world in all respects except the occurrence of m_1 , the consequent of statement 1 would obtain in this world. So according to Lewis' analysis of counterfactuals, if there is such a world then statement 1 is true.

Now consider statement 2. We are now concerned with possible worlds where event m_1 occurred but event p_1 did not. To begin with, we know that in all such worlds the biconditional law connecting P_1 to M_1 does not obtain at the moment of m_1 . Beyond this, how do we go about choosing the relevant possible world which is closest to actuality? That is, how do we get rid of the event p_1 ? We will follow Lewis' own directive: "To get rid of an actual event e with the least over-all departure from actuality, it will normally be best not to diverge at all from the actual course of events until just before the time of e ."¹³ Apparently, the least departure from actuality in Lewis' judgment would involve a very brief failure of the physical laws of the actual world--just enough of a failure to prevent event p_1 from occurring.¹⁴ I.e., the closest world to actuality in which p_1 does not occur will be a world where the course of events up to the moment just prior to p_1 is the same in the actual

world, but where p_1 does not occur because of a very brief failure in the laws of nature.

So it appears that the world closest to actuality in which the antecedent of statement 2 obtains, has the following features: 1) the law connecting P_1 to M_1 fails at the moment of m_1 , and 2) the course of events just prior to m_1 is the same as in the actual world, but p_1 does not occur because of a brief failure of physical laws. Three further features presumably would also apply: 3) the laws of our world resume, at the moment immediately after the occurrence of m_1 , 4) the subsequent course of events is as much like the actual subsequent course of events as possible, given that the laws of our world are again in effect and that p_1 did not occur, and 5) with the exception of the non-occurrence of p_1 , the situation at the time m_1 occurs is the same as it is in our world.

Having specified the characteristics of the world closest to actuality in which the antecedent of statement 2 obtains, we can now ask whether the consequent of 2 obtains in this world. Evidently it does. Physical theory has resumed by the time p_2 is supposed to occur, but p_1 has not occurred. The non-occurrence of p_1 , together with the truth of physical theory, is presumably sufficient to guarantee the non-occurrence of p_2 . And the non-occurrence of p_2 , together with the biconditional law linking P_2 to M_2 , is sufficient to guarantee the non-occurrence of m_2 . Thus, the consequent of statement 2 is true in the world we have described, a world which seems to be the closest to actuality in which the antecedent of 2 obtains. So according to Lewis' analysis of counterfactuals, if there is such a world then statement 2 is true.

Turning now from the development of Lewis' argument to its evaluation, I think we would have to admit that statements 1 and 2 are true, and hence that the

"spurious cause" argument is well-taken, if we were to admit that there are possible worlds of the sort just described. But the supervenience principle now comes to the fore. Because of the supervenience of the law connecting P_1 to M_1 , I contend, there are no possible worlds which meet the descriptions set out above. Let's return to the first putative world we described, which seemed to be closest to actuality of the worlds in which the antecedent of statement 1 obtains. This was a world just like ours, except that the law connecting P_1 to M_1 was momentarily violated. Such a world is precisely the sort which the supervenience principle rules out of existence: a T_0 K-world in which a particular microreductive connecting principle fails to obtain.

Now consider the second putative world we described, which seemed to be the closest world to actuality in which the antecedent of statement 2 obtains. Strictly, this putative world is not a T_0 K-world, since the required breach in physical law would involve a breach in the laws of T_0 . (No purely higher-level nomological failure occurs in any T_0 K-world, because of the supervenience principle.) But I submit that the force of the supervenience principle carries over to this putative world as well. Since event m_1 is "nothing more than" event p_1 , a momentary failure in the laws of T_0 which prevents the occurrence of p_1 will also, ipso facto, prevent the occurrence of m_1 . So the second putative world we described doesn't exist.

How do these observations affect the truth-values of statement 1 and 2? To begin with, it becomes clear that any possible world which satisfies the antecedent of statement 1 or statement 2 will be much further removed from actuality than Lewis apparently supposes. Specifically, it will have to be a world whose final reducing theory differs from T_0 , or which contains differ-

ent kinds of indecomposable entities than those of the actual world. Since the relevant worlds must be so dissimilar from the actual world, the question of the truth-values of statements 1 and 2 becomes hazy to the point of indeterminacy. How would we decide which of such far-removed worlds, so vastly dissimilar from ours, are closest to actuality? And how would we know whether or not p_1 would lead to p_2 and m_2 in these worlds? (Remember that we are discussing possible worlds with different fundamental scientific theories, or different fundamental entities, or both.)

The very insuperability of these problems points to another fact about counterfactual assertions like statements 1 and 2. Clearly, such statements are not intended to refer to counterfactual situations as far removed from actuality as the situations we are now discussing. They are intended to refer to counterfactual situations much closer to home. If the envisioned situations do not occur in any possible world that's even remotely similar to actuality in point of fundamental physical laws and basic indecomposable entities, then we cannot assign any determinate truth-value to the counterfactuals in question. Given the supervenience principle, this is the fate of statements 1 and 2. Because there are no sufficiently-close possible worlds where the counterfactual situations described by their antecedents obtain, their truth value is indeterminate.

The dual causation view I have espoused is therefore immune to the "spurious cause" objection, because of the supervenience principle. This principle stands behind the constitution relation for simultaneous physical and mental events. The resulting intimacy of the events p_1 and m_1 , and of p_2 and m_2 , is what prevents the mental event m_1 from being a spurious cause of subsequent events p_2 and m_2 . Events p_1 and m_1 can each be regarded as genuine causes of both p_2 and m_2 , because the event-

types P_1 and M_1 do not diverge across K-worlds in which fundamental physical theory is true. The common-sense doctrines of mental-mental causation and interactionism are thus preserved.

Finally, let me note that the dual causation position I have advocated becomes a multiple causation thesis, once the various scientific "levels" of microreduction are all taken into account.¹⁵ Events can have simultaneous causes from each of the levels: sociological, psychological, physiological, biological, chemical, and microphysical. Each event in the group of causes will bear the constitution-relation to every higher-level event in the group. None of the causes will be spurious, for the reason just described in the mental/physical case. Furthermore, the intimacy of the relation between these events allows us to speak of any of them as "the" cause of the subsequent event, even though there are several causes, strictly speaking. After all, a higher-level cause is "nothing more than" the lower-level cause which constitutes it.

Section 5

I will now briefly summarize my solution to the mind-body problem, and I will compare my solution to certain traditional answers to the problem. My solution, predicated upon the Unity of Science hypothesis, involves two basic components: 1) an account of the relation between simultaneous physical and mental events, and 2) an account of the relation between non-simultaneous mental and physical events.

With respect to 1, I claim that every concrete mental event is constituted by a simultaneous concrete physical event. This position is akin to traditional parallelism, in two respects: a) the constitution-related events are neither identical nor causally related,

and b) there is a regular concomitance of instances of the first event-type with instances of the second event-type. But there is also an element rather different from traditional parallelism: the strong intimacy of the constitution relation, as reflected in the supervenience principle. If event p_i constitutes event m_i , then there is a supervenient nomological connection between the respective constitutive properties, P_i and M_i . Thus, the fact that instances of P_i are regularly accompanied by those of M_i is not a happy co-incidence, as it seems to be under certain traditional forms of parallelism. There is no need to invoke a doctrine of pre-established harmony to explain the concomitance, as Leibnitz did, or to invoke the Occasionalist doctrine that God constantly intervenes to assure the concomitance, as Malebranche did. So my viewpoint does not suffer the problems of these earlier forms of parallelism.

With respect to 2, I have proposed a dual-causation interactionist account of the relation between mental events and physical events. My account is quite different from the classic interactionism of Descartes, since it involves no dualism of substances. On the contrary, persons are composed entirely of microphysical components, according to the Unity of Science hypothesis. My account also differs from forms of interactionism which hold that there are gaps in the chain of physical causation, and that mental causes fill these gaps. This view too is at odds with the Unity of Science hypothesis. Unified science would establish that there is a closed causal nexus at the level of microphysics, or at least that any lower-level indeterminacies are not "filled in" by higher-level laws.

The supervenience principle (S^*) is at the basis of my account. (S^*) establishes the intimacy of the constitution relation, by guaranteeing that there is

more to the constitution relation between two events p_i and m_i than the mere nomic correlation of the constitutive properties P_i and M_i . The intuitive N-D principles (N), (D), (F), and (E) carry over to events as well. Thus, there is "nothing more" to m_i than p_i , since there is nothing more to the property M_i than the property P_i . The occurrence of event m_i is determined by physical events plus physical laws, since the correlation of P_i and M_i is itself determined by the truth of physical theory. Event m_i is a non-fundamental event, since the law connecting P_i to M_i is non-fundamental. Event m_i is explainable by physical theory by means of a D-N explanation of p_i , because the law connecting P_i to M_i is itself in need of no explanation. Given these consequences, I think it is fair to label my account "materialist," even though I have denied that mental events and physical events are identical. A materialist need not, and should not, subscribe to the psychophysical identity thesis.

¹I use the term 'sum' in the sense of Goodman's calculus of individuals.

²Cf. Goldman, especially chap. 2.

³This point is made by Kim, "Reduction, Correspondence, and Identity," pp. 428-29.

⁴For the classic exposition of the deductive-nomological model, see C. Hempel and P. Oppenheim, "Studies in the Logic of Explanation," PS 15 (1948): 135-75.

⁵This characterization is adopted from the discussion in Kim, "Reduction, Correspondence, and Identity," p. 430.

⁶For a discussion of causation using the event-conception I am using here, see Kim, "Causation, Nomic Subsumption, and the Concept of Event," JP 70 (1973): 217-36.

⁷Human actions, too, are thought to be brought within the causal nexus of physical theory. But the case of actions perhaps falls under the second principle I mention, involving mental events. The concept of action is "mentalistic," since the notion of causation by wants and beliefs seems built into it.

⁸The examples are from J. Shaffer, "The Mind-Body Problem," The Encyclopedia of Philosophy, ed. P. Edwards (New York: Macmillan Publishing Co.), vol. 5, p. 341.

⁹Davidson, "Actions Reasons, and Causes," JP 60 (1963): 685-700; and "Causal Relations," JP 64 (1967): 691-703.

¹⁰Lewis, p. 25.

¹¹Lewis, "Causation," JP 70 (1973): 556-67.

¹²Ibid., p. 560.

¹³Ibid., p. 566

¹⁴Lewis makes clear that in his judgment, failure of a law may well involve less departure from actuality than failure of particular fact. The possible world I have described seems to be the sort he has in mind when he says: "Similarities in matter of particular fact trade off against similarities of law....Comprehensive and exact similarities of particular fact throughout large spacio-temporal regions seem to have special weight. It may be worth a small miracle to prolong or expand a region of perfect match." (Ibid., p. 560.)

¹⁵A multiple causation thesis may not be appropriate, given the remarks I make in the latter part of the appendix.

APPENDIX

Section 1

As an afterword I will briefly discuss two matters closely related to the concerns of this thesis, neither of which received adequate attention above. In the present section I will consider the matter of "sortal essentialism." Sortal essentialism, as I shall use the phrase, is the view that the microreductive thing-identities of unified science express essential properties of the complex T_0 things mentioned in these identities. I.e., such T_0 thing-sums possess, in all possible worlds, the same sortal properties countenanced by higher-level theories which they possess in the actual world. The Bohr-structures which we call hydrogen molecules are hydrogen molecules in all possible worlds, DNA molecules are genes in all possible worlds, sodium chloride is salt in all possible worlds, the physiochemical organisms which we call persons are persons in all possible worlds, etc. The matter of sortal essentialism has direct relevance to the problem of classifying things into "natural kinds," a problem which has come into prominence among philosophers who wish to apply the resources of intensional logic to philosophical issues. (The notion of a natural kind plays a major role in the latter parts of Kripke's "Naming and Necessity," as we saw in chapter III.)

Intuitively, sortal essentialism seems quite plausible. We have a strong feeling that the composition of a thing is central to its status as a certain sort of thing. Accordingly, we are inclined to treat sortal properties from higher-level theories as essential properties of the T_0 thing-sums to which they apply. On careful consideration, however, there seem to be limits to the intuitive appeal of sortal essentialism. The essentialism clearly extends to all T_0 K-worlds. But

when we move beyond these, intuitions begin to get hazy. Suppose, for example, that there is a possible world where theory T_0 obtains, but persons have an immaterial soul as the seat of consciousness. Would we say that the predicate 'is a person' applies, in this world, to physiochemical organisms like ours? It seems not. Rather, a person would be a unity of body and soul. Now consider possible worlds where T_0 does not obtain. Suppose that in one such world, w , T_0 things continue to exist.¹ Suppose further that the T_0 things in w happen to come together temporarily into a particular complex structure--say, the structure we call a "hydrogen molecule" in our world. Would we classify this temporary structure in w as a hydrogen molecule, given that the laws of T_0 are not in force? Probably not.

In short, we remain in conformity with pre-analytic intuition if we take the thesis of sortal essentialism as relativized to T_0 K-worlds. Relativized sortal essentialism is the view that the microreductive thing-identities of unified science are true in all T_0 K-worlds. The latter thesis, of course, is a special case of the supervenience principle, since microreductive thing-identities are among the connecting principles of unified science. So I have provided a justification for relativized sortal essentialism, by means of my above argument for (S*).

It should be emphasized, moreover, that sortal essentialism was not presupposed in my argument for (S*) in chapters IV and V. I did presuppose the supervenience of primary and auxiliary theoretical-observational correspondence principles, on the basis of the contention that micro-macro continuity is a methodological constraint on the formulation of such principles. And I also presupposed a "functionalist" account of cognitive and volitional mental properties, when arguing for (S*) as applied to psychology. But nowhere in my argument did I

assume in advance that higher-level sortal predicates apply, in all T_0 K-worlds, to the same entities to which they apply in the actual world.

This is as it should be, in my opinion. The reason why DNA molecules are genes in all T_0 K-worlds is that they can have the appropriate biological properties in all such worlds--not because of their composition alone. Human beings are persons in all T_0 K-worlds because they can undergo the appropriate sorts of mental properties in all such worlds--not because they have the same physiochemical composition as in our world. (In the latter example, I have assumed that 'is a person' is the primary sortal predicate of psychology.)

These remarks about the status of sortal essentialism do not conflict with remarks made by Kripke in "Naming and Necessity" concerning the status of compositional properties as essential properties of individuals. He suggests in several places that the composition of any given individual thing counts as an essential property of that thing.² This suggestion seems reasonable, and I do not wish to deny it. But I emphasize that the suggestion does not entail that having a particular compositional structure is an essential property of things of a certain kind, or that belonging to a certain thing-kind is an essential property of all entities with a given compositional structure. Kripke's position does not imply, for example, that being a gene is an essential property of DNA molecules, or that being a DNA molecule is an essential property of genes.

In sum, sortal essentialism emerges as a special case of the supervenience principle, and it was not presupposed in my arguments for this principle. I do not think that sortal essentialism should be adopted on a priori grounds. Complex T_0 things should not be assumed in advance to possess the same higher-level sortal properties in all T_0 K-worlds which they possess in the

actual world. Before we adopt sortal essentialism, we must be assured that the relevant complex T_0 things can have the same non-sortal properties, in all T_0 K-worlds, as they do in the actual world.

Section 2

I now wish to reconsider an approach to microreductively correlated properties which deserves more attention than it received above. This approach would treat such properties as contingently identical, given Kripke's account of identity statements. That is, it would treat denoting expressions which refer to higher-level theoretical properties as non-rigid designators, and it would treat higher-level theoretical predicates as "nonrigidly expressing" different properties relative to different possible worlds.

The notion of non-rigid expression of a property by a predicate has not been explicitly put forward elsewhere, so far as I know. The view I have in mind is this: when we apply a predicate to something, which property we thereby attribute to the object depends on which possible world we are describing. For example, the predicate 'has temperature T ' would attribute a particular mean molecular kinetic energy to objects, when the predicate is used relative to the actual world. But it would attribute a different property to objects when used relative to possible worlds in which thermodynamic theory has a different reduction basis than in the actual world.

I think there are reasons to reject this approach as it applies to the psychophysical connecting principles which reduce psychology to natural science. (I will return to this case below.) But in the case of microreductive connecting principles which reduce one natural science to another, the view seems fairly plausible.

Furthermore, it seems to mesh fairly well with the discussion of such connecting principles in the text. Take the case of temperature, for example. At the close of chapter III, I suggested that the difference between temperature and mean molecular kinetic energy consists in the fact that temperature is ascribable, in a possible world, to whatever entities in that world satisfy the constraints imposed by the laws and correspondence principles of thermodynamics. This way of describing temperature would seem to accord quite well with the view that the predicate 'has temperature T' is non-rigid. I.e., we could contend that this predicate expresses different properties relative to different possible worlds; in a given world w , it expresses whatever property in w satisfies the constraints imposed by the laws and correspondence principles associated with the expression 'temperature T'.

The non-rigid view also meshes quite well with my discussion in chapter IV, concerning the criteria for the truth of the connecting principles R in a T_0 K-world w . The criteria were these: 1) the auxiliary correspondence principles of R obtain in world w ; 2) in world w , the models for the theoretical entities countenanced by higher-level theories are constructable on the basis of the T_0 model; and 3) there is no comparably simple set R' of connecting principles in world w which satisfies conditions 1 and 2. It seems that we can view these criteria as "picking out" the properties which are non-rigidly expressed, in world w , by the various higher-level theoretical predicates in R . (If we adopt the non-rigid view with respect to all microreductive connecting principles within unified science, then the only rigid predicates will be those of theory T_0 . Higher-level theoretical predicates will all non-rigidly express the microphysical properties with which they are ultimately correlated.)

The following question quite naturally arises at this point: what is the substantive difference between this non-rigid view, and the view that higher-level theoretical predicates express properties rigidly? Whether or not there is a substantive difference, and what specifically the difference comes to, will depend largely on the nature of properties themselves. I have not directly addressed the question of the nature of properties in this thesis, and I shall not do so now. I do not have a theory of properties to propose.

Let's now consider the non-rigid view as applied to psychophysical connecting principles. First, take the case of theoretical-phenomenal connecting principles. Can we plausibly say that the predicate 'is in pain' non-rigidly designates some microphysical property of sentient organisms? I suggest not. What would be the criterion for determining which lower-level property this predicate expresses in other possible worlds? The appropriate criterion is this: the property non-rigidly expressed must be one which an entity undergoes if and only if it has pain. But clearly this won't do: we end up treating pain as a separate property in the very act of attempting to state criteria for the non-rigid use of the predicate 'is in pain'. Pain is a separate property from its neural correlate.

Secondly, consider the case of cognitive and volitional predicates in psychology. It seems implausible to treat them as non-rigid predicates, for the following reason. There are many underlying physical properties of a person which can accompany the same cognitive or volitional property, as we saw in chapter V. So, on the non-rigid view, we would have to say that we are constantly assigning different properties to people when we apply a given intentional predicate to them. This position is highly counterintuitive. If we say on two dif-

ferent occasions that Jones wants to go to Nassau, or that Jones believes that Nixon is a liar, it seems clear that we mean to ascribe the same property to him on both occasions, even though the underlying microphysical property of Jones might be quite different on the two occasions. Thus, the lack of a neat one-one mapping from the intentional to the physical vitiates the non-rigid approach to intentional predicates.

If we do adopt the non-rigid view with respect to microreductively correlated properties in natural science, while rejecting it in the case of psychophysical property-correlations, then the microreduction of psychology becomes somewhat unique. But the truth of the supervenience principle is not affected and this uniqueness thus does not affect the applicability of the intuitive principles (D), (E), (F), and (N) to the case of the reduction of psychology. Moreover, there may be advantages in adopting the non-rigid property-identity thesis as applied to connecting principles within natural science. For example, we do not need to expand the dual causation thesis of chapter VI into a multiple causation thesis. Also, we achieve a degree of ontological economy, since we countenance fewer distinct property-kinds within natural science.

In conclusion, let me emphasize the way the supervenience principle is related to the non-rigid property-identity thesis. The non-rigid identity thesis explicitly treats higher-level properties as non-rigid designators of those lower-level properties which meet the constraints imposed by higher-level laws, correspondence principles, and models. If property identity is understood in this way, then the assertion that microreductively correlated properties are identical does not entail the supervenience principle. The possibility is left open that there are T_0 K-worlds in which higher-

level theoretical predicates do not non-rigidly express the same T_0 properties which they express in the actual world. So if we were to adopt the non-rigid identity thesis without adopting the supervenience principle, the resulting contingent property identities would be fundamental laws of unified science, whose truth was not determined by T_0 . In order to avoid this result we would have to adopt the supervenience principle, over and above the non-rigid identity thesis.

¹Even this may be a problematic assumption. Will entities which are T_0 things in our world possess their T_0 sortal properties in worlds where the laws of T_0 do not obtain? Perhaps not.

²Kripke, p. 314 and footnote 57, p. 351.

REFERENCES

- Armstrong, D.M. A Materialist Theory of the Mind. New York: Humanities Press, 1968.
- Block, Norman, and Fodor, Jerry. "What Psychological States Are Not." PR 81 (1972): 159-81.
- Brandt, Richard, and Kim, Jaegwon. "The Logic of the Identity Theory." JP 64 (1967): 515-37.
- _____. "Wants As Explanations of Action." JP 60 (1963): 425-35.
- Carnap, Rudolph. "Empiricism, Semantics, and Ontology." Revue Internationale de Philosophie 11 (1950): 20-40.
- _____. "The Methodological Character of Theoretical Concepts." Minnesota Studies in the Philosophy of Science, 1. Ed. H. Feigl and M. Scriven. Minneapolis: University of Minnesota Press, 1956.
- Donagan, Alan. The Later Philosophy of Collingwood. Oxford: Oxford University Press, 1962.
- Davidson, Donald. "Actions, Reasons, and Causes." JP 60 (1963): 685-700.
- _____. "Causal Relations." JP 64 (1967): 691-703.
- _____. "Mental Events." Experience and Theory. Ed. L. Foster and J.W. Swanson. Amherst: University of Massachusetts Press, 1970.
- Dennett, Daniel. Content and Consciousness. New York: Humanities Press, 1969.
- _____. "Intentional Systems." JP 68 (1971): 87-106.
- Feigl, Herbert. The "Mental" and the "Physical": The Essay and a Postscript. Minneapolis: University of Minnesota Press, 1967.
- Fodor, Jerry. "Explanations in Psychology." Philosophy in America. Ed. M. Black. Ithaca: Cornell University Press, 1965.
- _____. Psychological Explanation: An Introduction to the Philosophy of Psychology. New York: Random House, 1968.

Goldman, Alvin. "The Compatibility of Mechanism and Purpose." PR 78 (1968): 468-82.

_____. "The Individuation of Action." JP 68 (1971): 761-73.

_____. A Theory of Human Action. Englewood Cliffs: Prentice-Hall, 1970.

Hempel, Carl. "On the 'Standard Conception' of Scientific Theories." Minnesota Studies, 4. Ed. M. Radner and S. Winokur (1970).

_____. "Reduction: Ontological and Linguistic Facets." Philosophy, Science, and Method: Essays in Honor of Ernest Nagel. Ed. S. Morgenbesser, P. Suppes, and M. White. New York: St. Martin's Press, 1969.

_____, and Oppenheim, Paul. "Studies in the Logic of Explanation." PS 15 (1948): 135-75.

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

Kim, Jaegwon. "Causation, Nomic Subsumption, and the Concept of Event." JP 70 (1973): 217-36.

_____. "Phenomenal Properties, Psychophysical Laws, and the Identity Theory." The Monist 56 (1972): 177-92.

_____. "On the Psycho-Physical Identity Theory." APQ 3 (1966): 227-35.

_____. "Reduction, Correspondence, and Identity." The Monist 52 (1968): 424-38.

Kripke, Saul. "Naming and Necessity." Semantics of Natural Language. Ed. D. Davidson and G. Harmon. New York: Humanities Press, 1972.

Lewis, David. "An Argument for the Identity Theory." JP 63 (1966): 17-25.

_____. "Causation." JP 70 (1973): 556-67.

Maxwell, Grover. "The Ontological Status of Theoretical Entities." Minnesota Studies, 3. Ed. H. Feigl and G. Maxwell (1962).

Meehl, Paul, and Sellars, Wilfred. "The Concept of Emergence." Minnesota Studies, 1.

Nagel, Ernest. "The Meaning of Reduction in the Natural Sciences." Philosophy of Science. Ed. A. Danto and S. Morgenbesser. New York: Meridian Books, 1960.

_____. The Structure of Science. New York: Harcourt, Brace, and World, 1961.

Oppenheim, Paul, and Putnam, Hilary. "Unity of Science as a Working Hypothesis." Minnesota Studies, 2.

Popper, Karl. "The Aim of Science." Ratio 1 (1957): 24-35.

_____. Conjectures and Refutations. New York: Basic Books, 1962.

Putnam, Hilary. "The Mental Life of Some Machines." Intentionality, Minds, and Perception. Ed. H. Castaneda. Detroit: Wayne State University Press, 1966.

_____. "Minds and Machines." Dimensions of Mind. Ed. S. Hook. New York: NYU Press, 1960.

_____. "Psychological Predicates." Art, Mind, and Religion. Ed. W. Capitan and D. Merrill. Pittsburg: University of Pittsburg Press, 1967.

_____. "On Properties." Essays in Honor of Carl Hempel. Ed. N. Rescher. New York: Humanities Press, 1970.

_____. "Robots: Machines of Artificially Created Life?" JP 61 (1964): 668-91.

Quine, W.V. "Natural Kinds." In his Ontological Relativity and Other Essays. New York: Columbia University Press, 1969.

_____. Word and Object. Cambridge, Mass.: MIT Press, 1960.

Rudner, Richard. "An Introduction to Simplicity." PS 23 (1961): 109-115.

Schaffer, Jerome. "The Mind-Body Problem." The Encyclopedia of Philosophy. Ed. P. Edwards. New York: Macmillan, 1967.

Shaffner, Kenneth. "Approaches to Reduction." PS 34 (1967): 137-47.

Sklar, Lawrence. "Types of Inter-theoretic Reduction." BJPS 18 (1967): 109-24.

Smart, J.J.C. Philosophy and Scientific Realism. New York: Humanities Press, 1963.

_____. "Sensations and Brain Processes." PR 68 (1959): 141-56.

Spector, Marshall. "Models and Theories." BJPS 15 (1965): 121-42.

Suppe, Fredrick. "On Partial Interpretation." JP 68 (1971): 57-76.

Tarski, Alfred. "The Concept of Truth in Formalized Languages." In his Logic, Semantics, Metamathematics. Trans. J.H. Woodger. Oxford: Oxford University Press, 1956.