We might therefore prefer to say that the T-terms keep the meanings they received at their first introduction. They should still be defined using the original version of T even after it has been superseded by revised versions.

This position will work only if we permit the T-terms to name components of the nearest near-realization of T, even if it is not a realization of T itself. For after T has been corrected, no matter how slightly, we will believe that the original version of T is unrealized. We will want the T-terms to name components of the unique realization (if any) of the corrected version of T. They can do so without change of meaning if a realization of the corrected version is also a near-realization of the original version.

According to this position, we may be unable to discover the meanings of theoretical terms at a given time just by looking into the minds of the most competent language-users at that time. We will need to look at the past episodes of theory-proposing in which those terms were first introduced into their language. The working physicist is the expert on electrons; but the historian of physics knows more than he about the meaning of 'electron', and hence about which things could truly have been called electrons if the facts had been different. If we were ignorant of history, we could all be ignorant or mistaken about the meanings of words in common use among us. This situation is surprising, but it has precedent: a parallel doctrine about proper names has recently been defended.[13] To know what 'Moses' means among us it is not enough to look into our minds; you must look at the man who stands at the beginning of the causal chain leading to our use of the word 'Moses'.

I do not wish to decide between these alternatives. Either seems defensible at some cost. I hope the truth lies in between, but I do not know what an intermediate position would look like.

DAVID LEWIS

Princeton University

EXPLANATION AND REDUCTION *

ON what has come to be a standard account of the matter, to *explain* some phenomenon is to derive a sentence describing it from other sentences at least one of which states a general law. Thus a deductive argument whose premises include the law for reflection in a plane mirror and a description of certain

[13] See David Kaplan, "Quantifying In," *Synthese*, XIX (1968): 178–214.

* This is a revised version of a paper read at the Pacific Division meeting of the American Philosophical Association, March, 1969. John Earman, Karel Lambert, and John Vickers have supplied helpful comments.

attendant conditions provides an explanation of the fact that light from a small source at the focus of a paraboloidal mirror is reflected in a beam parallel to the axis of the paraboloid. Similarly, to explain a law or theory is to derive it in turn from other laws and theories. Galileo's laws for freely falling bodies are explained, for example, by subsuming them under Newton's laws of motion and gravitation.

Many who hold to this account claim also that to derive a law or theory is to *reduce* the former to the latter. As the most prominent advocate of this claim, Ernest Nagel, puts it:

> . . . a reduction is effected when the experimental laws of the secondary science [i.e., the theory to be reduced] are shown to be the logical consequences of the primary science [i.e., the theory to which the reduction is to be made].[1]

Which is to say, of course, that reductions and explanations have the same logical structure, and both must satisfy the same formal conditions. Galileo's laws as derived from Newton's reduce to them, as eventually (on the received view) do Newton's to Einstein's. We can call the concept involved that of D- (for "derivation") reduction.

Now, in addition to this concept, there is a different, nonformal, connection between the notions of explanation and reduction. This connection stems from two principles which have been widely held, for the most part implicitly, to determine the adequacy of explanations of a certain type. The first, roughly put, requires that the properties of wholes be explained in terms of the properties of their parts. It is often referred to as the "principle of micro-reduction." The second principle, again to put it roughly, requires that the properties of these parts differ from those of the wholes they are invoked to explain. I will call it the "principle of property-reduction." In a way it serves primarily as an amendment to the first principle. For example, suppose that, in accordance with the principle of micro-reduction, we explained the transparency of water in terms of the molecules and eventually the atoms that compose it. Intuitively, the principle of property-reduction suggests, to say that water is transparent because its components are themselves transparent would be to postpone, rather than to provide, an adequate explanation of the phenomenon.[2] It is micro-reductive "explanations" of

[1] *The Structure of Science* (New York: Harcourt, Brace & World, 1961), p. 356; parenthetical page references to Nagel are to this book.

[2] The example is borrowed from P. Oppenheim and H. Putnam, "Unity of Science as a Working Hypothesis," in Feigl, Scriven, and Maxwell, eds., *Minnesota Studies in the Philosophy of Science* (Minneapolis: University of Minnesota Press, 1955), vol. II. As they put it, unless micro-reduction is restricted in this

this intuitively trivial kind that the principle of property-reduction is meant to exclude. We can call the concept involved that of P- (for "partition-and-property") reduction.

A case taken from the history of physics that seems to exemplify both of these concepts is the reduction of classical thermodynamics to statistical mechanics, supplemented by the kinetic theory of matter, which was carried out largely by Ludwig Boltzmann and James Clerk Maxwell in the latter part of the nineteenth century. At the price of vast oversimplification, one aspect of this reduction can be sketched as follows: once a concept peculiar to thermodynamics—temperature—has been in some sense correlated or identified with that of the mean kinetic energy of aggregates of gas molecules and certain idealizing statistical assumptions have been made, the laws of classical thermodynamics, e.g., Boyle's, can be derived from the fundamental postulates of statistical mechanics combined with the kinetic theory of matter. As a consequence of the derivation we could also say that, for example, the experimental law that the pressure of a fixed mass of gas at constant pressure is inversely proportional to its volume is explained by mechanics and kinetic theory. Clearly this is a case of D-reduction.

But it is a case of P-reduction as well. In the first place, certain macroscopic properties of objects—once again, temperature—are reduced to and explained in terms of the component parts—molecules—of those objects. The reduction of thermodynamics to statistical mechanics is thus an instance of micro-reduction. In the second place, these parts, as characterized by the micro-theory that introduces them, lack those macroscopic properties of objects which they are invoked to explain. In the example at hand, temperature—the concept of which does not appear among the fundamental assumptions of statistical mechanics or kinetic theory—is explained in terms of, among other things, the positions and momenta of molecules.

It should be added that, whereas many contemporary philosophers of science tend to emphasize the connection between D-reduction and explanation, physicists interested in thermodynamics stress the importance of P-reduction. P. T. Landsberg, for one, writes:

way, "it lacks any clear empirical significance." Let me add that their account of micro-reduction differs importantly from mine; in particular, they do not connect up micro-reduction and the notion of explanation in the same way. It should also be added that, in certain reaches of contemporary physics, there are cases of apparent macro-reductive explanation, where the behavior of parts is to be explained in terms of the wholes they compose. My purpose is not to deny this, but to point out certain features of micro-reduction taken as a principle of explanation.

> Since the nature of the thermodynamic variables, and their number, can vary within wide limits, the basic theoretical framework of thermodynamics must be kept very general. This has the advantage of giving the theory a wide range of application; but this is balanced by the drawback that *thermodynamic reasoning is in general unsuitable for giving insight into the details of physical processes.*
>
> This last observation, together with the remark that thermodynamics leaves macroscopic variables on one side, leads to the conclusion that a thermodynamic theory is necessarily incomplete. For any system to which thermodynamics can be applied, *a deeper-going theory should exist which yields insight into the detailed physical processes involved.*[3]

Very much the same sort of point is made by P. W. Bridgman:

> There can be no doubt I think, that the average physicist is made a little uncomfortable by thermodynamics. . . . He finds more congenial the approach of statistical mechanics, with its analysis reaching into the details of those microscopic processes which in their larger aggregates constitute the subject matter of thermodynamics. He feels, rightly or wrongly, that by the methods of statistical mechanics and the kinetic theory he has achieved the deeper insight.[4]

For both Landsberg and Bridgman the explanatory force of statistical mechanics seems to stem more from its P-reductive than from its D-reductive character. The former is the source of the "deeper insight" into the "detailed physical processes." Along the same lines, many physicists make a more general distinction between *explanatory* and *phenomenological* theories, once again taking statistical mechanics and thermodynamics as instances of each type.

I

Whatever the emphasis, it might be supposed on the basis of our paradigm (and a host of others apparently like it) that our two concepts of reduction/explanation are generally congruent. Cases of P-reduction are cases of D-reduction.[5] There are reasons, however,

[3] *Thermodynamics* (New York: Interscience, 1961), pp. 3–4 (my italics).

[4] *The Nature of Thermodynamics* (New York: Dover, 1956), pp. ix–x.

[5] Moritz Schlick provides a good example of how the two concepts are run together: "Explanation means the discovery of like in unlike—of identity in difference. And inasmuch as explanation reduces different species of natural phenomena to the same domain, these different species are included as special cases in the latter. Hence we may say, that explanation is the inclusion of the special in the general. Thus, heat and sound, for example, are both explained in so far as they are regarded as special cases of the motion of the smallest particles." *Philosophy of Nature* (New York: Philosophical Library, 1949), p. 18. Also G. Schlesinger's statement of the principle of micro-reduction: "whenever it is desired to unify the behaviour of physical aggregates and the behaviour of their constituent elements, construct those theories in which the former is

for thinking that the two concepts are in fact incompatible and, eventually, that trying to incorporate them into a single account of explanation results in confusion.

We can begin with our paradigm. If thermodynamics is to be derived from mechanics there must be some way of linking up the terms, like 'temperature', which do not occur in mechanics with others which do. This is what Nagel calls the "condition of connectability," and, he insists, every D-reduction must satisfy it. The point is one of elementary logic: a deductive argument is valid (with minor exceptions) only if there is some connection between the terms occurring in the premises and in the conclusion. Only if we add to the assumptions of statistical mechanics plus the kinetic theory the further assumption that temperature is correlated or identified with mean kinetic energy can the D-reduction of classical thermodynamics be carried out. Moreover, it should be clear that *any* P-reduction will require that an assumption of this kind be made if it is to be a D-reduction also.

The possible correlations or identifications appear to be of two broad and basic types. To use a distinction which, indirectly, I will try to undermine, they might be either "logical" in nature or "factual." The supposition that they are logical can, in turn, be taken in at least two different ways. One is to the effect that the meaning of 'temperature' as established in classical thermodynamics is synonymous with 'mean kinetic energy'. The meaning of the one expression can be extracted from that of the other via an analysis. In just this form the suggestion must be rejected. As used in classical thermodynamics 'temperature' is simply not synonymous with 'mean kinetic energy', nor could the ostensible reduction have been carried out by a philosopher.

The other way of taking the supposition that the connection is "logical" is that 'temperature' is identified with (or replaced by) some apparently corresponding expression in statistical mechanics, i.e., is given an explication in terms of statistical mechanics and kinetic energy such that in the sentences comprising thermodynamics 'temperature' is systematically replaceable by 'mean kinetic energy'. What we have done on this suggestion is akin to providing a paraphrase (in Quine's sense) of 'temperature' in terms of 'mean kinetic energy' relative to a set of statistical mechanical assumptions

derived from the latter. In other words, 'the properties of physical systems should be explained in terms of the properties of their parts and not vice-versa' or 'a physical system should be atomized and its properties micro-reduced, that is, shown to *follow* from the behaviour of its micro-parts'." *Method in the Physical Sciences* (New York: Humanities Press, 1963), p. 45 (my italics).

and of concepts shared by mechanics and thermodynamics. In this very limited sense, 'temperature' and 'mean kinetic energy' might be said to be "synonymous."

But this suggestion, although it is plausible on a number of counts, is incompatible with D-reduction as Nagel understands it. Two points especially must be noted. The first is Nagel's contention that

> . . . expressions belonging to a science possess meanings that are fixed by its own procedures of explication. In particular, expressions distinctive of a given science (such as the word 'temperature' as employed in the science of heat) are intelligible in terms of the rules or habits of usage of that branch of inquiry; and when those expressions are used in that branch of study, they must be understood in the senses associated with them in that branch, whether or not the science has been reduced to some other discipline (352).

The suggestion that, in the limited sense indicated, 'temperature' and 'mean kinetic energy' are synonymous is, therefore, tantamount to *redefining* 'temperature'. The second point is that if 'temperature' is redefined in this way, then although something syntactically similar to Boyle's law can be derived from statistical mechanics it is not the classical law, for the meaning of one of the terms used to state it has changed. As Nagel puts it,

> . . . if thermodynamics is to be reduced to mechanics it is temperature in the sense of the term in the classical science of heat which must be asserted to be proportional to the mean kinetic energy of gas molecules (357).

Taken together, these two points add up to what might be called a "condition of meaning invariance." [6] Such a condition is essential to the concept of D-reduction. For D-reduction requires, short of the fallacy of equivocation, that meaning remain invariant in the course of particular reductions. Were this not the case, then the claim that classical thermodynamics has been reduced could not be maintained, which is to say that D-reduction depends on 'temperature' having its *classical* meaning.

The supposition that the connection is "factual" can also be taken in at least two different ways. Nagel's own way of taking it is that the nonoverlapping concepts in the reduced and reducing theories are connected by means of certain "bridge-laws," empirical hypotheses

[6] This notion of meaning invariance derives from P. K. Feyerabend. See his "Explanation, Predictions, Theories," in B. Baumrin, ed., *Philosophy of Science: The Delaware Seminar* (New York: John Wiley, 1963), vol. II.

> . . . asserting that the occurrence of the state of affairs signaled by a certain theoretical expression 'B' in the primary science is a sufficient (or necessary and sufficient) condition for the state of affairs designated by 'A' (354).

When a "bridge-law" suitably correlating temperature and mean kinetic energy is added to statistical mechanics, then the laws constituting classical thermodynamics can be derived. But to which theory will such "bridge-laws" belong? Clearly, they must belong to the reducing theory (the primary science); otherwise the reduction will not be to the supposed reducing theory but to some third theory which contains it. But if the "bridge-law" is to be part of the reducing theory, then this theory must contain the terms and predicates needed for its formulation. In which case, however, the principles embodying the concept of P-reduction have been violated.

The other supposition along "factual" lines is that the connection in question is not correlation via "bridge-laws" but rather ("contingent") identification. This suggestion has much to recommend it. In my view, it also has difficulties. One of these is the unanalyzed notion of identification on which it trades. In the case at hand, for example, the identification of properties—temperature and mean kinetic energy—seems to be at stake. Yet what is it to "identify" *properties* with each other? I raise the question mainly because I have seen no satisfactory answer to it. A second difficulty has to do with the closely related facts (a) that the statistical mechanical analogues of various thermodynamical properties have additional (extensional) attributes, and (b) that the statistical mechanical analogues have a wider range of application or extension than the classically characterized thermodynamical concepts. The statistical mechanical analogue of the classical concept of entropy, to pick but one example, goes well beyond it in that it has a meaning in nonequilibrium situations, as contrasted with the classical situation in which nonequilibrium states are, on a number of assumptions, converted to equilibrium states in terms of which empirical temperature is defined. It is implausible to suggest, as some have, that as a result of our "identification" we "discover" that the thermodynamical properties have additional attributes or a broader extension. And yet, "contingent identification" would at least seem to require that the properties identified have the same attributes and extension. More generally, since, on the principle of property-reduction, the parts *must have different properties* than the wholes whose behavior they explain, almost any sort of identification would seem to violate these requirements.

What in fact seems to be the case, at least in the history of physics, is that theories have replaced, rather than D-reduced, one another, although the replacing theories have tended to be of a P-reductive kind. In the course of such replacement, central concepts—like "temperature"—have been redefined. And this fact leads us to think that a D-reduction has been carried out. But if, to repeat Nagel's point, the concept of temperature has been redefined, then it is not classical thermodynamics that has been D-reduced but something resembling it in certain crucial ways.

II

Not only is it itself a mistake to assume that our two concepts of reduction/explanation are congruent, but this mistake leads to others. The one I want to discuss here concerns the "doctrine of emergence." In advancing what is possibly the stoutest defense of this doctrine,[7] C. D. Broad uses several different arguments, all of which seem to come to this: macroscopic qualities of objects cannot be reduced to and explained in terms of other properties of their parts, because this would entail that given the parts and their properties other macroscopic qualities could be predicted in advance of their having been observed, and this is impossible. Broad asserts, for example, that we could not "possibly have formed the concept of such a colour as blue or such a shade as sky-blue unless we had perceived instances of it." As a result, even the archangel could not deduce from his knowledge of the microscopic structure of a given object that the object would appear blue or sky-blue as it does to human beings. If the existence of the so-called "secondary qualities," he continues,

> . . . or the fact of their appearance depends on the microscopic movements and arrangements of material particles which do not have these qualities themselves, then the laws of this dependence are of the emergent type (*ibid.*, 71/2).

As I have summarized his argument, it has at least three different aspects. One concerns Broad's premise, setting out conditions of concept formation, which is nowhere supported by him. Presumably it stems from Hume's Principle that "ideas," e.g., of colors, are not (logically) possible without one's already having had the appropriate "impressions" from which the "ideas" are first derived. The immediate difficulty is that Hume himself provides a crucial counterexample. We can, he says, form the idea of a particular shade of blue, " 'tho' it had never been conveyed to (us) by (our)

[7] In *The Mind and Its Place in Nature* (London: Routledge & Kegan Paul, 1925).

senses." There is no point, moreover, in maintaining that *no* previous visual experience of colors is necessary to imagine or have the concept of a particular color. For Broad's argument, on the present reading, depends on the far stronger premise that there is a one-one correspondence between "impressions" and derived "ideas."

In the second place, Broad's argument seems to involve a confusion between what, e.g., color an object has and what color it appears to have to particular observers. Obviously the archangel could not predict the latter on the basis of the microscopic structure of the objects alone.

But of greater importance, the argument also turns on running together our two concepts of reduction. Broad wants to maintain that there are certain things that mechanical-atomistic theories cannot explain. His reason is that macroscopic qualities of objects cannot be D-reduced in that they could not have been predicted. But this affects the possibility of giving a P-reductive explanation (and hence the adequacy of mechanical-atomistic theories) only if we assume that to explain is to derive, that meaning must remain invariant, and that explanation and prediction are symmetrical. If we do not make these assumptions, then Broad's claim does not stand. If we do make them, some sort of "emergentist" doctrine seems inevitable.[8] In fact, I have suggested that no P-reductions are D-reductions. If we maintain in addition that P-reductions are explanatory, then Broad's point is effectively undermined. That certain phenomena cannot be simultaneously P-reduced and D-reduced is true. But this does not mean that an upper bound can, in an a priori way, be placed on the explanatory force of P-reductive theories.

III

I want to say something more, finally, about the concept of P-reduction. It embodies, we suggested, two different principles, one of micro-, the other of property-reduction. These principles have not been very much analyzed, and, in particular, the concept of explanation they involve has not been elaborated at any length. We must limit ourselves here to saying something about one of the ways in which they are related.

[8] See F. C. S. Northrop, *Science and First Principles* (Cambridge: University Press, 1931), p. 251: "Certainly a theory which does not account for the presence of (secondary qualities) is incomplete or false. Yet if nothing exists but the atoms and fields of physical theory, as they are defined by the physicist, whites and blues and pains and pleasures would not be present. For no possible combination of colourless atoms or fields can ever give rise to a blue. By no process of logical gymnastics can one deduce a perceived colour or sound or pain, or any other secondary or tertiary quality, from (these) entities."

The physicist Werner Heisenberg formulates the principle of property-reduction as follows:

> It is impossible to explain . . . qualities of matter except by tracing these back to the behavior of entities which themselves no longer possess these qualities. If atoms are really to explain the origin of color and smell of visible material bodies, then they cannot possess properties like color and smell. . . . Atomic theory consistently denies the atom any such perceptible properties.[9]

The first use Heisenberg wants to make of this principle is in connection with the origin and development of atomism *chez* Democritus, especially as regards his distinction between "primary" and "secondary" qualities. It is not merely that atoms, as they are usually conceived, do not have any of the "secondary" qualities, but that to play a role in the explanation of these qualities they must not have them. This is Democritus's basic insight. If the phenomena cannot be "reduced" in this way, then they have not been adequately explained. It is of course a further question why just those qualities traditionally labeled "secondary" should from the outset have been taken as candidates for reduction and explanation in terms of the qualities traditionally labeled "primary."

The second use Heisenberg wants to make of the principle of property-reduction is more radical. Atomic theory, as consistently developed, denies to atoms *any and all* macroscopic properties. That is to say, it denies to them extension, position, and motion as well as color, etc. This point extends in two directions. In one direction, Heisenberg wants to argue in behalf of quantum mechanics that, insofar as elementary particles are characterized by no macroscopic properties (but, for example, by probability functions unpicturable in principle), it represents a consistent, and complete, extension of the atomic explanation of the behavior of physical objects. No macroscopic property is (theoretically) left unreduced and, consequently, unexplained. Those who find quantum mechanics unsatisfactory because on its account the elementary "particles" postulated by the theory are stripped of their "physical objecthood" simply have not come to terms with the implications of a basic principle of explanation. The other direction in which the point extends is toward the necessity of accounting for the fact that Democritus and his successors down through at least the end of the nineteenth century failed to extend the principle of property-reduction consistently. Heisenberg's explanation is simple:

[9] "Gedanken der antiken Naturphilosophie in der modernen Physik," *Die Antike,* XIII (1937): 119. Quoted by N. R. Hanson, *The Concept of the Positron* (Cambridge: University Press, 1963), p. 50.

> Democritus has left to the atom the quality of 'being', of extension in space, of shape and motion. He has left those qualities because it would have been difficult to speak about the atom at all if such qualities had been taken away from it.[10]

There is no doubt that much of the debate concerning the philosophical interpretation of quantum theory turns on this point: whether or not a theory can intelligibly lead to the "dematerialization" of the matter concept. Heisenberg's opponents sometimes claim that stripping the elementary "particles" of all macroscopic properties robs us of the possibility of coming to any genuine understanding of what is going on in nature.

But there is another issue involved here. Recall the statement that Newton gives of the third of his "Rules of Reasoning in Philosophy" in the *Principia:*

> The extension, hardness, impenetrability, mobility, and inertia of the whole result from the extension, hardness, impenetrability, mobility, and inertia of the parts; and hence we conclude the least particles of all bodies to be also extended and hard and impenetrable and movable and endowed with their proper inertia. And this is the foundation of all philosophy.[11]

This passage can be surveyed from a variety of points of view. From one, it sets out a kind of rule of property-induction which (ostensibly) allows an empiricist to be realistic about elementary particles. From another, it expresses Newton's retention, as a good corpuscularian, of the principle of micro-reduction. From yet a third point of view, the passage marks a rather sharp break with Heisenberg's formulation of the principle of property-reduction. Newton quite straightforwardly exempts certain properties from the scope of the principle. On a plausible interpretation, his grounds are that these properties are additive: very crudely, that there exists for them an empirical operation (such as placing rods end to end) formally similar to addition in arithmetic. My reasons for adhering to this interpretation, in order of ascending importance, are three.

First, Newton says that the extension, etc., of the whole results from the extension, etc., of the parts. The most natural interpretation of 'result from', I should think, is that it means something like 'is the sum of'.

Second, it is undeniable that Newton's interest in the qualities he lists is that they lend themselves to mathematization and mea-

[10] *Physics and Philosophy* (New York: Harper & Row, 1958), pp. 69–70.
[11] From the Motte translation (1729), revised, etc., by Florian Cajori (Berkeley: University of California Press, 1934), p. 399.

surement in a way in which those properties commonly listed as "secondary" do not.

Third, and this is for us the crucial point, the additivity of their properties gives us some grip on what it means to call, e.g., the elementary corpuscles "parts" of wholes. Objects are "sums" of particles in just this sense, that their length, mass, etc., is the sum of the lengths, masses, etc., of the particles that compose them.

There is, then, this connection between the principles of micro- and property-reduction. If we are to understand micro-reduction on anything like a traditional part/whole relation, then the principle of property-reduction must be restricted in the way indicated. On the other hand, if property-reduction is left unrestricted, other ways in which parts go together to form wholes will have to be specified. This, in turn, suggests that the principle of micro-reduction as I have formulated it will need a great deal of examination and eventual qualification.

GORDON G. BRITTAN, JR.

University of California, Irvine

BOOK REVIEWS

Many-valued Logic. NICHOLAS RESCHER. New York: McGraw-Hill, 1969. 359 p. $8.95.

This is a valuable survey of many-valued logic. The emphasis is on semantical and philosophical questions; technical matters are expounded fully and clearly, but the reader is referred to the literature for complicated proofs. In fact, the 96-page bibliography is one of the strong points of the book. It begins with a chronological listing extending to 1965; then there is an author listing, followed by a topically classified register (e.g., quantification theory in many-valued logic; many-valued logic and the logical paradoxes, etc.).

The bulk of the text is in chapter II, consisting of thirty sections devoted to an assortment of topics. Just a few of the subjects covered are: the many-valued logics of Lukasiewicz, Bochvar, Kleene, and Post; truth-functional completeness; semantical interpretations; varieties of negation; the laws of contradiction and excluded middle; axiomatizability; consistency and completeness; probability logic; modal structures; quantification. (Only nine pages deal with quantifiers. The rest of the book is concerned exclusively with propositional logic.)

The author gives a tolerant, balanced description of the various