An Inductive Logic of Theories

1. Theoretical Inference

In their contributions to this volume, Professors Hempel and Feigl have both discussed the "layer-cake" model of scientific theories from the point of view of the meaning and interpretation of theoretical language. The problem of interpretation has also concerned Professor Maxwell, and he has in addition mentioned the problem of confirmation of theories, but only to assert that this is wholly independent of and irrelevant to the problem of meaning. It is this last assertion of independence that I wish to question in this paper. I believe that the problem of theoretical inference or confirmation is insoluble in terms of a deductive model of theories, and indeed in terms of any analysis which makes a radical epistemological and semantic distinction between theoretical and observation predicates. It follows that an analysis of the meaning of theoretical terms such as that given by Maxwell will be insufficient to support an adequate confirmation theory of theories, and that the questions of what a theory *means* and how it comes to be *confirmed* are not after all independent. The argument of this paper will be of reductio ad absurdum form: first developing expressions for a deductive theoretical system and its models in the usual way, and then showing how they fail to allow for inductive inferences of a kind generally regarded as justifiable.

The problems of induction and confirmation have been given a new lease of life in recent discussions, but little has yet been done toward a logic of inductive inference in relation to scientific theories. Nevertheless, among those who still hold that there is such a thing as inductive logic, I suppose it would be generally agreed that there ought, in principle, to be some way of explicating inductive inferences to theories, and to the further observational consequences of theories. Putnam[1] has given a striking

[1] H. Putnam, "'Degree of Confirmation' and Inductive Logic," in P. A. Schilpp, ed., *The Philosophy of Rudolph Carnap* (LaSalle, Ill.: Open Court, 1963), pp. 761–793.

example of the kind of inference involved here, which would surely be regarded by most scientists as inductively justifiable. In an attempt to show that any confirmation theory of Carnap's type is bound to be inadequate for theoretical inference, he considers inferences of the kind that were made when the first atomic bomb explosion was predicted and subsequently tested with success. Here there was a body of evidence drawn from physics and chemistry which supported the nuclear theory, and from this theory were derived in great detail, and with very great confidence, predictions about what would occur in an experimental situation so far wholly unrealized in any previous experience, namely the slamming together of two subcritical masses of uranium to produce an explosion of given magnitude. Schematically the situation is this: Given total initial evidence e_1, we use this to support a theory t in which we have sufficient confidence to deduce future predictions e_2, namely the results of the atom bomb test. Putnam's own argument to the effect that no confirmation theory can explicate this inference seems to me insufficiently general, but there is a perfectly general argument which shows that no probabilistic confirmation theory of any type yet developed will allow us to infer with higher than prior confirmation from e_1 to e_2 *merely in virtue of the fact that both are deductive consequences of some theory*, where "theory" is understood in the deductive-model sense. This theorem has, in fact, already been mentioned by Carnap,[2] in his discussion of some points in Hempel's paper "Studies in the Logic of Confirmation."[3]

In this paper, Hempel expresses a principle which seems to underlie the kind of inference Putnam exemplifies. The principle is what Hempel calls the consequence condition for confirmation, namely "If some evidence e confirms a hypothesis t, then e confirms every L-consequence of t" (C_1). If this condition were satisfied it would cover Putnam's example, because e_2 would then be confirmed by e_1 which confirms t. Consider the case in which $t \rightarrow e_1 \cdot e_2$. This is the situation we have when we say that t explains the data e_1 and predicts e_2. If we now add a further condition C_2, "If t L-implies e_1, then e_1 confirms t," which Hempel calls the converse consequence condition, it is easy to show that C_1 and C_2 taken together are

[2] R. Carnap, *Logical Foundations of Probability* (Chicago: University of Chicago Press, 1950), especially pp. 471–472.

[3] C. G. Hempel, "Studies in the Logic of Confirmation," *Mind*, 54 (1945), 1–26, 97–121. Hempel has modified some of these points in his addendum to a reprint of the paper in C. G. Hempel, *Aspects of Scientific Explanation* (New York: Free Press, 1965), p. 50.

counterintuitive. Suppose $t \equiv e_1 \cdot e_2$. Then certainly $t \rightarrow e_1 \cdot e_2$, and using C_1 and C_2 we conclude that e_1 confirms e_2. But this is absurd, because e_2 may be any statement whatever. If t is produced just by arbitrarily conjoining any other statement e_2 to e_1, we should certainly not want a confirmation theory in general to allow e_1 to confirm e_2. Some further conditions must be imposed upon admissible e_1 and e_2, and either C_1 or C_2 or both must be modified.

This argument is very general, and concerns any confirmation theory which satisfies C_1 and C_2. In particular, any probabilistic c-theory satisfies C_2, but Carnap[4] has shown that in his probabilistic c-theory, confirmation does not satisfy C_1. Therefore he avoids the counterintuitive inference, but apparently at the cost of being unable to explicate inferences of Putnam's type. In general, in a probabilistic c-theory we cannot have

(1) $\quad c(e_2, e_1) > c_0(e_2)$

unless there is some probabilistic dependence between e_1 and e_2, and such dependence is not guaranteed by the fact that e_1 and e_2 are both implied by some t, because in the absence of further conditions upon t, this is trivially the case for every pair of statements e_1, e_2. I have suggested elsewhere[5] a method of rescuing Putnam-type inferences within a probabilistic c-theory, which involves specifying a relation between e_1 and e_2 that ensures the satisfaction of (1) only in case e_1 is intuitively relevant to the confirmation of e_2. The relation which constitutes such relevance between e_1 and e_2, I suggest, is a relation of *analogy*. I cannot discuss in detail here how this would be defined, but I will describe the conception briefly by taking an example which I develop more fully in the other paper.

2. The Analogical Character of Theories

Consider the inference made by means of Newton's theory of gravitation (t) from some initial data (e_1) which we will suppose comprise Kepler's laws, to the prediction (e_2) that a falling body will fall in the neighborhood of the earth with a certain acceleration. We suppose that the historical situation is that we accept Kepler's laws, but do not yet know the acceleration relation for falling bodies. This is not too far from the actual case, because, as has been pointed out many times, Galileo's law is contra-

[4] Carnap, *Logical Foundations of Probability*, p. 471.
[5] M. B. Hesse, "Consilience of Inductions," in I. Lakatos, ed., *The Problem of Inductive Logic* (Amsterdam: North-Holland, 1968), pp. 232–257.

dicted by Newton's theory, and we do in fact have more confidence in the relation e_2 predicted by that theory than in Galileo's earlier empirical approximation to it. This confidence seems to come from the support given to Newton's theory by the other data entailed and explained by it, which support we take to be in some way passed on to the prediction. But in general, as I have shown, this is not a justified inference. Unless more is said about the relation between e_1 and e_2 we shall not have the desirable increase of $c(e_2, e_1)$ compared with $c_0(e_2)$. When the condition (1) is satisfied I shall say that the corresponding inductive inference is justifiable.

Let us express Kepler's laws and the predicted acceleration relation very schematically as

(2) $\quad e_1\colon (x)(F(x) \cdot G(x)\colon \supset P(x) \cdot Q(x))$
$\quad\quad e_2\colon (x)(G(x) \cdot H(x)\colon \supset Q(x) \cdot R(x))$

Here a relation of "analogy" has been assumed between e_1 and e_2 in the following sense: The predicates F, G represent properties of the planets asserted by Kepler's laws to have certain motions which we denote by the conjunction of predicates P, Q. (It would of course be necessary to use metric predicates if this were a realistic reconstruction, but we simplify drastically for purposes of exposition by considering only monadic predicates.) Expressed in similar fashion is e_2, for the bodies referred to in e_2 share some properties with those referred to in e_1, but not all. All these bodies are solid, massive, opaque, and so on; but bodies near the earth differ from planets in size, shape, chemical composition, and so on. In the same way, Kepler's laws can be thought of as describing motions which are in some respects the same as and in some respects different from the motions described in e_2: All the orbits are ideally conic sections, but they are traversed at different speeds, and about different foci.

My suggestion is that the confidence we have in the prediction of e_2 is due to the relation of analogy between e_1 and e_2 which is constituted by the repetition of predicates G and Q in the expressions of e_1 and e_2. We regard e_2 as confirmed by e_1 because the bodies described by e_2 are sufficiently similar in some respects to those described by e_1 to justify the inference that their behavior will also be similar. Explication of this inference therefore requires a c-theory which will yield the inequality (1) in particular when e_1 and e_2 are as specified in (2), and in comparable cases. There is an example of such a c-theory (the η-theory) in Carnap and Stegmüller,[6] where Carnap's earlier theory is developed in such a way that this condition is satisfied. Explication of analogical inference of this kind has therefore been shown to be possible within a probabilistic confirmation theory, although the η-theory itself is not wholly satisfactory in other respects.

It is not necessary here to go into the question of how this relation of analogy would be defined in general. It is sufficient to say that the best tactics do not seem to be to try antecedently to define what we would mean in all cases by the *analogy* between two objects or two systems in virtue of the inferences we regard as intuitively justified. Such a procedure is liable to strain intuition too far. Rather, we should take some simple cases, such as the one just discussed, in which it is clear that the inference would generally be regarded as justified, find what conditions these cases would impose upon a c-theory, and then use the weakest possible c-theory satisfying these conditions to define the analogy relation in cases which are too complex to give much help to intuition. That is to say, whenever the inequality $c(e_2, e_1) > c_0(e_2)$ is satisfied for observation statements in such a c-theory, we shall say that there is necessarily a relation of analogy in the sense intended between e_1 and e_2. This relation of course has to be consistent with an assertion of analogy or its absence in cases where we do intuitively recognize the existence of analogy and the justifiability of analogical argument, or their absence.

The question that immediately arises, however, concerns the place of the theory t in this explication. We have not needed to mention t either in the expressions of e_1 or e_2, or in the statement of inequality of c-functions. Is t then wholly redundant? Further inspection of (2) reveals that this is not the case, for it is implied in (2) that there is a theory, indeed more than one theory, which has the traditional relation to the data and prediction of entailing their conjunction. In particular, if t is

(3) $(x)(F(x) \supset P(x)) \cdot (G(x) \supset Q(x)) \cdot (H(x) \supset R(x))$

then $t \rightarrow e_1 \cdot e_2$. Furthermore, t has the desirable characteristic of "saying more than" $e_1 \cdot e_2$, since it is not the case that $e_1 \cdot e_2 \rightarrow t$. What t does in effect is to pick out from e_1 and e_2 the predicates G, Q which are in common between them, and to assert that the essential correlation in both cases is that bodies which are G are also Q, and that the properties of the two domains of phenomena which are different are due to two other laws,

[6] R. Carnap and W. Stegmüller, *Induktive Logik und Wahrscheinlichkeit* (Vienna: Springer, 1959), Appendix B.

one of which (relating F and P) applies only to the e_1-domain, and the other (relating H and R) only to the e_2-domain.

It might be noted at this point that when we speak of Newton's theory as "explaining" Kepler's laws and the law of falling bodies, we do not as a rule claim that Newton's theory includes a deductive explanation of *all* the differences between planets and falling bodies, that is, we do not include in the explanatory theory *laws* $(x)(F(x) \supset P(x))$ and $(x)(H(x) \supset R(x))$ which mention all the properties the bodies do *not* share. Newton's theory contains laws explaining why some features of the motions of planets are different from those of falling bodies, but not all such features are mentioned in the theory; for example, their different chemical compositions do not appear in the antecedent of any law of Newton's theory, nor do their different initial velocities appear in the consequent of any such law. Kepler's laws, on the other hand, if they are considered as data to be explained, do not imply any distinction between properties which are in the later light of Newton's theory "relevant" or "irrelevant" to the search for explanation. Kepler's own understanding of planets in the assertion "All planets move in ellipses" certainly included for example the assumption that planets have magnetic properties, which he considered specially relevant to explanation of their motions. This aspect of the explanandum is, however, not mentioned in Newton's "explanation" even in the initial conditions for Kepler's laws. It follows that the expression (3) for t which was used above in deference to the requirement that the explanandum be *deducible* from the explanans plus initial conditions is too strong to reproduce the real situation, in which an "explanation" is *not* required to entail the explanandum as *that was originally formulated,* but is already the result of assumptions of relevance and irrelevance which are rarely made explicit in deductivist accounts of theories. Before the deductive account can be made to work at all, irrelevant features must in fact be dropped from the explanandum as unexplainable by that theory (although they may of course be explained by another theory). The analogical account of theories which has just been suggested has the merit of making these assumptions of irrelevance explicit from the beginning. According to this account, we should regard the theory t, not as in expression (3), but rather as $(x)(G(x) \supset Q(x))$, together with the statements of initial condition which differentiate the e_1-domain as an application of t from the e_2-domain.

The pattern of theoretical inference we have been studying now takes on a different aspect. We are no longer concerned with a dubious inductive inference from e_1 up to t and down to e_2, but with a direct analogical inference from e_1 to e_2. And t does not provide the upper layer of the cake, but rather, as it were, extracts the jam from e_1 and e_2, that is to say it reveals in these laws the relevant analogies in virtue of which we pass from one to the other inductively.

3. The Function of Models

In the light of these confirmation conditions for theoretical inference, let us investigate the adequacy of a typically deductivist construal of theories and their interpretive models. In particular, I want to consider how the use of models for theories provides examples of inference in which we need stronger logical relations between e_1 and e_2 than can be included in the usual deductivist scheme.

Consider an expression of a theoretical system for a domain of entities, in which all the constant theoretical terms ($T_1, T_2, \ldots$) whose "meanings" are problematic have been replaced by the variables $\tau_1, \tau_2, \ldots$:

(4) $(x)(y) \ldots (x,y, \ldots \epsilon S)\phi(\tau_1,\tau_2, \ldots O_1,O_2, \ldots)$

This is a representation of the theoretical calculus, together with a set of observation predicates $O_1, O_2, \ldots$, so that at this stage it is only a *partially* interpreted system, the τ's remaining uninterpreted. Now there will be a model of this system (let us call it the Q-model) which is represented by replacing the τ's again by the problematic theoretical terms $T_1, T_2, \ldots$, although it is rather difficult to see that this is in the ordinary sense an *interpretation*, because the problem of the meaning of theoretical terms arises precisely from the fact that we do not know what constant predicates the T's are, and so do not know what domain of entities and predicates satisfies this model. However, *if* we knew this, then in the usual logician's sense the Q-model would be a model of the system (4). In what follows "model" will be used of *linguistic* entities (systems of laws, theories, etc.) *not* of the sets of entities and predicates which satisfy these systems.

It seems to me that what the physicist normally means by a model for a theory is not the Q-model, but rather a system of laws satisfied by a set of entities and predicates different from the set of entities and predicates which were to be explained when he set up his theory. When he refers to a set of Newtonian particles as a model for gas theory, this is a set of entities different from the gases whose behavior he is endeavoring to explain

by the gas theory. When the crystallographer builds a structure of colored balls and steel rods on the laboratory bench, this is a set of entities different from the organic molecules he is attempting to construct a theory for. So it is necessary to talk in terms of *two* domains of entities: S will now be, not a universal domain, but what I shall call the *domain of entities of the Q-model*, and I shall denote by S* the *domain of entities of the P-model*, where the P-model is a model in the physicist's sense just indicated.

Let us elaborate expression (4) in order to take account of the relation of the Q-model to the P-model. We shall suppose that there are two sets of observation predicates, $O_1, O_2, \ldots$, and $O_1', O_2', \ldots$, where the first set enters into laws *known to be true* in S and S*, and the second into laws known to be true only in S*, and not yet examined in S. We shall suppose the Q- and P-models to have an *analogical relation* in the following sense: They share the O- and O′- predicates, and the Q-model involves also predicates $M_1, \ldots$ not applicable to S*, and the P-model involves predicates $N_1, \ldots$, not applicable to S (the "negative analogy" of the two models). We assume that the laws known to be true of S* and S are respectively

(5) e_1^*: $(x)(y) \ldots [(x,y, \ldots \epsilon S^*)\psi(N_1,O_1, \ldots)]$
(6) e_1: $(x)(y) \ldots [(x,y, \ldots \epsilon S)\psi(M_1,O_1, \ldots)]$

and the laws known to be true of S* and unexamined in S are respectively

(7) e_2^*: $(x)(y) \ldots [(x,y, \ldots \epsilon S^*)\psi'(N_1,O_1', \ldots)]$
(8) e_2: $(x)(y) \ldots [(x,y, \ldots \epsilon S)\psi'(M_1,O_1', \ldots)]$

The P- and Q-models can be expressed as

(9) t^*: $(x)(y) \ldots [(x,y, \ldots \epsilon S^*)\phi(P_1,N_1,O_1,O_1', \ldots)]$
(10) t: $(x)(y) \ldots [(x,y, \ldots \epsilon S)\phi(T_1,M_1,O_1,O_1', \ldots)]$

where the P-model is a *true interpretation* of the partially interpreted expression corresponding to (4), and we have

(11) $(x)(y) \ldots [\phi(P_1,N_1,O_1,O_1', \ldots) \rightarrow \psi(N_1,O_1, \ldots) \cdot \psi'(N_1,O_1', \ldots)]$

All predicates in these expressions are constants.

The problem is to show that there is a justified analogical inference, in the type of confirmation theory we have discussed, from $e_1^* \cdot e_1 \cdot e_2^*$ to e_2. But before considering this, there is an obscurity about the notion of observability which ought to be cleared up at this point. There is not only the very important distinction between observable *predicates* and observable *entities* but there is also a distinction concerning observability of predicates *in different domains*. It may very well be the case that in the P-model we have predicates (the P's), such as "mass," "velocity," "radius," which are observable in the domain of macroscopic physical objects, but not in that of microscopic objects. Where the P-model is used as a model for the theory about gases, the corresponding predicates in the Q-model (the T's) are not observable in any domain, and are only given as it were courtesy titles when we refer to them as "mass," "velocity," etc. If we are to use these adjectives at all in relation to the Q-model, we must at least recognize that they name properties unobservable in the domain S, though observable in S*. The domains of both Q-model and P-model, however, contain the O-predicates, and these are observable in both domains. (The O′-predicates may not be observable in S*, as we shall see presently.) In S* the O's will include the average pressure of a cloud of macroscopic particles hitting a surface, such as hailstones striking a wall horizontally. In S they will include the pressure of the gas measured by manometers. "Pressure" is the same predicate, observable in both domains.

The difference between the deductivist's construal and my own emerges when we consider the T-predicates. He generally wishes to say that these are entirely undetermined except by the Q-model, whose status as an interpretation is, as we have seen, highly problematic. Although there may be another model, the P-model, of the same calculus, there is for the deductivist no relation between the Q- and P-models other than that they are models of the same calculus and share the same O-predicates. Therefore, in this view, if we do use the words "mass," "velocity," etc., in relation to the T's, this is an equivocal use when compared with their use in relation to the P's. We cannot know whether for God the T's are the same predicates as the P's or not. Talk of the T's therefore seems to me at best to define a *class* of models of the partially interpreted system (4), and it is not clear that replacing the variable τ's by constant T's has added anything to the content of (4), because we do not know what these putative constants are and have in principle no means of finding out (unless of course the T's later become *observable*, but in the deductivist's view this could not be a *general* solution to the problem of the interpretation of theoretical terms, because not all such terms will become observable—if they did *his* problem would dissolve).

If we consider the deductivist construal in the light of the considerations about theoretical inference in the preceding sections, it is not clear

that it has any resources for explicating this kind of inference. Even if we waive the difficulties about interpretation for the moment, and suppose that the Q-model is an interpreted theory in the usual sense, we have shown above that inferences from one subset of observable consequences of this model to another subset are not in general inductively justifiable. In particular there would be no justifiable prediction from a set of experimental laws about gases, say Boyle's and Charles's laws, by way of the kinetic theory to other laws about gases if the kinetic theory is understood as a Q-model, that is, if there is no more than an equivocal sense in which we can speak of its being "about" masses and velocities of molecules. Suppose, however, we now bring in the P-model. In the gas example this is a model of Newtonian particles whose laws of motion are known to be true, or at least accepted for purposes of exploitation in the theory of gases. We suppose the models expressed as in (9) and (10), and that ϕ entails the known experimental laws shared by both models as in (11). It should incidentally be noticed that since $\psi'(N_1, O_1', \ldots)$ is entailed by ϕ, it is known to be true in S* even if it has not been directly examined, or even if it is for all practical purposes unobservable in S*. For example, it is unlikely that the analogue of Boyle's law in Newtonian particle mechanics has ever been *observed* to be true; nevertheless it is believed because Newton's laws are believed in that domain.

In a confirmation theory of the type described in section 2 we may have a relation of analogy between $e_1{}^*$ and e_1 as expressed in (5) and (6). (Compare the e_1, e_2 of section 2 (2), where $N_1, \ldots$, stand for F, P; $M_1, \ldots$, stand for H, R; and $O_1, \ldots$, stand for G, Q.) It is very important to be clear at this point that the relation of analogy here spoken of is *not only* the formal analogy in virtue of the fact that both models are models of the same calculus, but includes what I have elsewhere[7] called a *material analogy* in virtue of the sharing of the O-predicates. Similarly, we may have material and formal analogies between $e_2{}^*$ and e_2 expressed in (7) and (8), in virtue of the sharing of O′-predicates and their relations expressed in ψ'. Since $e_2{}^*$ is true, and if the negative analogy between S and S* is not too strong, there will then be a justifiable analogical inference to e_2, which is strengthened by the truth of both $e_1{}^*$ and e_1. Moreover, the same argument yields a justifiable inference to

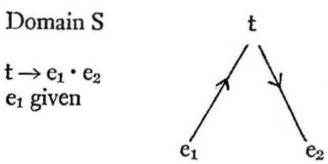
[7] M. B. Hesse, *Models and Analogies in Science* (London: Sheed, 1963; Notre Dame, Ind.: University of Notre Dame Press, 1966).

(12) $(x)(y)\ \ldots\ [(x,y,\ \ldots\ \epsilon S)\phi(P_1,M_1,O_1,O_1',\ \ldots)]$

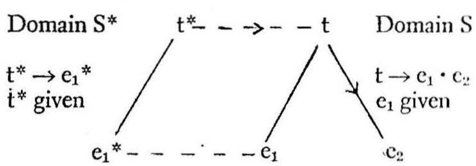
where the P-predicates are observable in S* but unobservable in S.

Comparing (12) with (10), we see that the analogical inference leads to the suggestion that the T-predicates should be *identified* with the P-predicates, rather than being regarded as problematic constants whose reference is in principle unknown. Although their referents in S are unobservable, their "meaning" is derived from the observables of S*, that is, they mean the same as they do in the P-model, and satisfy the same laws. With any construal of the T's not involving some identification of this kind, it is not clear that there can be justifiable analogical inference to T-statements in S. (We shall return to this point in the next section.) With the identification, however, the inference to predictions can be made even stronger, for as was remarked above, it is not necessary that the O′-predicates should be observable in S*, only that the P's should be. In such a case it is clear that the inference to e_2 depends essentially upon knowing the truth of t^* empirically, and making an analogical inference to the probable truth of t and hence e_2, and that this depends upon the identification of the P- and T-predicates in S.

Diagrams may help to elucidate the structure of these inferences and to relate them to what has been said in section 1 about the status of the theory in predictive inference. In the diagrams arrows on the lines indicate alleged justifiable inferences and dotted lines relations of analogy. In the case of the deductive construal, we have seen that these are illusory, and that there is no justifiable inference from e_1 to e_2 unless there is an analogical relation between these laws independent of t. If there is no such relation, however, we may be able to make the inference if we can find a P-model as represented in the right-hand diagram. Here broken lines indicate analogical relations, and the inductive inference depends on the analogy between the laws e_1, e_1^* in the two domains, and the assertion of the truth of t^*. Then there is a justifiable inference to t, and hence to e_2, since $t \rightarrow e_2$. Truth is, as it were, fed into the *theory* of S from the P-model,



Deductive construal



Analogical construal

and so passed on to e_2, whereas in the deductive construal there is no justifiable inference through the theory of S because this theory acquires no probable truth from any source other than its own entailments. Theories cannot be pulled up by their own bootstraps, but only by support from external models.

It will immediately be objected to this account that there are many examples for which analogical inferences of the kind described not only would be unjustified in the light of further evidence, but would never be regarded as justifiable even before further evidence is collected. Analogical arguments are notoriously weak and liable to failure and must generally be treated with extreme caution. This of course is true, but it must be borne in mind that the examples above have presupposed the principle of total evidence. *If* all the evidence we have is summarized as in t^* and e_1, then the inferences may be intuitively reasonable. But if we have other evidence to the effect, for example, that S and S* differ from each other in many further characteristics, or if we know of other domains in which the inference to ψ' breaks down, then such information may well weaken the inferences to the point of disconfirmation. I suspect that when apparent counterexamples to these inferences are produced, they will be found to involve one or other of these types of additional evidence. In principle an adequate confirmation theory must be capable of dealing with such complexities, and must explicate the weakness as well as the strength of analogical arguments.

4. Identification of Theoretical Predicates

In the absence of a detailed and adequate confirmation theory it is not possible to be precise in reply to objections of the kind just mentioned. But it is perhaps permissible to speculate a little further upon the characteristics which an adequate c-theory might exhibit. In particular, it may be possible to suggest some compromise between the position outlined here and the deductive account, though still within the framework of a confirmation theory. There are two ways in which my position has been opposed too sharply to deductivism and should now be modified.

First, it may be doubted whether we wish to *identify* a theoretical predicate such as "mass" of a molecule or electron with "mass" of a macroscopic particle. On the other hand I think it has been correct to say that "mass" cannot be simply *equivocal* without destroying the possibility of theoretical inference. What we need to reconstruct is a notion of analogical meaning of the word "mass" in the two domains, where "analogical" is used as a middle term between "univocal" and "equivocal," as in some Thomist philosophy. "Mass" is not always used "in the same sense" when predicated of different systems, but it is not on the other hand a pun when it is used of positrons, neutrons, quasars, and the like.[8] The problem of determining how far the meaning of "mass" can be extended analogically and the problem of deciding what analogical inferences are justifiable are closely related. Predicates can be stretched just as far as analogical argument remains justifiable, and conversely. How far this is would have to be decided by looking at the whole complex of evidence in all domains in which the predicate is applied. If, for example, the difference in domain is only one of scale, we shall probably be quite satisfied simply to identify the P- and T-predicates, but when the difference of scale is accompanied by other differences as radical as those between, say, the macroscopic and the nuclear domains, we may become increasingly unwilling to allow any analogical extension of meaning from one to the other.

The second respect in which my account might be modified in the direction of deductivism is in regard to the attributes of the S- and S*-domains which are allowed to weigh in analogical inference. It is convenient here to refer to a discussion by Sellars[9] in which he argues that to identify the theoretical predicates with the P-predicates from an antecedent observation language is to fall into the "myth of the given," and to misrepresent the *novelty* which may be introduced by using P-models in connection with theories. To the construal of T-predicates favored by Nagel[10] Sellars objects that it makes the T's new but not meaningful, and to my construal in my *Models and Analogies in Science* he objects that it makes them meaningful but not new.

If I have understood Sellars correctly, his proposal for the resolution of

[8] I have discussed analogy, context meaning, and related questions in my "Aristotle's Logic of Analogy," *Philosophical Quarterly*, 15 (1965), 328–340; "The Explanatory Function of Metaphor," in Y. Bar-Hillel, ed., *Logic, Methodology and Philosophy of Science* (Amsterdam: North-Holland, 1965), pp. 249–259; "A Self-Correcting Observation Language," in B. Van Rootselaar and J. F. Staal, eds., *Logic, Methodology and Philosophy of Science*, vol. III (Amsterdam: North-Holland, 1968), pp. 297–309; and "Theory and Observation," to appear in the forthcoming vol. 4 of the University of Pittsburgh series in the philosophy of science edited by R. G. Colodny.

[9] W. Sellars, "Scientific Realism or Irenic Instrumentalism," in R. S. Cohen and M. W. Wartofsky, eds., *Boston Studies in the Philosophy of Science*, vol. III (New York: Humanities, 1965), pp. 171–204.

[10] E. Nagel, *The Structure of Science* (New York: Harcourt, Brace and World, 1961).

this dilemma is as follows. Not all analogy is analogy of *particulars* in virtue of their sharing identical attributes, as has been assumed in the discussion of analogy in section 2 above. Attributes themselves may be similar or analogous, that is, first-order predicates may themselves be predicated by second-order predicates, and may be analogous in virtue of sharing such second- or even higher-order predicates. Sellars gives as examples the second-order predicate "perceptible," which applies to first-order predicates, and the second-order predicate "transitive," which applies to first-order relations such as "before," "to the left of." Either such second-order attributes can be *mentioned*, as in " 'transitivity' is true of *before*," or they can be *shown*, as by exhibiting the transitivity postulate satisfied by "before." The function of a P-model is to introduce second- and higher-order predicates which are shared with the Q-model, and thus to convey some interpretation to the T-predicates: "Thus as a first approximation, it can be said that models are used in theory construction to specify new attributes as *the attributes which* share certain higher order attributes with attributes belonging to the model, fail to share certain others (the negative analogy)—and which satisfy, in addition, the conditions laid down by the relevant correspondence rules."[11] Thus, Sellars claims, both meaning and novelty are allowed for in the relation of Q- and P-models, and the P-model remains heuristically useful at least so long as its higher-order attributes remain implicit. When they are themselves formalized, presumably the P-model can be abandoned, and postulates representing higher-order attributes can be explicitly added to the Q-model and corresponding calculus.

I confess that many features of this suggestion remain obscure to me. In the first place, if the higher-order attributes can be referred to by intensional expressions such as "transitive," "perceptible," it is not clear that we have escaped the "myth of the given." These expressions are already in the descriptive language, and a more sophisticated analysis of "analogy," involving a type logic, could presumably take account of analogies depending on the sharing of these predicates, as it can of first-order predicates. If novelty depends on introduction of new predicates, there is no novelty here. If, on the other hand, the introduction of novelty depends essentially on the P- and Q-models merely *exhibiting* analogy of higher-order attributes, it is not clear that anything other than *formal* analogy has been introduced. *Each* model of a calculus "exhibits" such

[11] Sellars, "Scientific Realism or Irenic Instrumentalism," p. 181.

analogy with every other model of the same calculus; indeed once the calculus has been fully expressed, mention of models other than the Q-model is wholly redundant to this kind of analogy, since the higher-order attributes of the Q-model are already shown by the calculus itself. And in this case the construal of the T's as "the attributes which share certain higher-order attributes with attributes belonging to the [P-]model" does not seem to differ in principle from Nagel's account, which Sellars rejects, or from the deductivist's which we have seen reason to reject above. For in both these accounts it follows from the status of the Q- and P-models that there are *some* second- or higher-order attributes that they share, namely the relations exhibited in the calculus of which they are models.

It is possible, however, that Sellars has in mind a situation which is somewhere between the two extremes just mentioned, namely a P-model which tacitly introduces higher-order attributes in virtue of which we vaguely accept an "analogy" between it and the explanandum, but which have so far been unanalyzed, and for which we may not have names in the language. For example, we may recognize an analogy between a loud noise and a bright flash, and may exploit it in a P-model for light drawn from the phenomena of sound, without necessarily having in our language a concept "intensity" which applies to both noise and flash. However, the question now arises whether the P-model is introduced here because of antecedent recognition of an analogy, even though this was inexpressible in the existing language, or whether the adoption of this P-model itself *introduces* a new higher-order attribute which sound and light phenomena share.[12] The answer to this question is, surely, "six of one and half a dozen of the other." But such liberality must not be taken to the point of admitting *any* model as a candidate for the P-model. Unless *some* analogy of predicates, whether first or higher order, is recognized, which is not merely the relation of isomorphism between two models of some same calculus, use of a P-model in theoretical inference becomes vacuous, as I shall now try to show.

Lying behind Sellars's attempt to reconcile the meaningfulness of theoretical concepts with the possibility of novelty is some obscurity about the function of models in *inference*. Sellars has not given sufficient weight to the fact that my plea in *Models and Analogies in Science* for recognition of the logical role of models depends essentially upon taking *predictive*

[12] A closely related question is discussed in M. Black, *Models and Metaphors* (Ithaca, N.Y.: Cornell University Press, 1962), pp. 37–38.

power as a necessary condition for theories. From this point of view, acceptance of his shared higher-order predicates as ingredients in the role of models in relation to theories will depend upon whether or not we regard such shared predicates as justifying analogical inference. Take the example of transitivity, and assume that the P-model contains the first-order relation "before," and that the Q-model contains a relation R which is either said or shown to be transitive. According to Sellars we need know nothing about R except that it satisfies the postulates of the theory and is transitive. Waiving now the question of what relation R *is*, we must nevertheless ask how the P-model helps us to make analogical inferences in the theory. The answer is, surely, not at all. For suppose we risk an analogical inference from P-statements involving "before" to Q-statements involving R. There is a large class of relations all members of which are consistent with what is known about R, but clearly the analogical inference will not be valid for all of them. Suppose R is, in God's private eye, "larger than," then an analogical inference involving R would be equally as justifiable or unjustifiable as a similar inference involving "smaller than," but the conclusion cannot be true in both cases. Any analogical inference may of course lead to false conclusions for *empirical* reasons, but in this case one or the other conclusion must be false for *logical* reasons. On the other hand, if R is *known to be* a particular transitive relation having other affinities with "before," such as "to the left of," there might very well be a justifiable inference from the P-model to the Q-model if the theory were concerned, for example, with a geometry of space-time. In other words, whether higher-order predicates can function to justify analogical inference is a question only to be decided by examining particular predicates, and the possibility of doing this presupposes either that the predicates are already in the language or that they can be coined as required to name particular attributes in virtue of which an analogy is suspected. It cannot be the case that every shared higher-order predicate is sufficient to generate a justifiable analogical inference, for this would certainly lead to inconsistency, as in the "transitivity" example above, and would even be vacuous, if it could be shown that any system shares *some* higher-order attributes with every other system, a proposition which should not be too difficult to prove if our ontology is generous enough.

Finally, a remark about novelty and the myth of the given. Sellars's objection to identifying the predicates of the Q- and P-models seems to follow from his rejection (page 184) of the assumption (which he ascribes to Nagel) that there is a one-to-one correspondence between predicates and the extralinguistic attributes to which they apply. If this assumption were true, then indeed it would be difficult to see how analogical inference from observable predicates of the P-model would leave room for any theoretical novelty. Sellars instead wants to allow for enrichment and revision of the observational vocabulary by its interaction with theory. But use of models as analogues neither presupposes such a one-to-one correspondence nor rules out the kind of interaction of observation and theory that Sellars requires. Indeed, if we understand predicates as *analogical* in their applications to different situations rather than as either univocal or equivocal (as suggested above and discussed in the references there given), the possibility of novelty is safeguarded by the indefinite variety of analogical extensions of existing predicates. It is, after all, rare for new descriptive predicates to be *coined*, and much more common for new situations to be described by complex combinations of old predicates. I do not wish to deny, however, in anything I have said here, that totally new concepts may sometimes be required in theoretical science. It only seems to me to follow from the foregoing arguments that *if* totally new concepts are introduced, there is no possibility of theoretical inference of the kind discussed here, and consequently I suspect that such occasions of total novelty are rare.

To summarize the conclusions of this paper with regard to inductive and deductive construals of theories: Though one may admit that deductivism represents our existing knowledge about the theory of a given set of observable laws, it does not explicate the kinds of inference to prediction that we normally find justifiable. And these inferences ought not to be dismissed to the context of discovery or heuristics or psychology or history, but ought to be seen as part of the content of theories as we state them. Therefore they should feature in an adequate construal of theories. My alternative to the deductive account has been an attempt to show how this may be done.

GROVER MAXWELL

Structural Realism and the Meaning of Theoretical Terms

A *theory*, in the sense used in this paper, is a set of statements some of which both refer to unobservables and are capable of functioning essentially in the derivation from the theory of statements that are observationally decidable. Theories will be considered to be *semantically autonomous*; this means that there will be no (metalinguistic) rules whose function is to relate unobservables to observables (e.g., by relating theoretical terms to observation terms). This requirement amounts to the same as Carnap's later practice of eliminating "correspondence rules" ("C-rules") in favor of C-postulates (and, we might add, C-theorems). Such a C-statement is any nonmolecular statement that contains both theoretical terms and observation terms.

It will be assumed that, in spite of notorious difficulties, a distinction between the unobservable and the observable can be made. The program proposed in this paper will be applicable no matter how, within reasonable bounds, such a distinction is drawn. My own view on this matter, which I shall not try to defend here, is that all items should be considered theoretical unless they occur in direct experience; since I reject any form of direct realism, this means that the observable is instantiated only in inner events of observers. One question that the paper attempts to answer is this: Granted that a theory contains two kinds of terms one of which is relatively semantically unproblematic (commonly called observation terms) and the other (theoretical terms) for which an account of meaning seems to be required, how is such an account to be given? Russell's *principle of acquaintance* implemented by a device proposed by Ramsey provides, I believe, a satisfactory answer. I shall state the former in semantical terms as follows: All the descriptive (nonlogical) terms in any meaningful sentence refer to items with which we are acquainted. (Russell's usual formulation is to the effect that every ingredient of any proposition that