Borrower: S4N

Call #: Q175 .H56

Lending String: *FDA,IAS,ICY,IBS,KNM

Location: DIRAC

Patron: Jung, DongWook

Mail
Charge
Maxcost: $10.00IFM

Journal Title: Models and analogies in science,

Volume: **Issue:**
Month/Year: **Pages:** 157-177

Shipping Address:
Interlibrary Loan Division
Seoul National University Library,
San 56-1, Shillim-dong, Kwanak-ku
Seoul 151-742, Korea
2 1254 00319 8940

Article Author: Hesse, Mary B.

Article Title: Mary B. Hesse, 'The Explanatory Function of Metaphor', Models and Analogies in Science. University of Notre Dame Press, pp. 157-177.

Fax: 82-2-889-7616
Ariel:
Odyssey:
Email: jnong@snu.ac.kr

Imprint: [Notre Dame, Ind.] University of Notre D

__ **Saved on "P" Drive**
__✓ **NOS** (Not on Shelf)
__ **NFAC** (Not Found as Cited)
__ **LACKING**

ILL Number: 63469389

PROCESSED BY _____

not from the point of view of his own reliance on univocity as a primitive notion. How, if at all, this conclusion would ease the problem of metaphysical analogy is beyond the scope of this paper. In Aristotle's case, I think it would have little direct effect, since he in any case *presupposes* analogical predication in order to discuss the metaphysical terms. Its effect would rather be on his theory of induction and of science. But the conclusion should give no comfort at all to those neoscholastics and others who try to elucidate analogy in metaphysical and theological contexts from an Aristotelian standpoint. There are, I submit, *no* further resources in Aristotle for this undertaking, precisely because the elucidation of analogy was not his problem and because progress appears to depend on the rejection of his assumptions about univocity. He was right, however, in seeing a close link between the (for him) pseudo analogies of science and the true analogies of metaphysics, and this insight can be salvaged, indeed should be energetically exploited, because we now know far more than he could have known about the analogical uses of language implicit in the growth of science. The relevance of this to metaphysics and theology may not be negligible.

The Explanatory Function of Metaphor[1]

The thesis of this paper is that the deductive model of scientific explanation should be modified and supplemented by a view of theoretical explanation as metaphoric redescription of the domain of the explanandum. This raises two large preliminary questions: first, whether the deductive model requires modification, and second, what is the view of metaphor presupposed by the suggested alternative. I shall not discuss the first question explicitly. Much recent literature in the philosophy of science (for example, 4, 5, 10, 14)[2] has answered it affirmatively, and I shall refer briefly at the end to some difficulties tending to show that a new model of explanation is required, and suggest how the conception of theories as metaphors meets these difficulties.

The second question, about the view of metaphor presupposed, requires more extensive discussion. The view I shall present is essentially due to Max Black, who has developed in two papers, entitled, respectively, "Metaphor" and "Models and Archetypes" (3), both a new theory of metaphor and a

1. Presented at the Congress of the International Union for the Logic, Methodology and Philosophy of Science, Jerusalem, March, 1964. Reprinted by permission.

2. Numbers in parentheses refer to literature listed in the Bibliography at the end of this volume.

parallelism between the use of literary metaphor and the use of models in theoretical science. I shall start with an exposition of Black's *interaction view* of metaphors and models, taking account of modifications suggested by some of the subsequent literature on metaphor (1, 2, 11, 13, 15). It is still unfortunately necessary to argue that metaphor is more than a decorative literary device and that it has cognitive implications whose nature is a proper subject of philosophic discussion. But space forces me to mention these arguments as footnotes to Black's view, rather than as an explicit defence *ab initio* of the philosophic importance of metaphor.

The Interaction View of Metaphor

1. We start with two systems, situations, or referents, which will be called, respectively, the "primary" and the "secondary" systems. Each is describable in literal language. A metaphoric use of language in describing the primary system consists of transferring to it a word or words normally used in connection with the secondary system: for example, "Man is a wolf," "Hell is a lake of ice." In a scientific theory the primary system is the domain of the explanandum, describable in observation language; the secondary is the system, described either in observation language or the language of a familiar theory, from which the model is taken: for example, "Sound (primary system) is propagated by wave motion (taken from a secondary system)"; "Gases are collections of randomly moving massive particles."

Three terminological remarks should be inserted here. First, "primary system" and "secondary system" and "domain of the explanandum" will be used throughout to denote the referents or putative referents of descriptive statements; "metaphor," "model," "theory," "explanans," and "explanandum" will be used to denote linguistic entities. Second, use of the terms "metaphoric" and "literal," "theory" and "observation" need not be taken at this stage to imply a pair of irreducible dichotomies. All that is intended is that the literal and observation languages are assumed initially to be well understood and unproblematic, whereas the metaphoric and theoretical are in need of analysis. The third remark is that to assume initially that the two systems are "described" in literal or observation language does not imply that they are exhaustively or accurately described or even that they could in principle be so in terms of these languages.

2. We assume that the primary and secondary systems each carries a set of associated ideas and beliefs that come to mind when the system is referred to. These are not private to individual language-users but are largely common to a given language community and are presupposed by speakers who intend to be understood in that community. In literary contexts the associations may be loosely knit and variable, as in the wolf-like characteristics that come to mind when the metaphor "Man is a wolf" is used;

in scientific contexts the primary and secondary systems may both be highly organized by networks of natural laws.

A remark must be added here about the use of the word "meaning." Writers on metaphor appear to intend it as an inclusive term for reference, use, and the relevant set of associated ideas. It is, indeed, part of their thesis that it has to be understood thus widely. To understand the meaning of a descriptive expression is not only to be able to recognize its referent, or even to use the words in the expression correctly, but also to call to mind the ideas, both linguistic and empirical, that are commonly held to be associated with the referent in the given language community. Thus a shift of meaning may result from a change in the set of associated ideas as well as in change of reference or use.

3. For a conjunction of terms drawn from the primary and secondary systems to constitute a metaphor it is necessary that there should be patent falsehood or even absurdity in taking the conjunction literally. Man is not, literally, a wolf; gases are not in the usual sense collections of massive particles. In consequence some writers have denied that the referent of the metaphoric expression can be identified with the primary system without falling into absurdity or contradiction. I shall return to this in the next section.

4. There is initially some principle of assimilation between primary and secondary systems, variously described in the literature as "analogy," "intimations of similarity," "a programme for exploration," "a framework through which the primary is seen." Here we have to guard against two opposite interpretations, both of which are inadequate for the general understanding of metaphors and scientific models. On the one hand, to describe this ground of assimilation as a *program* for exploration or a *framework* through which the primary is seen, is to suggest that the secondary system can be imposed a priori upon the primary, as if *any* secondary can be the source of metaphors or models for *any* primary, provided the right metaphor-creating operations are subsequently carried out. Black does indeed suggest that in some cases "it would be more illuminating . . . to say that the metaphor creates the similarity than to say it formulates some similarity antecedently existing" (p. 37), and he also points out that some poetry creates new metaphors precisely by developing itself the system of associations in terms of which "absurd" conjunctions of words are to be metaphorically understood.

There is, however, an important distinction to be brought out between such a use of metaphor and scientific models, for, whatever may be the case for poetic use, the suggestion that *any* scientific model can be imposed a priori on *any* explanandum and function fruitfully in its explanation must be resisted. Such a view would imply that theoretical models are irrefutable. That this is not the case

is sufficiently illustrated by the history of the concept of a heat fluid or the classical wave theory of light. Such examples also indicate that no model even gets off the ground unless some antecedent similarity or analogy is discerned between it and the explanandum.

But here there is a danger of falling into what Black calls the *comparison view* of metaphor. According to this view the metaphor can be replaced without remainder by an explicit, literal statement of the similarities between primary and secondary systems, in other words, by a simile. Thus, the metaphor "Man is a wolf" would be equivalent to "Man is like a wolf in that . . ." where follows a list of comparable characteristics; or, in the case of theoretical models, the language derived from the secondary system would be wholly replaced by an explicit statement of the analogy between secondary and primary systems, after which further reference to the secondary system would be dispensable. Any interesting examples of model-using in science will show, however, that the situation cannot be described in this way. For one thing, as long as the model is under active consideration as an ingredient in an explanation, we do not know how far the comparison extends—it is precisely in its extension that the fruitfulness of the model may lie. And a more fundamental objection to the comparison view emerges in considering the next point.

5. The metaphor works by transferring the associated ideas and implications of the secondary to the primary system. These select, emphasize, or suppress features of the primary; new slants on the primary are illuminated; the primary is "seen through" the frame of the secondary. In accordance with the doctrine that even literal expressions are understood partly in terms of the set of associated ideas carried by the system they describe, it follows that the associated ideas of the primary are changed to some extent by the use of the metaphor and that, therefore, even its original literal description is shifted in meaning. The same applies to the secondary system, for its associations come to be affected by assimilation to the primary; the two systems are seen as more like each other; they seem to interact and adapt to one another, even to the point of invalidating their original literal descriptions if these are understood in the new, postmetaphoric sense. Men are seen to be more like wolves after the wolf metaphor is used, and wolves seem to be more human. Nature becomes more like a machine in the mechanical philosophy, and actual, concrete machines themselves are seen as if stripped down to their essential qualities of mass in motion.

This point is the kernel of the interaction view and is Black's major contribution to the analysis of metaphor. It is incompatible with the comparison view, which assumes that the literal descriptions of both systems are and remain independent of the use of the metaphor and that the metaphor is reducible

to them. The consequences of the interaction view for theoretical models are also incompatible with assumptions generally made in the deductive account of explanation, namely, that descriptions and descriptive laws in the domain of the explanandum remain empirically acceptable and invariant in meaning to all changes of explanatory theory. I shall return to this point.

6. It should be added as a final point in this preliminary analysis that a metaphoric expression used for the first time, or used to someone who hears it for the first time, is intended to be *understood*. Indeed, it may be said that a metaphor is not metaphor but nonsense if it communicates nothing and that a genuine metaphor is also capable of communicating something other than was intended and hence of being *mis*understood. If I say (taking two words more or less at random from a dictionary page) "A truck is a trumpet," it is unlikely that I shall communicate anything; if I say "He is a shadow on the weary land," you may understand me to mean (roughly) "He is a wet blanket, a gloom, a menace," whereas I actually meant (again roughly) "He is a shade from the heat, a comfort, a protection."

Acceptance of the view that metaphors are meant to be intelligible implies rejection of all views that make metaphor a wholly noncognitive, subjective, emotive, or stylistic use of language. There are exactly parallel views of scientific models that have been held by many contemporary philosophers of

science, namely, that models are purely subjective, psychological, and adopted by individuals for private heuristic purposes. But this is wholly to misdescribe their function in science. Models, like metaphors, are intended to communicate. If some theorist develops a theory in terms of a model, he does not regard it as a private language but presents it as an ingredient of his theory. Neither can he, nor need he, make literally explicit all the associations of the model he is exploiting; other workers in the field "catch on" to its intended implications—indeed, they sometimes find the theory unsatisfactory just because some implications the model's originator did not investigate, or even think of, turn out to be empirically false. None of this would be possible unless use of the model were intersubjective, part of the commonly understood theoretical language of science, not a private language of the individual theorist.

An important general consequence of the interaction view is that it is not possible to make a distinction between literal and metaphoric descriptions merely by asserting that literal use consists in following linguistic rules. Intelligible metaphor also implies the existence of rules of metaphoric use, and since in the interaction view literal meanings are shifted by their association with metaphors, it follows that the rules of literal usage and of metaphor, though they are not identical, are nevertheless not independent. It is not sufficiently clear in Black's

paper that the interaction view commits one to the abandonment of a two-tiered account of language in which some usages are irreducibly literal and others metaphoric. The interaction view sees language as dynamic: an expression initially metaphoric may become literal (a "dead" metaphor), and what is at one time literal may become metaphoric (for example, the Homeric "He breathed forth his life," originally literal, is now a metaphor for death). What is important is not to try to draw a line between the methaphoric and the literal, but rather to trace out the various mechanisms of meaning-shift and their interactions. The interaction view cannot consistently be made to rest on an initial set of absolutely literal descriptions, but rather on a relative distinction of literal and metaphoric in particular contexts. I cannot undertake the task of elucidating these conceptions here (an interesting attempt to do so has been made by K. I. B. S. Needham (12), but I shall later point out a parallel between this general linguistic situation and the relative distinctions and mutual interactions of theory and observation in science.

The Problem of Metaphoric Reference

One of the main problems for the interaction view in its application to theoretical explanation is the question what is the *referent* of a model or metaphor. At first sight the referent seems to be the primary system, which we choose to describe in metaphoric rather than literal terms. This, I believe, is in the

end the right answer, but the process of metaphoric description is such as to cast doubt on any simple identification of the metaphor's reference with the primary system. It is claimed in the interaction view that a metaphor causes us to "see" the primary system differently and causes the meanings of terms originally literal in the primary system to shift toward the metaphor. Thus "Man is a wolf" makes man seem more vulpine, "Hell is a lake of ice" makes hell seem icy rather than hot, and a wave theory of sound makes sound seem more vibrant. But how can initial similarities between the objective systems justify such changes in the meanings of words and even, apparently, in the things themselves? Man does not in fact change because someone uses the wolf metaphor. How then can we be justified in identifying what we see through the framework of the metaphor with the primary system itself? It seems that we cannot be entitled to say that men *are* wolves, sound *is* wave motion, in any identificatory sense of the copula.

Some recent writers on metaphor (2, 11, 15) have made it the main burden of their argument to deny that any such identification is possible. They argue that if we allow it we are falling into the absurdity of conjoining two literally incompatible systems, and the resulting expression is not metaphoric but meaningless. By thus taking a metaphor literally we turn it into a myth. An initial misunderstanding may be removed at once by remarking that "identification"

cannot mean in this context identification of the referent of the metaphoric expression, taken in its *literal* sense, with the primary system. But if the foregoing analysis of metaphor is accepted, it follows that metaphoric use is use different from the literal sense, and furthermore it is use in a sense not replaceable by any literal expression. There remains the question what it is to identify the referent of the metaphoric expression or model with the primary system.

As a preliminary to answering this question, it is important to point out that there are two ways, which are often confused in accounts of the "meaning of theoretical concepts," in which such identification may fail. It may fail because it is in principle meaningless to make any such identification, or it may fail because in a particular case the identification happens to be false. Instances of false identification, e.g., "Heat is a fluid" or "The substance emitted by a burning object is phlogiston," provide no arguments to show that other such identifications may not be both meaningful and true.

Two sorts of argument have been brought against the view that metaphoric expressions and models can refer to and truly describe the primary system. The first depends on an assimilation of poetic and scientific metaphor and points out that it is characteristic of good poetic metaphor that the images introduced are initially striking and unexpected, if not shocking; that they are meant to be entertained

and savored for the moment and not analyzed in pedantic detail nor stretched to radically new situations; and that they may immediately give place to other metaphors referring to the same subject matter which are formally contradictory, and in which the contradictions are an essential part of the total metaphoric impact. Any attempt to separate these literal contradictions from the nexus of interactions is destructive of the metaphor, particularly in the interaction view. In the light of these characteristics there is indeed a difficult problem about the correct analysis of the notion of metaphoric "truth" in poetic contexts. Scientific models, however, are fortunately not so intractable. They do not share any of the characteristics listed above that make poetic metaphors peculiarly subject to formal contradictoriness. They may initially be unexpected, but it is not their chief aim to shock; they are meant to be exploited energetically and often in extreme quantitative detail and in quite novel observational domains; they are meant to be internally tightly knit by logical and causal interrelations; and if two models of the same primary system are found to be mutually inconsistent, this is not taken (*pace* the complementarity interpretation of quantum physics) to enhance their effectiveness but rather as a challenge to reconcile them by mutual modification or to refute one of them. Thus their truth criteria, although not rigorously formalizable, are at least much clearer than in the case of poetic metaphor. We can perhaps signal-

ize the difference by speaking in the case of scientific models of the (perhaps unattainable) aim to find a "perfect metaphor," whose referent is the domain of the explanandum, whereas literary metaphors, however adequate and successful in their own terms, are from the point of view of potential logical consistency and extendability often (not always) intentionally imperfect.

Second, if the interaction view of scientific metaphor or model is combined with the claim that the referent of the metaphor is the primary system (that is, the metaphor is true of the primary system), then it follows that the thesis of meaning-invariance of the literal observation-descriptions of the primary system is false. For, the interaction view implies that the meaning of the original literal language of the primary system is changed by adoption of the metaphor. Hence those who wish to adhere to meaning-invariance in the deductive account of explanation will be forced to reject either the interaction view or the realistic view that a scientific model is putatively true of its primary system. Generally they reject both. But abandonment of meaning-invariance, as in many recent criticisms of the deductive model of explanation, leaves room for adoption of both the interaction view and realism, as I shall now try to spell out in more detail.

Explanation as Metaphoric Redescription

The initial contention of this paper was that the deductive model of explanation should be *modified* and *supplemented* by a view of theoretical explanation as metaphoric redescription of the domain of the explanandum. First, the association of the ideas of metaphor and of explanation requires more examination. It is certainly not the case that all explanations are metaphoric. To take only two examples, explanation by covering-law, where an instance of *A* which is *B* is explained by reference to the law "All *A's* are *B's*" is not metaphoric; neither is the explanation of the working of a mechanical gadget by reference to an actual mechanism of cogs, pulleys, and levers. These, however, are not examples of *theoretical* explanation, for it has been taken for granted that the essence of a theoretical explanation is the introduction into the explanans of a new vocabulary or even of a new language. But introduction of a metaphoric terminology is not in itself explanatory, for in literary metaphor in general there is no hint that what is metaphorically described is also thereby explained. The connection between metaphor and explanation is, therefore, neither that of necessary nor sufficient condition. Metaphor becomes explanatory only when it satisfies certain further conditions.

The orthodox deductive criteria for a scientific explanans—for example, in Hempel and Oppenheim (6)—require that the explanandum be deducible from it, that it contain at least one general law not redundant to the deduction, that it be not empir-

ically falsified to date, and that it be predictive. We cannot simply graft these requirements on to the account of theories as metaphors without investigating the consequences of the interaction view of metaphor for the notions of "deducibility," "explanandum," and "falsification" in the orthodox account. In any case, as has been mentioned already, the requirement of deducibility in particular has been subjected to damaging attack, quite apart from any metaphoric interpretation of theories. There are two chief grounds for this attack, both of which can be turned into arguments favorable to the metaphoric view.

In the first place, it is pointed out that there is seldom in fact a deductive relation strictly speaking between scientific explanans and explanandum, but only relations of approximate fit. Furthermore, what counts as sufficiently approximate fit cannot be decided deductively but is a complicated function of coherence with the rest of a theoretical system, general empirical acceptability throughout the domain of the explanandum, and many other factors. I do not propose to try to spell out these relationships in further detail here, but merely to make two points relevant to my immediate concern. First, the attack on deducibility drawn from the occurrence of approximations does not imply that there are *no* deductive relations between explanans and explanandum. The situation is rather this. Given a descriptive statement D in the domain of the explanandum, it

is usually the case that the statement E of an acceptable explanans does not entail D, but rather D', where D' is a statement in the domain of the explanandum only "approximately equivalent" to D. For E to be acceptable it is necessary both that there be a deductive relation between E and D', and that D' should come to be recognized as a *more acceptable* description in the domain of the explanandum than D. The reasons why it might be more acceptable—repetition of the experiments with greater accuracy, greater coherence with other acceptable laws, recognition of disturbing factors in arriving at D in the first place, metaphoric shifts in the meanings of terms in D consequent upon the introduction of the new terminology of E, and so on—need not concern us here. What is relevant is that the nondeducibility of D from E does not imply total abandonment of the deductive model unless D is regarded as an invariant description of the explanandum, automatically rendering D' empirically false. That D cannot be so regarded has been amply demonstrated in the literature.

The second point of contact between these considerations and the view of theories as metaphors is now obvious. That explanation may modify and correct the explanandum is already built into the relation between metaphors and the primary system in the interaction view. Metaphors, if they are good ones, and ipso facto their deductive consequences, do have the primary system as their referents, for

they may be seen as correcting and replacing the original literal descriptions of the same system, so that the literal descriptions are discarded as inadequate or even false. The parallel with the deductive relations of explanans and explananda is clear: the metaphoric view does not abandon deduction, but it focusses attention rather on the interaction between metaphor and primary system, and on the criteria of acceptability of metaphoric descriptions of the primary system, and hence not so much upon the deductive relations that appear in this account as comparatively uninteresting pieces of logical machinery.

The second attack upon the orthodox deductive account gives even stronger and more immediate grounds for the introduction of the metaphoric view. It is objected that there are no deductive relations between theoretical explanans and explanandum because of the intervention of correspondence rules. If the deductive account is developed, as it usually is, in terms either of an uninterpreted calculus and an observation language, or of two distinct languages —the theoretical and the observational, it follows that the correspondence rules linking terms in these languages cannot be derived deductively from the explanans alone. Well-known problems then arise about the status of the correspondence rules and about the meaning of the predicates of the theoretical language. In the metaphoric view, however, these problems are evaded, because here there are

no correspondence rules, and this view is primarily designed to give its own account of the meaning of the language of the explanans. There is *one* language, the observation language, which like all natural languages is continually being extended by metaphoric uses and hence yields the terminology of the explanans. There is no problem about connecting explanans and explanadum other than the general problem of understanding how metaphors are introduced and applied and exploited in their primary systems. Admittedly, as yet we are far from understanding this process, but to see the problem of the "meaning of theoretical concepts" as a special case of it is one step in the solution of this problem.

Finally, a word about the requirement that an explanation be predictive. It has been much debated within the orthodox deductive view whether this is a necessary and sufficient condition for explanation, and it is not appropriate here to enter into that debate. But any account of explanation would be inadequate which did not recognize that, in general, an explanation is required to be predictive or, what is closely connected with this, to be falsifiable. Elsewhere (8) I have pointed out that, in terms of the deductive view, the requirement of predictivity may mean one of three things.

1. That general laws already present in the explanans have as yet unobserved instances. This is a trivial fulfilment of the requirement and would not, I think, generally be regarded as sufficient.

2. That further general laws can be derived from the explanans *without* adding further items to the set of correspondence rules. That is to say, predictions remain within the domain of the set of predicates already present in the explanandum. This is a weak sense of predictivity that covers what would normally be called *applications* rather than extensions of a theory (for example, calculation of the orbit of a satellite from the theory of gravitation but not extension of the theory to predict the bending of light rays).

3. There is also a strong sense of prediction in which new observation predicates are involved, and hence, in terms of the deductive view, additions are required to the set of correspondence rules. I have argued (7, 8) that there is no rational method of adding to the correspondence rules on the pure deductive view, and hence that cases of strong prediction cannot be rationally accounted for on that view. In the metaphoric view, on the other hand, since the domain of the explanandum is redescribed in terminology transferred from the secondary system, it is to be expected that the original observation language will both be shifted in meaning and extended in vocabulary, and hence that predictions in the strong sense will become possible. They may, of course, turn out not to be *true,* but that is an occupational hazard of any explanation or prediction. They will, however, be rational, because rationality consists just in the continuous adaptation of

our language to our continually expanding world, and metaphor is one of the chief means by which this is accomplished.