Mackie, J. L. 1974: Mill's methods of induction. In *The Cement of the Universe* (Oxford: Oxford University Press).

Ryan, A. 1974: *J. S. Mill* (London: Routledge and Kegan Paul).

Scarre, G. 1989: *Logic and Reality in the Philosophy of John Stuart Mill* (Dordrecht: Kluwer Academic Publishers).

Skorupski, J. 1989: *John Stuart Mill* (London and New York: Routledge).

Whewell, W. 1968: *William Whewell's Theory of Scientific Method*, ed. Robert E. Butts (Pittsburgh: University of Pittsburgh Press).

44

Models and Analogies

MARY HESSE

Models in classical physics

Questions about the structure and justification of theories, the interpretation of data, and the problem of realism have been in the forefront of debate in recent philosophy of science, and the topic of models and analogies is increasingly recognized as integral to this debate. Models of physical matter and motion – for example, models of atoms and planetary systems – were already familiar in Greek science, but serious analysis of "model" as a concept entered philosophy of science only in the nineteenth century. This was largely the result of proliferation in classical physics of theoretical entities such as "atom," "electro-magnetic wave," and "electron," for which there appeared to be no directly observable evidence (see THEORETICAL TERMS).

The senses of "model" discussed in classical physics were of two types, which may be distinguished as "material" and "formal" (Hesse 1966). A *material model* is, or describes, a physical entity – familiar examples are billiard balls, a fluid medium, a spring, or an attracting or repelling electric particle. A *formal model* is the expression of the form or structure of physical entities and processes, without any semantic content referring to specific objects or properties. For example, a "wave equation" in mathematical symbols may express the laws of a simple pendulum, of sound or light waves, of quantum wave functions, etc., while remaining neutral to any specific application. Another example is the formal structure of a computer program (the software), which may be realized in a number of different hardware setups, and has provided useful formal models of brain structure in Artificial Intelligence. Formal models are syntactic structures; material models are semantic, in that they introduce reference to real or imaginary entities.

"Analogy" will be taken here to refer to some relation of similarity and/or difference between a model and the world, or (less question begging) between a model and some theoretical description of the world, or between one model and another. Models are relata of analogy relations; that is, a model is an analogue. Analogy relations themselves may be formal or material: they may be merely analogies of structure, such as that between a light wave and a simple pendulum, or they may introduce material similarities, as when gas particles are held to be like billiard balls in all mechanical properties relevant to Newton's laws.

Analogy relations, like similarity, come in degrees and in different respects, and are therefore not generally transitive. This makes rigorous treatment difficult, but it is useful as a start to distinguish three types of material analogy relation: positive, negative, and neutral. A *positive (material) analogy* picks out those features of the analogues that are identical or strongly similar; a *negative analogy* picks out those known to be different or strongly dissimilar; a *neutral analogy* picks out those for which there is no evidence yet as to similarity or dissimilarity. For example, DNA models built of painted balls and metal struts are positively analogous to DNA molecules in spatial structure and connectedness, but negatively analogous in size, material, shape, and color of the constituents, etc. These models have a neutral analogy with molecules insofar as their further detailed properties are used to explore as yet unknown features of genetic material. The dividing line between these three sorts of analogy will of course shift as research goes forward – the better the model, the more of the neutral analogy will eventually be accepted as positive, whereas a poor model will become more and more negatively analogous.

So much for the somewhat rough definitions that sufficed for the description of models in classical physics and chemistry. Models served there to introduce unobservable entities and processes into physical theory by analogy with familiar observable entities and processes, thus providing pictures of the explanatory entities held to underlie phenomena. The problem of justifying these explanatory models led to a polarization of epistemological views. Realists held that successful models are positive analogues of the real world; positivists denied the reality of the theoretical entities referred to, and regarded models merely as working pictures to be dispensed with in accepted theories, having at best a formal analogy with the world (see REALISM AND INSTRUMENTALISM).

The philosophical debate about models was initiated by Norman Campbell (1920, ch. 6) in the course of a critique of the so-called hypothetico-deductive (HD) theory of theories (see THEORIES). According to this view, of which Maxwell's electromagnetic theory is a paradigm case, an explanatory theory in physics consists of a set of mathematical equations, some but not all the terms of which are interpreted by means of directly observable or measurable properties, such as shape, position, momentum, time interval, weight, texture, color, light intensity, temperature, etc. These interpretations were called "bridge principles" or (with Campbell) the "dictionary." A theory is confirmed if, with the bridge principles, laws and predictions can be deduced and shown to give a good fit with the experimental *explananda*; if the fit is poor, the theory is disconfirmed or refuted. In the positivist version of HD, models are used only as aids to discovery, and are not a logically essential part of the theory.

Campbell argued, on the contrary, that models, as interpretations of unobservable terms, are essential elements of theory, because a merely mathematical formalism gives no meaningful information other than that contained in the experimental laws and properties themselves. Taking the "billiard ball" model of gases as his main example, he showed how the experimental laws are "explained" (unified and made intelligible) by this model and, most importantly, how theoretical inference proceeds by modification and extension of the model to give new predictions. The logic of such inference is analogical argument from the properties of the model's familiar source (observable mechanical particles) to the *explanandum* (gases). For example, the original point model of particles that explained Boyle's and Charles's laws is extended to particles of finite size, thus predicting the corrections to Boyle's law which are necessary to obtain greater experimental range and accuracy for real gases. Thus models are shown to be essential to argument in physics, not merely dispensable heuristic devices.

Two kinds of issue arise from this analysis, one epistemological and the other ontological. Campbell has an implicit epistemological thesis in his argument for the essentiality of models: namely, that they *justify* reliance on predictions from models in virtue of the known positive analogy between model and *explanandum*. That there is such reliance was pointed out by Putnam (1963, p. 779) in a striking example from the construction of the first atomic bomb. Although laboratory-sized tests of the nuclear reactions involved had been performed successfully, large-scale tests had not, and their failure would have been catastrophic. Such tests would not have been carried out unless there had been *some* intuitive confidence that analogical extrapolations from evidence and theory justified expectation of success (Hesse 1974, ch. 9).

Underlying all such intuitions there is a metaphysics of the "analogy of nature," and this brings with it ontological questions about the status of models. If the kinetic model, for example, has no relation to real analogies in nature beyond those already observed, there is no basis for prediction to its analogical extrapolations. Does this imply, however, that there *are* molecules as described in the theory? Campbell's reply to this question was subtle. The model of molecules is not identical with the substructure of gases, only *materially analogous* to it. Models are entities which share the properties of mechanical particles insofar as these are required to explain already known phenomena (the positive analogy), and to predict phenomena yet to be examined (the neutral analogy). But analogies always have negative elements, and realistic identification of models with nature is therefore unjustified. Campbell, then, was anti-realist about theoretical entities and some of their first-order properties, but realist about their positive analogy relations.

The semantic conception of theories

Campbell's view anticipates more recent emphasis on the tentative and dynamical character of theory making, in contrast with the HD account, in which theories tend to be seen ahistorically, as static formal systems. Subsequently, however, there has been a greater concern with static ontology than with dynamic epistemology, and the analysis of models has become part of the general philosophical debate about realism. The syntactical HD account has been transformed into the so-called semantic conception of theories (SCT), in which emphasis shifts from formal theory structure to the set of semantic or metamathematical models for the theory (Suppe 1989, pp. 86ff). Each model of this set is an interpretation of the formal system that makes the axioms of the system true. The models may be real entities or, more often, imaginary idealizations of real entities, such as frictionless planes, point particles, or workshop mock-ups of the next mark of stretched limosine; or they may be mathematical entities such as geometrical spaces as models of some geometric axiom set. The semantic content of a theory is then said to be the whole class of its models – that is, all possible interpretations. If the theory is empirically acceptable, the real world will be (probably only approximately) among these models. This "family of models" is a highly abstract conception, carrying no information other than the structure of its parent formal system. Even if the models are conceived in some sense as real entities, the properties they have over and above their formal structure are irrelevant to the theory; as "models of the theory," they are logically equivalent, and therefore do not compete with each other for "reality" or "truth."

The semantic conception makes a welcome move from talk about linguistic formulations to talk about things and processes, and thus comes nearer to talk of models as this actually occurs in science. But SCT adds little of philosophical interest to the topic of models itself, and nothing to the epistemological issues of the previous section (van Fraassen 1989, p. 216). Emphasis is still on the properties of a theory as frozen in a particular structural formulation. It is significant how many accounts of SCT refer to theories as expressed in "textbooks" (Cartwright 1983, p. 46; Giere 1988, p. 78). Like HD, SCT has nothing to say about theory change, or about general theory frameworks or "paradigms," because these are rarely formalizable in deductive axiom systems, and therefore do not define a set of semantic models (Suppe 1989, p. 269). The old problem of the "meaning of theoretical terms" gets pushed into the philosophy of language rather than philosophy of science. How models are thought up, and how their descriptive terminology is understood, becomes no different in this view from the introduction of any new terms into language, whether in new dialects, novels, science fiction, or literature in general (van Fraassen 1980, 221). But such distinctions between the philosophy of models and the philosophy of language are unjustified. Connections have already been found, for example, between the use of models as scientific metaphors and the linguistic analysis of metaphor in general (Black 1962, chs 3, 13; Hesse 1966, pp. 157ff) (see METAPHOR IN SCIENCE). Such comparisons have important implications for the philosophy of both science and language, and it makes no sense to exclude discussion of the development of scientific language from analysis of the structure of science.

An even greater weakness of the semantic conception lies in its tacit acceptance of the distinction made in classic HD between theoretical and observation terms. It is now generally accepted that this is a grave oversimplification (see OBSERVATION AND THEORY). As long ago as 1960, Suppes pointed out that the subject matter of science is not raw observation, but *models of data*. In the case of mathematical science, these come as sets of measurable quantities representing observable properties derived from idealizations of the real world, and not from raw experience. For example, theories of mechanics are related to experience by means of a set of variables interpreted as particles, time intervals, and space, mass, and force functions. These represent idealized mechanical entities and their measurable properties. Suppes himself did not go on to discuss *un*observable terms, but subsequently the much more general thesis of theory-ladenness of observation has blurred sharp distinctions between "observable" and "unobservable," and made his analysis relevant to theoretical models also. The question of what the particular sensory equipment of *Homo sapiens* can or cannot directly observe has lost most of its interest in relation to the nature and structure of theory. Scientific knowledge can now be conceived as a hierarchy of models, some of which are more particular and lie closer to the data, some of which are theoretical and more distantly related to the world.

What, though, *is* this theory–world relation? Answers to the question within SCT depend on how far it is construed as a realist or anti-realist theory of science. The generally received view is realist, at least in the sense that the real world is supposed to be (approximately) among the models of a good theory, and attempts have been made to specify criteria of "goodness" which will reduce the indefinitely large set of possible models to a manageable few. These criteria, by definition, have to be nonempirical, because it is assumed that the family of models which constitute a successful theory are all consistent with the data so far (or, rather, with models of these data). Different data define different theories. Nonempirical criteria that have been suggested include unification of phenomena, formal simplicity, and economy and non-*ad hoc*-ness of theory, but there has so far been little success in showing that these criteria are relevant to *truth*, or in showing that sequences of theories in a particular domain tend to converge upon a unique "best explanation" (see EXPLANATION, and INFERENCE TO THE BEST EXPLANATION).

Ronald Giere has suggested a more flexible realist version of SCT, which he calls "constructive realism." Here it is explicitly recognized that there is some looseness of fit between theoretical models, models of data, and the real world. Even in the HD conception, numerical approximation and statistical likelihood already disrupt the purely deductive character of theories. More generally, Giere identifies *similarity* as the primary relation between all types of models and the real world (Giere 1988, p. 81). This is logically an intransitive relation, and cannot yield "truth" or "correspondence." Giere declines to discuss it in logical terms at all, but regards the recognition of sufficient similarity in relevant respects as a wholly natural cognitive process, depending both on human biological capacities and on socially accepted conventions and paradigms (Giere 1988, pp. 94ff). In this type of realism there is no guarantee of convergence of finality in the process of theory making; it is an ontological analysis of what a theory *is*, not of how it is developed or justified.

Giere's constructive realism brings SCT closer to real science, and also to the type of anti-realism or "constructive empiricism" adopted by van Fraassen (1980). The difference between these two views seems to relate chiefly to the nature of the theory–observation distinction. Where Giere sees a seamless hierarchy of models of theory and data, van Fraassen makes a distinction (which cannot be more than pragmatic) between the empirical adequacy of a theory and the nonrealistic models whose relation to experience is mediated through the deductive apparatus of the theory and its bridging principles. Thus the relation of theory to world remains one of satisfaction of propositions, that is of "truth" or "correspondence," but at the empirical level only. Theory models are not held to carry truth-values in relation to the world in any interesting sense. Both Giere and van Fraassen, however, continue to neglect problems of theory change and model choice, preferring to refer these either to cognitive neurophysiology or to the general philosophy of perception and language. In other words, the ghosts of the formal, static, HD and SCT approaches still linger.

The analogical conception of theories

In order to address issues of meaning and justification, we need to abandon two dogmas still lurking in SCT. The first is the undue concentration on ontology and realism at the cost of banishing linguistic and epistemological questions from philosophy of science. The second is the emphasis on static, "textbook" formulations of theory, to the neglect of the ongoing process of theory making and the consequent problems of theory choice and theory change. Recent discussions have made a sharp break with both these dogmas, chiefly as a result of detailed studies of the historical and contemporary dynamics of theory and experiment.

The new approaches emphasize empirical study of science itself, rather than "logical reconstructions" of it. The theory–world relation is explicitly described in terms of physiological and cognitive science, instead of being regarded as a deep and intractable philosophical problem, and the attempts to find rigorous logical relations throughout scientific theory are replaced by various degrees of approximation, looseness of fit, similarities, and analogies. The new approaches have to face the objection that all this necessarily results in fuzzy thinking. They have to show that, although reality and science *are* irreducibly fuzzy, nevertheless philosophical talk about them can be conducted in rigorous, precise, and intelligible terms, but without falling into unrealistic and inapplicable logic. They have begun to do this by reintroducing similarity, analogy, and related concepts into serious philosophical discussion, thus releasing model talk from the metamathematical straitjacket in which SCT has encased it (e.g., Gooding 1990; Harré 1986, ch. 11).

These points emerge explicitly in the analysis put forward by Nancy Cartwright in what she calls the "simulacrum theory of explanation," described as follows: "To explain a phenomenon is to find a model that fits it into the basic framework of the theory and that allows us to derive *analogues* for the messy and complicated phenomenological laws which are true of it" (Cartwright 1983, p. 152, italics added). Here models cease to be abstract metamathematical entities, and are seen in a more historical light as what groups of scientists adopt as manageable paradigms. The indefinitely large "families of models" are thus reduced to very few working models with their empirical bridge principles. Cartwright argues that these have no claim to reality status or truth – they are fictions, used piecemeal, exploited, and superseded to suit convenience. So far, her conception is similar to that of van Fraassen, and, like him, she maintains a split between lower and higher levels of theorizing. But she differs from him in allowing that *un*observable causes and entities do exist, and that causal laws have truth-values, at least locally (Cartwright 1983, pp. 160f).

Cartwright's argument is bolstered by a wealth of detailed examples from physics. It remains unclear, however, just how the concept of "real cause" is distinguished from fictitious theory models. Cartwright does not seem to hold a strong *modal* concept of "natural kinds" or of laws (1983, p. 95); so it is not easy to see what the notion of "true causal relations" contributes that cannot equally be said (in local contexts) in terms of empirically adequate laws and models of the data (see NATURAL KINDS). It seems preferable to tell the same story all along the theory–observation spectrum. To talk in terms of propositions for a moment, models can then be regarded as satisfying theoretical propositions with truth-value throughout; but at higher levels of theory these are almost certainly *false*, whereas nearer the phenomenal level they are likely to be approximately and locally *true*, because they are subject to multiple sources of evidence and test. We then have a conception of theory as essentially an embodiment of analogies, both formal and material, which describe regularities among the data of a given domain (models of data and phenomenal laws), with analogies between these and models of data in other domains, and so on in a hierarchy of levels of a unifying theoretical system. The "meaning of theoretical terms" is given by analogies with familiar natural processes (e.g., mechanical systems), or by hypothetical models (e.g., Bohr's planetary atom). In either case, descriptive terms of the analogues are derived metaphorically from ordinary language.

A useful model, then, represents the real world, not by correspondence or isomorphism, but by analogy, and this may be strong or weak, depending on how much evidence there is from different analogous domains. The justification of predictions from models to new domains becomes a question of the strength of analogical argument within the whole theory–data network. That strong analogues justify prediction in turn depends on a metaphysical and inductive assumption of the analogy of nature; that is, past similarities, differences, and regularities are taken to indicate real and persisting structural regularities. This assumption is weaker than that of "natural kinds" related by universal and causally necessary laws, but in order to operate counterfactual predictions, it does of course have a modal component (see LAWS OF NATURE). This may be expressed as: "If a number of objects were found to be more similar than different in specific respects, they would justifiably be expected to be more similar than different in other respects." This is the basis of theorizing with idealized models when these are applied to the real world in a series of analogical steps. For example, analogy takes us from the initial state of a falling body in air to the concept of a sphere falling in a vacuum, with its lawlike initial and final states, and then by analogy to the (approximate) final state of the real body. Similar counterfactural arguments are required for exploration and application of all models which are hypothesized but not necessarily assumed to exist.

A good test of the analogical conception of theories (ACT) is provided by quantum physics. This has always been a difficult case for model theory, because it is generally accepted that no familiar mechanical (or any other) models are adequate interpretations of its formalism. The so-called Copenhagen Interpretation takes a robustly positivist view, according to which the essence of quantum theory is its mathematics, for which no consistent and comprehensive analogies with other physical processes can, or need, be found. Realists, on the other hand, continue to look for "hidden variable" models which will restore comprehensive dynamical reality to the theory, though so far without much success. Meanwhile "particle," "wave," and "field" language continues to be used, and physicists have learned to use these partial models piecemeal in appropriate experimental situations, without assuming anything other than analogical relations with reality.

In terms of ACT, none of this should be surprising. ACT argues only for the reality of certain formal and material analogies in nature. This does not imply any uniquely "true" models of reality, and the history of quantum theory shows that it need not imply that we can articulate any models at all that are adequate for a given theory and its data. It is ironic that SCT, with its abstract "family of models," was being developed at exactly the same time as it was found that in quantum theory there are insoluble problems in articulating even one comprehensive model for its mathematics. There are, however, piecemeal and mutually conflicting models at various levels of the theoretical hierarchy, and these can be seen to function like any other models in aiding intuition and manipulation, and permitting justified local extrapolations to novel data by analogical inference. Quantum theory therefore provides a strong argument for the adoption of ACT, rather than SCT.

A major problem for ACT remains. Structures of similarity and analogy within theories and between theory and world have been amply illustrated in the historical and philosophical literature, but they are still largely unanalyzed primitives within the new conception. Attempts by Carnap and others to formalize a logic of analogical argument have quickly got lost in a "combinatorial jungle" of relative similarities and differences (see EVIDENCE AND CONFIRMATION). Something other than standard logic is required to do justice to the new metaphysical position, but it must be something that has its own rigor, and preferably more general application than just to the philosophy of scientific theories.

Only one recent approach seems to offer hope along both these dimensions. This is the development in cognitive science of parallel distributive processing (PDP), which has been analyzed from the point of view of philosophy of science by Paul Churchland (1989) (see COGNITIVE APPROACHES TO SCIENCE). A PDP system is itself a model (though not yet a quite satisfactory one) of human and animal brains – indeed, of any system that learns economically from experience. To take Churchland's simplified example, suppose the problem is to discriminate between sonar reflections from mines and from rocks as received by ships at sea. The input terminals of the PDP system are presented with vector sets specifying discriminating features of mines and rocks. These pass in parallel to a level of "hidden units" along pathways which weight the input in variable ways. The system can be "taught," by a complex hidden network of feedbacks, to build up prototype profiles of "mine" and "rock" respectively, in such a way that differential responses are made and corrected at the output. Eventually these responses are triggered off appropriately without explicit correction from the teacher when new data are fed in. Even the learning phase does not necessarily require a human programmer, but can be conceived as the result of natural processes of feedback, such as conditioned response to danger, or Darwinian selection.

Tests of the system exhibit speedy and successful learning, but its philosophical interest lies rather in the learning principles presupposed. These show it to be an excellent model of models of scientific theorizing, as this is construed in ACT. The principal virtue of PDP is that it models the process of analogical classification much more faithfully than any previous models in logic or probability theory. It does this simply by building in the assumption (similar to that of Wittgenstein's family resemblances) that perception, discrimination, and successful extrapolation naturally take place by clustering objects and properties with sufficient similarities for our purposes, and distinguishing clusters from one another according to differences with respect to our purposes (Hesse 1988).

To summarize: models have been discussed in philosophy of science from two opposing points of view. The "standard" approach – for example, the semantic conception of theories, is formal and ahistorical, defining a model as one of the entities and processes that satisfy the formal axioms of a theory. The theory itself consists of its formal structure plus the family of all its models. Realist versions of SCT strive to define a "good" theory as one whose models can be taken approximately to represent the real world. Anti-realist versions regard the models as fictions having no direct relation to reality, but to be used purely heuristically for the discovery and explanation of phenomenal laws. Both realist and anti-realist versions of SCT tend to analyze theories as static "textbook" entities, and both tend to make a sharp distinction between a theory with its models and data derived from observation and experiment. SCT consequently neglects epistemological problems of theory development and theory choice.

The alternative approach has been called here the "analogical conception of theories." According to this view, theories are historically changing entities, and consist essentially of hypothetical models or analogues of reality, not primarily of formal systems. Theoretical models, models of data, and the real world are related in complex networks of analogy, which are continually being modified as new data are obtained and new models developed. Analogies with familiar entities and events introduce descriptive terms for theoretical concepts, by processes similar to the use of metaphor in language. Inferences within theories, and from theory to data and predictions, are analogical rather than propositional. Their justification must be sought in some metaphysical principle of the "analogy of nature," a principle that is weaker than the usual assumptions of "natural kinds" or "universal laws." It has been suggested that a suitable philosophical model for the difficult concept of "analogy" may be found in artificial learning systems such as parallel distributive processing.

References

Black, M. 1962: *Models and Metaphors* (Ithaca, NY: Cornell University Press).

Campbell, N. R. 1920: *Physics, the Elements* (Cambridge: Cambridge University Press; subsequently published as *Foundations of Science* (New York: Dover Publications, Inc., 1957).

Cartwright, N. 1983: *How the Laws of Physics Lie* (Oxford: Clarendon Press; New York: Oxford University Press).

Churchland, P. M. 1989: *A Neurocomputational Perspective, the Nature of Mind and the Structure of Science* (Cambridge, MA: MIT Press).

Giere, R. N. 1988: *Explaining Science, a Cognitive Approach* (Chicago and London: University of Chicago Press).

Gooding, D. 1990: *Experiment and the Making of Meaning* (Dordrecht: Kluwer Academic Publishers).

Harré, R. 1986: *Varieties of Realism* (Oxford: Blackwell).

Hesse, M. B. 1966: *Models and Analogies in Science* (Notre Dame, IN: University of Notre Dame Press).

—— 1974: *The Structure of Scientific Inference* (London: Macmillan).

—— 1988: Theories, family resemblances and analogy. In *Analogical Reasoning*, ed. D. H. Helman (Dordrecht: Kluwer Academic Publishers), 317–40.

Putnam, H. 1963: "Degree of confirmation" and inductive logic. In *The Philosophy of Rudolph Carnap*, ed. P. A. Schilpp (La Salle, IL: Open Court; London: Cambridge University Press), 761–83.

Suppe, F. 1989: *The Semantic Conception of Theories and Scientific Realism* (Urbana, IL, and Chicago: University of Illinois Press).

Suppes, P. 1960: A comparison of the meaning and uses of models in mathematics and the empirical sciences. *Synthese*, 12, 287–301.

van Fraassen, B. C. 1980: *The Scientific Image* (Oxford: Oxford University Press).

—— 1989: *Laws and Symmetry* (Oxford: Clarendon Press).