PROGRAMS, THEORIES, AND MODELS

Paul Thagard

Cognitive Science Center, University
of Michigan, Ann Arbor

University of Michigan-Dearborn

April, 1982

This paper makes use of the philosophical literature on theories and models to develop an account of the role of AI programs in psychological theorizing. It is often said that programs *are* theories (e.g. Winston 1977, p. 258). I argue that programs do *not* constitute theories or models in any precise sense, but that the important contribution of programs to psychological theory can be described by adopting a new conception of theories as *definitions* of kinds of systems, developing a cognate conception of model, and interpeting AI programs as simulations of models which approximate to theories.

A program - a set of instructions which a computer can follow - is clearly not a theory according to what used to be the standard philosophical view that theories are sets of sentences axiomatized in a formal system (see e.g. Hempel 1965, pp. 182-183). However, a more plausible interpretation of theories is available.

The alternative conception of scientific theories was originally proposed by P. Suppes (1957, 1967) and has been developed by various authors and applied to fields as diverse as physics, biology, and economics (see e.g. Sneed 1971, F. Suppe 1972, 1977; van Fraassen 1970, 1972; Stegmueller 1976, 1979; Beatty 1980; Hausman 1981). It has been variously referred to as the "semantic" conception and the "structuralist" view of theories. There are important differences among these various accounts, but in what follows I shall eclectically adapt whatever features of the different formulations seem best to apply to cognitive science. In order to avoid confusion, I shall simply refer to the "new" conception or account of theories.

Whereas the traditional view of theories took them to be sets of sentences in an axiomatic system, the new account takes a theory to be a kind of definition. In Suppes' original account, a theory was a definition of a set-theoretic predicate, but for present purposes I shall employ a simpler version of the new account due to Giere (1979). For Giere, a scientific theory is a definition of a kind of natural system (p. 69). He illustrates his account by applying it to the theory of Newtonian mechanics. On the traditional view, this theory might be taken as consisting essentially of Newton' three laws of motion plus the law of universal gravitation. On Giere's view, Newtonian theory is a definition of a kind of particle system: "A natural system is a classical Newtonian particle system if and only if it is a system of objects satisfying Newton's three laws of motion and the law of universal gravitation." (p. 69). As a definition, such a theory is neither true nor false: in itself, it makes no empirical claim. However, it can be used to make empirical claims, for example that the solar system is a system of the kind defined by the theory. Giere calls such claims "theoretical hypotheses", but I shall term them simply "theoretical claims". A theoretical claim has the form: real system R is a system of the kind defined by the theory T.

Whereas a program is clearly not a set of sentences comprising a theory on the traditional view, it is very tempting to think of a program as specifying a kind of cognitive system and hence as qualifying as a theory on the new conception. For example, Kosslyn's imagery programs might be understood as specifying a kind of system for processing information using mental images. John Anderson's programs define a different sort of processing system, oriented around propositions. In either case, we might make the claim that the real human information processing system is a kind of system specified by the program. Such a claim can be empirically evaluated.

A program implicitly characterizes a processing system by specifying what knowledge structures are to be used and what procedures are to operate on them. Although this makes it appealing to say that a program can be a theory according to the new conception, there are two important reasons for resisting the appeal. First, although a program can loosely be said to "characterize" a processing system, it can not be said to *define* a system in the way required by the new conception of theories. Second, we would never want to make the theoretical claim that any real system is just like the system produced by the program, since any program contains a host of implementation-dependent characteristics which we know to be extraneous to real human cognition.

To handle the latter difficulty, I want to develop the concept of a *model*. This is a dangerous choice of term, since "model" has been used with even more ambiguity and vagueness than has "theory". However, the term "model" is often used in cognitive science in much the way I want to define it, and I hope to give a definition sufficiently precise to distinguish models from theories.

As Giere and others have pointed out, "model" and "theory" are sometimes used synonymously, but I think we can outline two features which generally distinguish models from theories. (Cf. Kosslyn 1980, Pylyshyn 1978.) First, models are intended only to have analogies with real systems; they are not expected to characterize them with complete accuracy (cf. Hesse, 1963). Second, models are often intended to have a relatively narrow range of application: we can have models for specific phenomena, whereas theories are usually intended to have wide generality. I shall now show how these features of models can be characterized within the general framework of the new conception of theories. We will still not be able to say that a program *is* a model, but the account of models will bring us closer to describing the role of programs in model building and theory construction.

On my interpretation models are like theories in being definitions of a kind of system, and so are in themselves neither true nor false. However, as indicated above, we expect models to include in the definition of a kind of system features which we would not attribute to real systems. Models define systems which we know not to be exactly like

real world systems. Accordingly, the claims which models are used to make must be different from the claims which theories are used to make. Recall that theories are used to make theoretical claims that a real system *is* a system of the kind defined by the theory. Since a model contains specifications which are known to be false of the target real systems, it can not successfully be used to generate such theoretical claims. For example, a processing model based on the computer metaphor may define a kind of system in which processing is serial, even though the theorizer believes that processing in the brain is parallel. That discrepancy would be enough to defeat any theoretical claim which said that the brain is a processing system of the kind described in the model. We need to be able to use the model to make a weaker claim.

As Hesse (1963) and Kosslyn (1980) have pointed out, the relation between a model and what it models is one of *analogy*. We do not assume that a model exactly describes the target phenomena, only that the phenomena are in important respects *like* what is described in the model. Under the new conception, we can say that a model defines a kind of system, but that we only expect the systems so defined to be analogous to real systems. Hence instead of a theoretical claim we use a model to make what I shall call a "modelling claim", which has the form: a given real system R is very much like the kind of systems defined by model M. This is clearly less precise than the identity claim made in a theoretical claim.

Models are thus less ambitious than theories. Not only do they include in their definitions characteristics which real systems are not expected to have, they are likely to define a narrower set of characteristics than would a theory, which would be expected to give a more complete account of the behavior of a system. Theories are also expected to apply generally to a number of different kinds of systems, whereas models can be either general or specific (Kosslyn 1980). A general model of cognitive processing is one which would be like a theory in having numerous applications, generating numerous modelling claims. But models, unlike theories, can be specific in that they are intended to apply only to a particular sort of system, and a modelling claim is made only about that kind of system. Construing models as definitions of kinds of systems is clearly compatible with both their general and specific uses.

All this has been preparatory to asking the central question: are computer programs psychological *models*? Since models differ from theories in admitting unrealistic characteristics as part of their system definitions, it is tempting to construe programs at least as models of human information processing. But the second impediment remains: a computer program may exemplify a system, but it does not define a kind of system, and therefore can not qualify as a model in the precise sense developed above. Still, we continue to get closer to being able to specify the role of programs in the construction of psychological theories and models.

Zeigler (1976) usefully distinguishes between a real system, a model, and a computer, and says that whereas the relation between the model and the real system is one of *modelling*, the relation between the computer and the model is one of *simulation*. A computer simulates a model which models a real system. Indirectly, then, we can say that a computer is a simulation of a real system. Zeigler's notion of model is different from the one discussed here, but his basic distinctions can be translated into the terms of the current discussion.

When a program is run on a computer, the computer is a simulation of a system. In particular, the system simulated is intended to be a system of the kind defined by the model. A model defines a kind of system, and the program, when executed, performs like a system of the sort defined. The program thus embodies many important features of the model. Hence a program can be used indirectly to make claims about the real system about which a modelling claim is made. Since the program simulates a system of the kind defined by the model, and since the model can be used to make the claim that the real system is a system of the kind defined by the model, we can use the program to make a *simulation claim*: the real system R is analogous to the system S simulated by execution of program P. In short, a simulation claim can have the form "program P simulates R". However, it must be kept in mind that the claim in both these forms is shorthand for a description of a much more complex relation involving models as definitions of systems. In sum, a program can not be said to be a theory or a model, but provides, when executed, a simulation of a system of a kind defined by a model which approximates to a theory.

REFERENCES

Anderson, J.R. (1976), *Language, Memory and Thought*, Hillsdale, New Jersey: Erlbaum Associates.

Beatty, J. (1980), "Optimal-Design Models and the Strategy of Model Building in Evolutionary Biology," *Philosophy of Science 47*: 532-561.

Giere, R. (1979), *Understanding Scientific Reasoning*, New York: Holt, Rinehart and Winston.

Hausman, D. (1981), *Capital, Profits, and Prices: An Essay in the Philosophy of Economics*, New York: Columbia University Press.

Hempel, C.G. (1965), *Aspects of Scientific Explanations*, New York: The Free Press.

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

Kosslyn, S. (1980), *Image and Mind*, Cambridge: Harvard University Press.

Moor, J. (1978), "Three Myths of Computer Science," *British Journal for Philosophy of Science 29*: 213-222.

Pylyshyn, Z. (1978), "Computational Models and Empirical Constraints," *Behavioral and Brain Sciences 1*: 93-99.

Suppe, F. (1972), "What's Wrong with the Received View on the Structure of Scientific Theories?" *Philosophy of Science 39*: 1-19.

Suppe, F. (1977), *The Structure of Scientific Theories* (second edn.), Urbana: University of Illinois Press.

Suppes, P. (1957), *Introduction to Logic*, New York: Van Nostrand.

Suppes, P. (1967). "What is a Scientific Theory?" in S. Morgenbesser (ed.), *Philosophy of Science Today*, New York: Basic Books, 55-67.

van Fraassen, B. (1970), "On the Extension of Beth's Semantics of Physical Theories," *Philosophy of Science 37*: 325-339.

van Fraassen, B. (1972), "A Formal Approach to the Philosophy of Science," in Colodny (ed.), *Paradigms and Paradoxes*, Pittsburgh: Pa.: University of Pittsburgh Press, 303-366.

Winston, P. (1977), *Artificial Intelligence*, Reading, Mass.: Addison Wesley.

Zeigler, B. (1976), *Theory of Modelling and Simulation*, New York: Wiley.