Journal Title: Mind, language, and society
Volume: Issue: 1984
Pages: 77-84
Article Author: Paul Thagard
Article Title: Computer Programs as Psychological Theories

FOR RESEND REQUESTS:
Fax or Ariel this sheet within 5 days

NOTICE: This material may be protected by Copyright Law. (Title 17 U.S. Code)
COMPUTER PROGRAMS AS PSYCHOLOGICAL THEORIES

Paul Thagard

One of the primary methodological tools of cognitive science is the use of computer programs to simulate aspects of intelligent behavior. It is widely agreed that the writing and running of complex programs is a valuable aid to the construction and testing of theories and models of human thinking. Given the view of human thought as information processing which dominates current work in cognitive psychology, programs are a natural and powerful aid to theorizing in that field.

But there is much confusion concerning the exact role of programs in psychological theorizing. One hears it said that programs are theories, as by P.H. Winston: "Occasionally, after seeing what a program can do, someone will ask for a specification of the theory behind it. Often the correct response is that the program is the theory". Programs are also sometimes said to provide models of psychological processes. Occasionally the strong claim is made that without a program one does not even have a theory. Entirely vague in these discussions is what is meant by 'theory' and 'model', and how programs can comprise theories or models.

This paper makes use of the philosophical literature on theories and models to develop an account of the role of programs in psychological theorizing. Two conceptions of theories are considered: the traditional syntactic one which takes theories to be sets of sentences, and a newer semantic conception which takes theories to be definitions of kinds of systems. I argue that on neither conception can a program be said to be a theory. Nor can a program be said to be a model in the sense of 'model' most consonant with the two conceptions of theories. However, the semantic conception of theories as definitions is compatible with an account of the relation of programs to theories and models, and such an account is developed. Roughly, a program is a simulation of a model which approximates to a theory.

1. Theories as sets of sentences

A decade or two ago, the "received view" in philosophy of science on the nature of scientific theories was that a theory is a set of sentences, ideally given an axiomatic formalization in a logical system (e.g. Hempel, 1965: 182-183). Such a theory is tested by deriving observation sentences from it and checking to see whether the expected observations occur. As a set of sentences, a theory can, when interpreted, be said to be true or false.

A program is clearly not a theory in this sense. A program is "a set of instructions which a computer can follow [...] to perform an activity" (Moore,


2 Although the semantic conception is useful for discussing the relation of computer programs and psychological theories, no unqualified endorsement is intended here. An alternative pragmatic conception of theories is developed in Thagard (forthcoming).
1978: 214). Such instructions are not intended to be true or false, nor do they in themselves serve to make predictions about observable phenomena. Descriptions of routines in a program are not intended to refer: their function is to guide processing. The pages of code which comprise a complex program are therefore very unlike the relatively small sets of sentences which make up a theory such as Newtonian mechanics, whose main components are the three laws of motion and the principle of gravitation. A formula such as \( f = ma \) can be construed as giving a description of the world, unlike any set of lines in a program. Since programs are not sets of sentences, and do not serve directly for the deduction of observable consequences, they are not theories in the traditional sense.

Commensurate with the logistic, sentential view of theories is an account of models as set-theoretic interpretations of the sentences in a theory. If a theory is a set of sentences in an uninterpreted formal system, then a system of things, or statements about those things, which provide an interpretation of the sentences, can be said to be a model of the theory (Nagel, 1961: 95). This is close to the sense of model used in mathematical logic. Although the precision of this model is attractive, it can be argued that scientific models are more complex than this formal account allows (Achinstein, 1968). In any case, it is evident that programs are not models of the sort just described. A program is not in itself a system of things, nor does it serve to provide an interpretation for anything. Hence on the traditional view programs are no more models than they are theories.

Of course, the fact that programs are not theories or models on one philosophical account of these notions does not show that programs are not theories or models on some more sophisticated account. And the traditional account of theories and models has received substantial criticisms (for surveys, see Suppe, 1972, 1977). I shall now consider an alternative conception of theories and models which more plausibly applies to programs in cognitive science.

2. Theories as definitions

The semantic conception of scientific theories was originally proposed by Suppes (1957, 1967) and has been developed by various authors and applied to fields as diverse as physics, biology, and economics. It has also been referred to as 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.

Whereas the traditional view of theories took them to be sets of sentences in an axiomatic system, the semantic account takes a theory to be a kind of definition. In Suppes' set-theoretic predicate, both versions of the semantic account, scientific theory is a definition. In his account by applying it to the traditional view, this theory is Newton's three laws of motion. In his (1979: 69) view, Newtonian system: "A natural system is one if it is a system of objects under the law of universal gravitation, or not: in itself, it makes no sense to speak of empirical claims, for they are defined by the theory, but I shall term them simply 'real system' or in the form: real system R is..."

Whereas a program is defined on the traditional view, it is not a definition of a kind of cognitive system or a semantic conception. For example, a program is not "defined" as a kind of object or image. John Anderson's program system, oriented around programs, is characterized by the real human information specified by the program. Such a program is defined.

A program implicitly characterizes a what knowledge structure asks questions on them. Although this makes no sense of a theory according to the sender, it is for resisting the appeal. Programs "characterize" a processing the way required by the sender, but in no way want to make the theory a system produced by the program, a characteristic which we know to consider these objections in a system.

First, we can at best define a system using a strict sense of system if only if certain theoretical claims are met as a theory, such as qualification of a kind of system it is a quibble. It might be argued that the strict sense required for long way in saying what a program. I propose to have a modified thing that real systems is one would say that a real system is described in the program. Such claims can be made "theoretical claims", nor do they have a weak sort of claim constituting definition. This is clearest if we consider taking programs to be theories.
be true or false, nor do observable phenomena. Thus to refer: their function comprises a complex program of sentences which make in components are the three formula such as *f = mz* valid, unlike any set of sentences, and do not sequences, they are not new of theories is an account of the sentences in a theoretically formal system, which provide an interpretation of the theory (NAGEL, 1961 in mathematical logic). It can be argued that account allows (ACHINERGMAS are not models of the system of things, nor does it. Hence on the traditional stories or models on one philosophical programs are not theories. And the traditional adoption of an alternative conception to programs in cognitive is was originally proposed by various authors and applied to science. It has also been used. There are important differences between I shall consider eclectic interpretations seem best to apply; them to be sets of sentences: 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 semantic account due to GIERE (1979). For GIERE (1979: 69), a scientific theory is a definition of a kind of natural system. 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's three laws of motion plus the law of universal gravitation. On GIERE's (1979: 69) 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." 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 *A* 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 a program as specifying a kind of cognitive system and hence as qualifying as a theory on the semantic 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 semantic conception, there are two important reasons for resisting the appeal. First, although a program can loosely be said to "characterize" a processing system, it cannot be said to define a system in the way required by the semantic 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 characteristics which we know to be extraneous to real human cognition. I shall consider these objections in turn.

First, we can at best say that a program simulates a system: it does not define a system using a structure which says that something is a certain sort of system if and only if certain conditions hold. A program which simulated the operation of classical particle systems such as the solar system would not qualify as a theory, since it does not provide the sort of general definition of a kind of system we saw above in Giere's example. This is more than a quibble. It might be argued that although programs are not definitions in the strict sense required for a Giere style theory, they obviously do go a long way in saying what a processing system is supposed to be like. We might propose to have a modified sort of theoretical claim, which, instead of saying that a real system is of the kind defined by a theory in Giere's sense, would say that a real system is of the kind just like the processing system described in the program.

Such claims can be made, but I do not think we should consider them to be "theoretical claims" nor does the fact that programs can be used to make the weak sort of claim constitute an elevation of a program to theory status. This is clearest if we consider the second main reason for not wanting to take programs to be theories in the semantic sense. The empirical claim that
a real system is just like the processing system characterized in the program
is one we would never want to defend: the program will always contain all
sorts of characteristics which we know are extraneous to real human cogni-
tion. Such characteristics derive from features of the programming language
and implementation which have nothing to do with the account we want to give
of human cognition. For example, if our program is written in LISP, it con-
stitutes a series of definitions of functions. The purpose of writing those
functions is not to suggest that the brain actually uses them, but to simu-
late at a higher level the operation of more complex processes such as image
or proposition processing. All this is obvious, but it should suffice to show
that we do not intend to use a program to make a theoretical claim of the
kind discussed by Giere. And since programs are not strictly definitions, it
seems pointless to try to hang on to the view that programs are theories in
the semantic sense. We need a richer set of notions to indicate how to des-
cribe the uses of programs within the framework of the semantic conception
of theories.

3. Models and simulations

To handle the difficulty that programs contain all sorts of extraneous
details not meant to characterize a processing system, 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 distin-
guish 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, 1966). Sec-
ond, models are often intended to have a relatively narrow range of applica-
tion: we can have models for specific phenomena, whereas theories are usually
intended to have wide generality. I shall now show how these features of mod-
els can be characterized within the general framework of the semantic con-
ception of theories. We will still not be able to say that a program is a mod-
el, but the account of models will bring us closer to describing the role of
programs in model building and theory construction.

Following HAUSMAN (1981), I shall take models to be like theories (on the
semantic conception) as being definitions of kinds of systems. Hausman uses
this notion of models to provide an interesting and cogent interpretation of
the significance of economic models, and much of his analysis can be carried
over to the case of psychology. However, whereas Hausman uses 'model' to mean
essentially what is meant by 'theory' in the semantic sense, I want to retain
that notion of theory, and use 'model' to refer to a somewhat different sort of
definition of a kind of system. Models, then, 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 defi-
nition of a kind of system features which we would not attribute to real sys-
tems. 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 dif-
ferent 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 about the computer metaphor may de-
rial, even though the theory of parallel might suggest that the brain is a pro-
We need to be able to use this
to explain a model.

As HESSE (1966) and KOS-
model exactly describes the
important respects like what a
conception, we can say that we
expect the systems so de
instead of a theoretical claim
"modelling claim", which has the
the kind of systems defined by
theory claim made in a theoret-
effectiveness by speaking of an
ystems defined by the model
not often attained in cogni-
like real systems in some respect
"analogical". And those respec-
ting the systems defined by
have seen that all models are
sims they are intended to
extended positive analogy, an
models and defined systems which it

Models are thus less and
in their definitions character-
have, they are likely to be a
theory, which would be ex-
havior of a system. Theories
of different kinds of systems
(KOSSLYN, 1980). A general
would be like a theory in how
modelling claims. But models
are intended to apply only to the
claim is made only about the
finitions of kinds of systems in
specific uses.

All this has been pre-
computer programs psychologically
theories for two reasons: the
e not in themselves defini-
ories in admitting unrealisti-
tions, it is tempting to con-
formation processing. But the

5 This is a matter of degree
that even theories involve
terms of ideal notions such
system will precisely be to
relax the strictness of the
the use of idealizations is
to be idealizations or ap
characterized in the program will always contain all us to real human cog- the programming language in account we want to give written in LISP, it con- purpose of writing those uses them, but to simu- processes such as image; it should suffice to show theoretical claim of the strictly definitions, it programs are theories in to indicate how to de- the semantic conception of

n all sorts of extraneous ten, I want to develop the term, since 'model' has han has 'theory'. However, e in much the way I want to ently precise to distin-

and 'theory' are sometimes features which generally 0; PILLSBURY, 1979). First, al systems; they are not cy (cf. HESSE, 1966). Se- ly narrow range of applica-whereas theories are usually r how these features of mo- work of the semantic con- say that a program is a mo- r to describing the role of

to be like theories (on the s of systems. Hausman uses aid cogent interpretation of his analysis can be carried Hausman uses 'model' to mean atic sense, I want to retain c a somewhat different sort are like theories in being oneselves neither true nor els to include in the defi- not attribute to real sys- be exactly like real world used to make must be dif- make. Recall that theories ystem is a system of the s specifications which are 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 (1966) and KOSSLYN (1980) have pointed out, the relation be- tween 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 semantic 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 less precise than the identity claim made in a theoretical claim. We could perhaps give an illusion of exactness by speaking of an "isomorphism" between the real system and the systems defined by the model, but that degree of mathematical precision is not often attained in cognitive models. Systems defined by the model will be like real systems in some respects; HESSE (1966) calls this the "positive analogy". And those respects in which the real systems are known not to be like the systems defined by the model are called the "negative analogy". We have seen that all models are expected to have some negative analogy to the systems they are intended to model. A modelling claim applies only to the intended positive analogy, and ignores the expected discrepancy between real and defined systems which is part of the negative analogy.

Models are thus less ambitious than theories\(^5\). 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*? We saw that programs fail to qualify as theories for two reasons: they contain unrealistic characteristics, and they are not in themselves definitions of systems. 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

---

\(^5\) This is a matter of degree. The above discussion brushes over the problem that even theories involve idealizations. A theory may define a system in terms of ideal notions such as 'point mass', whereas no object in a real system will precisely be a point mass. There may therefore be a need to relax the strictness of theoretical claims. However, models go far beyond the use of idealizations in theories to include items which are known not to be idealizations or approximations of real systems.
exemplify a system, but it does not define a kind of system, and therefore cannot 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 programs 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.

Let us review. A theory is a definition of a kind of system, claimed to apply to real systems. A model is also a definition of a kind of system, but real systems are only claimed to be analogous to the systems defined by the model. A program, when executed, provides a simulation of a system of the kind defined by a model. The program is laden with even more negative analogies than the model, because of the irrelevant details of program construction. Nevertheless, the positive analogy represented in the parts of the program designed to be analogous to real systems provides an important theoretical contribution. Neglecting some of the fine distinctions made above, we can summarize as follows: programs simulate models which approximate theories. A more accurate summary is: a program P, when executed on a computer, provides a simulation of a system of a kind defined by a model M, where M defines systems which are crude versions of the systems defined by a theory T, and the set of systems defined by T is intended to include the real system R.

4. Benefits and dangers of programs in psychological theorizing

So far, I have been concerned to characterize at a very abstract level the relations between psychological theories and computer programs. Now, more practically, I shall briefly discuss the functions of programs in the process of research, and also outline some of the dangers attached to theorizing using programs.

The primary function of programs in psychological research is to provide a vocabulary for making the processing assumptions of theories explicit. On modern programming theory, programs consist of data structures and algorithms which operate on them. Similarly, much current psychological theory discusses mental structures and associated processes. Designing and writing a program
system, and therefore
laced above. Still, we
role of programs in the
real system, a model,
tween the model and the
the computer and the
odel which models a real
 is a simulation of a
from the one discussed
 to the terms of the cur-
t is a simulation of a
ed to be a system of the
system, and the program,
ined. The program thus
 a program can be used
 which a modelling claim
kind defined by the mo-
 im that the real system
 use programs to make a
 the system S simulated by
 can have the form 'pro-
3 that the claim in both
more complex relation
nd of system, claimed to
of a kind of system, but
ystems defined by the
cn of a system of the
en more negative analogs
 of program construct-
in the parts of the pro-
es an important theoreti-
cations made above, we can
approximate theories. A
 on a computer, provides
1 $M$, where $N$ defines sys-
 by a theory $T$, and the
he real system $R$.

**Theorizing**

at a very abstract level
puter programs. Now, more
f programs in the process
ached to theorizing

al research is to provide
 theories explicit. On
structures and algorithms
logical theory discusses
and writing a program

forces the theorizer to be very explicit about what structures and processes
are being postulated.

Once a program is running, it provides a kind of existence proof for the
processing system which is postulated by the psychological theory, showing
that the system is at least a possible one. Ideally, the program should pro-
vide much more: a mechanism for generating and testing the predictions of
the theory. It is not always possible to tell what the consequences of a complex
theory are, and running the program allows you to see the results of all the
interactions of the structures and processes postulated. Thus programs can be
very useful for both the construction and testing of psychological theories.

However, there are several ways in which programs can hurt rather than
benefit theorizing. We should not let our favourite programming system shape
too strongly the nature of the psychological theory postulated, ignoring the
psychological phenomena which are the target of the theoretical exercise. Re-
call the proverb: "To one whose only tool is a hammer, everything looks like
a nail". It is hard to resist the temptation to use the most familiar compu-
tational tools at hand in constructing a simulation, without reflecting upon
whether the most familiar data structures and algorithms are up to the psy-
chological task. Programs should aid in theoretical development, but should
not determine it.

Moreover, the mere fact that a program works should not be construed as
evidence that psychological reality has been captured. A running program is,
as we saw, a kind of existence proof, but experimental tests are required for
judging the degree of fit between the theory and psychological reality. But
so long as computer programs aid rather than determine theorizing, and are
used to assist testing of theories rather than to replace testing, simula-
tions can be a most valuable tool for psychology.

* * * * *
REFERENCES


WINSTON, P.H. (1977): Artificial intelligence. Reading/MA.


THE PROBLEM OF

The main aim of this paper is to deal with some problems connected with Noam Chomsky's notion of "a language purports to be expressive of a language" (Chomsky, 1981). Linguistic grammar is a formal system, and the speaker-hearer who has learned this position faces some serious problems of philosophical importance.

(i) Are linguists just linguists, or do they have a role as philosophers of language? The answer to this question is not clear, especially in view of the attitude of many linguists who seem to think of their work as "pure" science.

(ii) What is the evidence with regard to the linguistic competence of the speaker-hearer? It is widely known that "later philosophy" of Ludwig Wittgenstein (1953)...

* I am grateful to Rainer Brock for his comments on this paper which not only helped me to improve my argument, but also made me aware of the importance of this issue.

1 Ludwig Wittgenstein: 555 paper Wittgenstein's works (listed in the bibliography and section). Except for given. The remarks contained following I will not give more details on what would be preferable.

2 This view is not only put forward by Chomsky (1966) and others with regard to the extension of the notion of a similar assumption, namely, competence which should be included in Section 2 below, some different approach.

3 Another philosophical problem deals with here, viz., the coherence of the system.

4 I hesitate to speak of Wittgenstein's views in his famous book "Laying out the terms," but in any case, and in this context, the following discussion...

5 Kenny (1973) and Winch (1967) have argued that there is no such thing as a language...