University of Oregon

Institute of
Cognitive & Decision Sciences

Why are Formal Models Useful in Psychology?

Douglas L. Hintzman

Technical Report No. 90-10

To appear in:

W. E. Hockley & S. Lewandowsky (Eds.), *Relating Theory and Data: Essays on Human Memory in Honor of Bennet B. Murdock.* Lawrence Erlbaum Associates, publisher

Address Correspondence to: Douglas Hintzman
Department of Psychology
University of Oregon
Eugene, OR 97403-1227

Why are Formal Models Useful in Psychology?

Douglas L. Hintzman

University of Oregon

Abstract

This chapter explores the value of formal (mathematical and computer) models in psychology. Research on factors that have been shown to bias and limit unaided human reasoning is briefly reviewed, and it is noted that psychologists are susceptible to these errors, just as their subjects are. Characteristics of formal models are discussed in relation to such errors, in an effort to identify the ways in which models can and cannot aid scientific thought. Some limitations of the modeling approach are also discussed. It is argued that because psychological models greatly oversimplify the domains to which they are applied, model evaluation is a complex matter. The measure of a model's value lies not in its ability to fit data, but in how much we can learn from it.

When I was a Senior at Northwestern University, I was enrolled in an honors seminar. One of our first assignments was to give an oral report on an article from *Psychological Review.* For my report, I chose a paper by someone at the University of Vermont, named Bennet B. Murdock, Jr. (Murdock, 1960). The paper concerned a method of quantifying the distinctiveness of stimuli that vary along a single dimension. One aspect of the paper was application of the method to explaining the shape of the serial-position curve of serial learning. It was the first attempt I had seen to derive formally, from a priori considerations, an empirical phenomenon of human memory, and I was quite impressed.

Lest Ben be blamed for what I say here or castigated for determining the direction of my career, I should add that my attitude toward the role of formal models in psychology has been shaped my numerous other experiences. As a first-year graduate student at Stanford, I began working on a computer simulation model of paired-associate learning that eventually became the topic of my dissertation (Hintzman, 1968). This work allowed me to experience first-hand the limitations of intuitive reasoning. Time after time, I made changes in the program with the expectation of achieving a particular outcome, only to learn that the revised system did not behave as I had planned. I also recall writing a term paper for a graduate course, presenting a new theory of visual illusions, based on the recently-discovered receptive fields of visual cortical cells. The class instructor was as excited about the theory as I was, until one day he dropped by my office to tell me he couldn't make it work algebraically. I couldn't either. Fortunately the end of the term was past and I already had my "A".

These experiences and others have made me skeptical of unaided, intuitive judgments concerning how specific theoretical assumptions relate to particular empirical results. Other people who work with formal models seem to share this distrust.

The flaw in my account of visual illusions was caught before any real damage had been done. There are several stages where such errors can be detected as an idea makes its uncertain way from private hunch to generally accepted principle: in the investigator's own elaboration and explication of the hunch, in discussions with colleagues, in the reviewing and editorial process, and--if these fail--in published commentary appearing before the idea is embraced by the scientific community as a whole. But an assertion can be so intuitively compelling that it is accepted without close examination. In these cases, it may take a formal model to convince researchers that the assertion is wrong, and even then the belief may be hard to kill. The widespread misconception that serial and parallel processes can easily be distinguished on empirical grounds is one example (see Townsend, 1990). Another is the idea that if two variables interact, then they must affect the same processing stage (see McClelland, 1979, 1988).

A number of experiments have been done in which subjects first learn to classify category exemplars, and then are tested on the exemplars and also on category prototypes which they have not seen before. Classification performance can be higher for the new prototypes than for the old exemplars. Even where this difference is not present initially, it has been reported to emerge over time. The standard interpretation--which I once accepted--has been that a representation of the central tendency of the category is abstracted and stored, and that this representation has a slower forgetting rate than do traces of the exemplars themselves. We now know that a simple model that stores only exemplars can account for such results (Hintzman, 1986; Hintzman & Ludlam, 1980).

Many experiments have been reported in which subjects search for elements such as letters--either in a set committed to memory or in a visual display. Such experiments produce a variety of results: search times may increase linearly with set size, or increase nonlinearly, or not increase at all; and search times on trials when the target is absent may show the same slope or a greater slope than on positive trials. The standard view in cognitive psychology has been that these different patterns require for their explanation different sorts of mechanisms--incorporating either a serial or a parallel search, for example, and either a self-terminating or an exhaustive stop rule. However, formal models show that basically the same mechanisms can produce any of these results (Broadbent, 1987; Townsend, 1990).

Students of memory are currently interested in relationships and comparisons among memory tasks. A popular idea has been that certain patterns of results indicate that two tasks are performed by different memory systems. One such pattern is a functional dissociation, in which a manipulated variable has different effects on the two tasks. The other is stochastic independence displayed by the contingency table relating successes and failures on the tasks. Formal models, however, show that such data patterns are not diagnostic of different systems. A single memory system can predict functional dissociations (Anderson & Reder, 1987; Humphreys, Bain, & Pike, 1989), and two tasks can show stochastic independence even if the same system performs both tasks (Hintzman, 1987; Nosofsky, 1988). (For further discussion of these issues, see Hintzman, 1990.)

Another example from the field of memory concerns how memory for an original event is influenced by the interpolation of conflicting information between the original learning and the test. In a typical experiment, subjects must choose between the original and the interpolated information on a forced-choice recognition test. These subjects appear to display poorer recognition memory than do controls who did not see the interpolated material. The result has been widely interpreted as showing that the inconsistent information either is incorporated destructively into the original memory trace or interferes with its retrieval. However, McClosky and Zaragoza (1985) showed, using numerical examples, that the result is entirely consistent with a simple Markov model that assumes coexistence and noninterference between traces of the original and interpolated events.

I can't resist adding a somewhat different example. A recent textbook on learning has a chapter on sociobiology, which contains the following claim regarding sexual promiscuity: "While adultery rates for men and women may be equalizing, men still have more partners than women do, and they are more likely to have one-night stands" (Leahey & Harris, 1985, p. 287). It is clear from the context that this does not hinge on the slight plurality of women to men (which would make it trivial), and that homosexual partners do not count. I challenge anyone to set up a formal model consistent with the claim--that is, there must be equal numbers of men and women, but men must have more heterosexual partners than women do. (While you are at it, derive the prediction about one-night stands.) An effort to set up such a model could have helped the authors avoid making a mathematically impossible claim.

My general point is that formal models are of proven value in psychology. They can clear up misconceptions and reveal underlying truths that are not obvious at first glance. The typical member of this audience may see the value of modeling as beyond dispute; but this audience is not a representative sample, and many psychologists are quite skeptical about the modeling approach. I propose that we try to understand why--and in what ways--formal models advance our understanding. This may help us increase the efficiency of our science by putting models to better use. My hope in this chapter is to at least provoke some needed thought and discussion on this important but neglected topic.

Some preliminary comments are in order. First, I discuss only explanatory models--the theoretical side of the research enterprise. Formal models of data are used almost universally in psychology, for example in our standard statistical techniques. It might also be worthwhile to ask why models of data are useful (and more widely accepted than the explanatory kind) but I won't do that here. Second, in many people's minds, formal modeling is synonymous with quantitative modeling. However, for reasons that will become apparent later, I want to make a distinction here. Quantitative models, which attempt to account for the precise numerical values obtained in an empirical investigation, represent an important subset of formal models, but the general class is much broader than that. Third, the question arises as to just what the class of formal models includes. Like many concepts, this is a fuzzy one. Diagrams, flow-charts, etc. may or may not qualify as formal models, depending on the extent to which they involve symbols that are manipulated according to definite rules. By restricting the discussion to the clear cases of mathematical and computer models, we can avoid arguing about exactly where the fuzzy boundaries lie.

The following discussion has four parts. First, I list several sources of error in unaided human reasoning; second, I discuss the nature of formal models; third, I attempt to relate models to reasoning errors, to uncover where the advantages of modeling might lie. Finally, I consider the evaluation of formal models, and argue that there are limitations as well as advantages in their use.

Human Reasoning

A growing body of psychological research attests to the flaws and foibles of human thought. Some phenomena that seem directly relevant to errors in scientific reasoning are as follows:

1. Working memory capacity constrains the number of concepts or entities we can manipulate mentally at the same time (e.g., Johnson-Laird, 1983). Chunking, automatization of rules, and external aids such as diagrams can relieve the burden somewhat (Kotovsky, Hayes, & Simon, 1985), but the limitations are still severe. Bruner, Goodnow, and Austin (1962) referred to this problem evocatively as "cognitive strain."

2. Imagining a dynamic system in action may require keeping track of the current states of several variables. Humans have difficulty updating the current values of variables and purging from memory outdated ones (Bjork, 1978).

3. Because memory is content-addressable, similarity is of overriding importance in retrieval. Humans reason by analogy with familiar situations (Nisbett, Fong, Lehman, & Cheng, 1987). We tend to judge likelihood based on ease of retrieval (Tversky & Kahneman, 1973). We are prone to confuse similar concepts and percepts, and even similar-sounding words.

4. Human cognition is fault-tolerant, in that it will come to quick-and-dirty conclusions even when crucial information is missing. People are generally not aware of the extent to which default expectations and objective data have been intermixed in the conclusions they have reached (e.g., Johnson, Bransford & Solomon, 1973).

5. The mapping of meanings to words and of words to meanings is not one-to-one. One consequence is that a verbal argument can maintain apparent coherence while subtly relying, at different points, on different (and possibly conflicting) interpretations of the same or synonymous words. Another consequence is that people may reason differently about essentially the same situation if it is described in slightly different ways ("framing effects"; Tversky & Kahneman, 1981).

6. Humans are biased to accept as true statements that they have encountered frequently before, independently of whether the statements are actually true or false (Hasher, Goldstein, & Toppino, 1977). (Consider the sociobiology example--"males have more one-night stands.")

7. People are better at reasoning about neutral material than about material that is emotionally charged (Lefford, 1946). We tend to base acceptance or rejection of an argument's validity on whether or not we like the conclusion (Janis & Frick, 1943; Lord, Ross, & Lepper, 1979). It seems that researchers like the conclusion, "I was right", and dislike the conclusion, "I was wrong." In one recent experiment, research scientists were asked to review for publication an experimental paper on ESP. By inserting descriptions of results that either agreed or disagreed with the scientists' preconceptions, the experimenter manipulated their evaluation of the experimental method (Koehler, 1989).

8. Once people know something, they find it difficult or impossible to remember what it was like not to know it (Fischhoff, 1975; Fischhoff & Beyth, 1975). This is called hindsight bias, or the "knew-it-all-along" effect, but as applied to researchers it might be called the "that's-just-what-I-would-have-predicted" effect. This tendency can protect researchers against recognizing ways in which their theories are flawed.

9. Humans often treat mere labels or slogans as though they were explanations. Ironically, this is so even when the label itself implies that the phenomenon is unexplained--e.g., UFO and ESP. Examples from psychology include "direct perception" (which sometimes seems synonymous with ESP), and "schema" (often credited with complex powers that are described, but not explained).

10. In hypothesis testing, humans have a confirmation bias, in that they seek information consistent with their favored hypothesis. They tend not to look for data that would disconfirm the hypothesis, or to ask whether an alternative hypothesis might also be consistent with the data (Mynatt, Doherty, & Tweney, 1977; Wason & Johnson-Laird, 1972). The failure to consider other hypotheses, even though they are crucial, has been called "pseudodiagnosticity" (Beyth-Marom & Fischhoff, 1983; Doherty, Mynatt, Tweney, & Schiavo, 1979).

Surely this is only a partial list, but for present purposes it is more than enough. Human reasoning is open to many sources of error. I want to emphasize just one point, which will figure in the arguments I will make: Knowing the "correct" answers, psychologists sometimes chuckle at the errors that subjects in reasoning experiments make. But we are as human as our subjects, and we would be foolish indeed to think that these cognitive limitations don't also apply to us.

Why Formal Models?

Why should psychologists use formal models? One might think that so fundamental a question would be posed and answered as a routine matter in the introductory section of every elementary treatise on mathematical psychology. I have searched widely for such an account, with little success. Bjork (1973) argued that models making quantitative predictions are more easily falsified; but that may be a mixed blessing, for reasons that I discuss later. Townsend and Kadlec (1989) say that psychology needs mathematics because its phenomena are so complex; but do not say why complexity should matter. One might argue that textbooks do not explain the usefulness of formal models because it is obvious; but textbooks say many obvious things, and it is hard to see why something so central would be left out. I have heard psychologists deny that there *are* any good reason for formal models in psychology and claim that modelers are just slavishly (and inappropriately) imitating physics, so the answer must not be obvious. Lacking a clear answer regarding psychology per se, let us step up a level in our conceptual hierarchy and ask why formal models work in any branch of science.

There is a long history of thought about why mathematics is useful in science as a whole. The topic is surveyed rather thoroughly by Kline (1985). The Pythagoreans resolved the mystery by holding that number relationships are the substance and form of nature--thus, mathematics and nature are essentially the same thing. Plato held that reality had been designed according to mathematical principles, so that only mathematics, and not our imperfect senses, can tell us what nature is really like. In the Middle Ages, people didn't think about the problem, because all occurrences in nature were considered acts of God; but Renaissance thinkers held that "God is a mathematician," thus justifying mathematics and science as quests to glorify God. These accounts strike me as woefully inadequate. Deep down, they just say there is a correspondence between mathematics and nature because a correspondence exists. What appears to be a current version of this theme is something called the computational viewpoint of physical processes. "The basic notion here is that the material world and the dynamic systems in it are computers [and] the laws of nature are algorithms that control the development of the system in time, just like real programs do for computers." (Pagels, 1988, p. 45). Claiming that the material world is a computer seems as circular an explanation as saying that nature is number or God is a mathematician.

Another approach has been to view mathematics as a human invention, rather than something having independent existence. Aristotle, in contrast to Plato, saw mathematics as merely descriptive. But this leaves unanswered the key question of why mathematics works. Kant asserted that the mind imposes structure on nature--hence the same entity that creates mathematics creates our perception of nature. This position seems to endow the mind with uncanny coherence (and perhaps an overwhelming confirmation bias). At best, it fails to explain why theories so often are wrong. The dominant modern view seems to be conventionalism. The idea here is that mathematicians invent the mathematical models--of which there are in principle an infinite number--and scientists just pick the models that work best in particular domains. Thus, the correspondence is explained by a kind of Darwinian selection. If a model fits the data we keep it, if not we either modify it or throw it out and try another. The problem with this account is that some mathematical models keep working--not just on observations similar to the ones they were selected to explain, but also on completely novel observations, which confirm long chains of deductions that were never tested before. This is true in the physical sciences, if not in psychology, and it is something that conventionalism seems unable to explain. On occasion, the power of mathematics has been declared inexplicable--for example, by Pierce, Shrödinger, and Einstein (Kline, 1985). Maybe this is why mathematical psychology textbooks don't explain why mathematics works.

It may be useful to characterize briefly what mathematics is. Its essence lies in the concept of proof. A mathematical proof begins with a set of assumptions or axioms represented by strings of discrete symbols, and a set of transformation rules that can be applied to the symbol strings. There is also a theorem or conclusion to be proved, also expressed as a symbol string. The proof consists of a step-by-step demonstration that one can get from the axioms to the theorem by applying the rules. The axioms of a proof must be clearly stated and mutually consistent. According to Davis and Hersh (1981), "The demands of precision require that the meaning of each symbol or each symbol string be razor sharp and unambiguous. The symbol … is perceived in a way which distinguishes it from all other symbols … , and the meaning of the symbol is to be agreed upon, universally" (p. 124). Moreover, in a calculation (e.g., a proof) "a string of mathematical symbols is processed according to a standardized set of agreements and converted into another string of symbols. This may be done by a machine; if it is done by hand, it should in principle be verifiable by a machine" (p. 125). Although an actual published proof will contain many gaps (where the intervening steps are presumably obvious), the implicit promise is that they can be filled in on demand. (These intuitive leaps are where errors are most often found.) The nature of a proof and its central role led Suppes (1984) to characterize mathematics as a "radically empirical" science, because the evidence (the proof) is "presented with a completeness not characteristic of any other area of science" (p. 78).

In short, mathematics has the earmarks of a system for imposing consistency on reasoning. Indeed, Descartes saw in Euclid's geometry a way to perfect human reasoning: An argument was to be broken down into steps so small that none of them could be doubted. Contrary to Kant, Suppes (1984) says, "The certainty we find in mathematics arises not from any intuitive or a priori consideration but simply from the discreteness and easily exhibited character of the evidence offered in support of a particular (empirical) claim" (p. 79). In a computer simulation, the steps are those of the algorithm being computed, which can be examined in a print-out of the program. We can be virtually certain that the program is being followed consistently because it is being run on a (reliable) machine.

If the essence of mathematics is consistency, as I claim, how does that help explain why mathematics works? At root, the answer may be simply that reality is consistent, too. This is, of course, a fundamental assumption of science. At the deepest level, nature's consistency presumably derives from there being only a few types of elementary particles and forces behind all phenomena in the universe. In higher or more complex domains like psychology, consistency derives from similarities within the classes of objects studied--such as human brains. This view seems to make the power of mathematics explicable while retaining the basic assumptions of conventionalism. Consistency implies an underlying redundancy in causal mechanisms, even when surface manifestations appear quite distinct. Thus a mathematical model that we have selected and retained because it mimics a range of phenomena in a particular domain has a good chance of succeeding on new phenomena in that domain. Although the surface manifestations may appear new, the underlying sources of redundancy have not changed. A one-to-one mapping between constructs of the model and entities in the world should not be required for this to work, although presumably it would help.

Trying this argument out on colleagues, I have been accused of contradicting myself. I said that humans are inconsistent in their reasoning, and that nature is consistent, but also that humans are part of nature. How can humans be inconsistent, then? I think this objection displays the problem of multiple meanings, which I listed earlier as reasoning problem #5. Humans are consistent, in that human reasoning--like all natural phenomena--shows regularities. These regularities might be captured in a formal model of human reasoning; but a successful model probably would not assume that humans reason logically--that is, in such a way as to avoid contradictions in the <u>contents</u> of their beliefs. The objection, I think, confuses two meanings of "consistent"--one applying to the laws that govern thinking, and one to what the thoughts are about.

Formal Models and Sources of Error

Let us now consider how the characteristics of formal models relate to the problems with human reasoning that were listed earlier.

1. Working memory problems are largely alleviated by executing steps one at a time. In mathematical modeling, intermediate results can be recorded on notepads and

consulted when needed; in computer simulation, they are held in memory as long as required.

2. Updating is likewise not a problem, particularly if a computer is used. The ability to keep track of the current values of large numbers of variables that are continually changing makes computers useful in simulations of all kinds of complex systems--e.g., weather patterns, national economics, and military encounters.

3. Symbols are discrete, so even conceptually similar entities should not be confused. Different versions of a concept are distinguished by subscripts or superscripts, in a mathematical model, and by storing the information in different memory locations, in a computer.

4. Default expectations may play a significant role in inventing a mathematical proof or in devising a program to accomplish some goal. But the mathematical method is designed to catch any steps in the proof itself that are not explicitly supported by the axioms and the rules. In a simulation model, if crucial information is missing or garbled, the program usually will not run.

5. The requirement that symbols be clearly defined and separately identified within a formal system helps to eliminate contradictions. The consequences of a contradiction (e.g., proofs of both P and not-P) can be unambiguously identified in a mathematical argument, and circular references are flagged as errors in a computer program. Likewise, the requirement that a problem be formally stated can help to eliminate framing effects. (The question of which is the "correct" formulation is empirical, however, as the conventionalist view of mathematics suggests.)

6-7. Familiarity, emotionality, and agreement or disagreement with the conclusions should play no role in a formal deduction, per se. They can, however, powerfully influence one's starting assumptions and what one tries to prove. (The mathematician, Kurt Gödel, is said to have devised a proof of the existence of God [Pagels, 1988].) Such factors can also influence whether one looks for a bug in a program or an error in a mathematical argument. These are good reasons for researchers to write their own versions of programs and to check each other's proofs.

8. With a formal model, hindsight can be rigorously checked. That is, if the theory was explicitly stated to begin with, then "postdictions" and predictions should be derivable from the axioms in exactly the same way. All theorems are implicit in the axioms, regardless of whether they or the relevant data were realized first.

9. The requirement that one derive observations from deeper assumptions immediately exposes labels and slogans as devoid of explanatory power. Formal models would be an invaluable aid to thinking for this reason alone.

10. Where confirmation bias is concerned, the value of formal models is not so clear. On the positive side, a formal model can force one to recognize that one's assumptions are inconsistent with an empirical outcome. Moreover, experience with models may help one realize that the connection between theoretical assumptions and behavior are sometimes nonobvious, so that alternative explanations should be considered. Simulation models, in particular, can help one to develop new intuitions about the behavior of systems having properties such as variability, parallelism, and nonlinearity. These are certainly properties of the brain, and they are inherently hard to understand. Models have heuristic value, in that experience with several model systems can help one anticipate how new combinations of assumptions are likely to interact.

But in other respects models can magnify the confirmation bias. Model building can take an enormous amount of intellectual work, and so modelers have a greater stake than other theorists in avoiding disconfirmation. The fear of rejecting one's model manifests itself as "conservative focusing" (c.f. Bruner, et al., 1962). Technically speaking, the model is being exposed to possible falsification, but rather than testing predictions that seem unlikely on a priori grounds, the modeler chooses to only slightly modify the experimental conditions under which the model has already shown success. However well such a strategy may serve to further one's career, it seems inimical to scientific progress. If the purpose of research is to discard mistaken ideas and replace them with better ones, then the sooner we recognize our errors, the better science we will do. If theorists are reluctant to "go for the jugular", they may need encouragement. It is sometimes suggested that Psychology needs a Nobel prize, but I think that would be a mistake. I propose an award "to the researcher who has most advanced psychological science by admitting error." (It wouldn't have to be given every year.)

As this discussion of confirmation bias suggests, formal modeling has limitations, as well as strengths. Explanatory models formalize the deductive process, but that is only one crucial part of the scientific enterprise. Theorists must be clear and consistent in their assumptions, but beyond that they can make any postulates they want. They can change their postulates if they don't like how they behave--indeed, the "hypothetico-deductive method" dictates that this is how things are done. A theorists can even reinterpret the way the model relates to the world, while leaving the model itself exactly the same. The invention, interpretation, and evaluation of a model are matters completely outside the formal system itself. As such, they are subject to all the weaknesses of human reasoning discussed earlier, as well as to its strengths.

Evaluating Models

The problem of model evaluation raises the question of which predictions are fair game. A scientific model of any complex domain must include some assumptions that are arbitrary but necessary to get the modeling exercise off the ground. For example, mathematical modelers often assume linear or exponential functions, solely because they are mathematically tractable; and simulation modelers employ discrete time steps because that is how computers work. In my MINERVA 2 model, stimulus items and memory traces are represented as random vectors of +1's, -1's, and 0's. Such assumptions are not the focus of interest. They are adopted for their familiarity, tractability, and ease of implementation, and because you have to be explicit about everything to have a system that works. Now an important question is, if you are going to "go for the jugular," do you focus on a prediction that crucially depends on these arbitrary assumptions, rather than on what the theorist considers the central ones? I see tests of the list-strength effect, predicted by my model and others, as important (Ratcliff, Clark & Shiffrin, 1990); but if someone were to show empirically that stimuli cannot be random vectors, I wouldn't be especially impressed. The question of which predictions to test can be difficult to answer, however, because the line between focal and arbitrary assumptions is fuzzy, and it is often unclear to what extent different assumptions contribute to a particular prediction that a model makes.

The explicit recognition that psychological models contain assumptions that are arbitrary has some implications that have been largely overlooked. One is that assessing a model's ability to account for precise, quantitative features of the data may often be difficult to justify. I think quantitative data fitting has a legitimate place in the modeling enterprise--particularly when the core assumptions of a model lead to definite quantitative predictions, independently of the arbitrary ones. But data fitting sometimes appears to be carried out almost as a ritual, or as an end in itself. I have read statements like this: "Although the goodness-of-fit statistic was significant ($p<.01$), we are generally quite pleased with the model's account of the data." Such remarks could be taken as manifestations of confirmation bias, but I think they are more an admission that we lack the kind of precision that the data-fitting exercise implicitly assumes.

The observation that scientific models oversimplify their domains is especially true in psychology. In virtually every experimental situation, there are significant sources of variance that our models do not even attempt to capture in a realistic way. These can include such obvious factors as subject and item differences, as well as more subtle ones such as subject-item interactions, practice and fatigue effects, and the evolution of strategies during the experimental session. If these are significant sources of variance in an experiment, and one's model does not take them into account, then why would one expect the precise, quantitative predictions of the model to be correct? This may be an example of the inappropriate imitation of physics. Physicists do not just assume for the purpose of model building that all electrons are tokens of the same type--they *believe* it. By contrast, psychologists do not believe that all subjects or all words are identical, or that strategies never change. I think that when psychologists discount their models' failures to fit data, they are implicitly acknowledging that such precision was not a realistic expectation in the first place. A model could easily fail to fit data from an experiment even though its core assumptions--those that are the focus of interest--are correct. Peripheral assumptions, made only to promote the model's tractability, may be to blame.

Here is an example of how an arbitrary assumption can mislead. Experienced modelers, as well as nonmodelers dabbling in stochastic reasoning, routinely make what I call the *uniformity assumption*. The pool of observations, combined over subjects and items, is used to compute a parameter estimate--say, the probability of recall, $P=.5$. From that point on, it is implicitly assumed that this value applies to each subject and each item individually, as though recall attempts were as uniform as tosses of coins. Put differently, it is assumed that the P's for different subject-item combinations have a variance of zero. For many derivations this assumption does no harm, and it makes the mathematics easy; but suppose a theorist wants to derive the probability of a subject recalling a given item at least once in two recall attempts, assuming that the attempts are independent. The theorist computes $P+P-P^2 = .75$, as the textbooks all dictate. The problem is that this derivation requires a variance of zero. To see this, consider the most extreme case, in which half the subject-item combinations (set A) have $P=0$ and half (set B) have $P=1$, so P still has the mean .5, but the variance is .25. For set A, $P+P-P^2 = 0$, and for set B, $P+P-P^2 = 1$, so the "correct" prediction is .5 instead of .75. The model may be rejected because the predicted value of .75 is too high, even though all the postulates are correct except an implict one--the uniformity assumption. The stark simplicity of this example may make the error seem obvious, but it crops up in the memory literature repeatedly in more disguised forms. Researchers who would never argue explicitly that all subjects, items, or subject-item combinations are the same, sometimes go to great lengths to perform quantitative tests that depend crucially on that being true. Once the assumptions of a model have been laid down, their initial arbitrariness tends to be forgotten--as though the credibility of the model as a whole can be divorced from the credibility of its parts.

So a model can *fail* to fit data even if its core assumptions are right. Are we on firmer ground if we have a model that fits? Consider the astonishing success of the one-element model (Bower, 1961), which fitted numerous statistics from a handful of paired-associates experiments with incredible precision, despite having only one free parameter--the probability of all-or-none learning of a pair on a given study trial. The model made no allowance for confusions among items, and it assumed--implausibly--that the learning rate was the same across all subjects and items and trials. Moreover, a wide range of empirical evidence shows that the model's basic assumption of all-or-none learning is wrong. The one-element model does a good job of fitting simulation data from my Minerva 2 model (Hintzman, 1984), even though the assumptions of the models differ in several significant ways. The lesson I draw from all this is that a model can fit data with impressive precision even though its basic assumptions are wrong.

So far, I have argued that neither failure nor success in fitting psychological data quantitatively is a reliable guide to the truth of a model's core assumptions. This may seem to undermine the whole modeling enterprise, but it simply reaffirms that evaluation is a matter of human judgment. It cannot be reduced to a simple algorithm, such as computing (or comparing) measures of goodness of fit. In evaluating a model, many questions need to be asked: Which assumptions deserve to be taken seriously, and which are arbitraary? To what extent does a particular prediction depend, either quantitatively or qualitatively, on assumptions of these two types? Are the core assumptions of the model plausible? How to they fit with data and theory in related domains? Depending on the particular situation, there are any number of questions that one should ask. Why should this be so complicated? My answer is that the point of modeling is not really to fit data, although modelers often seem to assume that it is. The point is to learn things. In a model, one has an artificial system through which the interrelationships between assumptions and behavior can be explored. By comparing the behavior of such systems with the behavior of subjects, we can confirm, refine and revise our ideas of how mental processes work. Fitting data may aid in this activity or may even distract from it, depending on the specific context in which it is done.

If oversimplification can cause certain problems, the alternative is worse. Computer modelers, in particular, can get carried away with the power and flexibility of the programming medium, as though the goal were to create in the computer a complete duplicate of the human mind. Such models quickly become so unwieldy that no one can tell why they fail or succeed. This completely misses the point of constructing a formal model in the first place. If theories evolve by shedding their bad assumptions and keeping or adding good ones, then they must be simple enough to be understood. Piling ad hoc assumption on top of ad hoc assumption can only impede progress, by obscuring what the core assumptions of the model imply.

Because science is a collective activity, evaluating models also requires good communication. Modelers sometime complain that experimenters continue to draw inferences about underlying mechanisms from supposedly diagnostic patterns in their data long after it has been shown that such inferences are wrong (see Townsend, 1990). Non-modelers complain about wading through technically difficult papers only to discover, after much effort, that the models were irrelevant to their interests, naive, or absurd. The number of experimenters willing to read a Psychological Review paper is inversely related to the number of equations it contains and to the strangeness of the symbols it employs. And the number who will read an article in the Journal of Mathematical Psychology is zero. Our discipline would benefit if experimentalists put more effort into keeping up with developments in modeling, but modelers need to take the initiative, too. They could make more use of diagrams to get basic ideas across, and put more effort into explaining in words what their mathematical expressions mean. If non-modelers were called on to review modeling papers (and modelers expected this), many communication difficulties might be avoided. Highly technical papers written for a mathematical audience might be supplemented with more readable summaries of the central implications in mainstream cognitive journals. In the long run, our science will benefit if students get more experience with formal models, as both producers and consumers; but progress in this direction has been slow.

Concluding Remarks

Broadbent (1987) has complained that standards of precision in theory have not kept pace with those in experimental methodology, and that they may even have regressed:

> Terms are used such as 'access to the lexicon', 'automatic processing', 'central executive', 'resources'; formal definitions of such terms are rare, and even rarer are statements of the rules supposed to be governing their interaction. As a result one is left unclear about exactly what kinds of experimental data would invalidate such theories, and whether or not they are intended to apply to some new experimental situation. (p. 169)

Watkins (1990) recently struck a similar note:

> When a theory does attract criticism, the critic almost always turns out to have misunderstood, and the theory stands as originally proposed. ... On the rare occasion a criticism demands action, fine tuning will almost always suffice. Thus, the chances of a theory having to be abandoned or even appreciably revised as a consequence of criticism are vanishingly small, and hence researchers can be confident that their theories will stay alive just as long as they continue to nourish them. (p. 328)

Interestingly, while these two authors appear to describe the same symptoms, their diagnoses are quite different. Broadbent (1987) argues that researchers should avoid the ambiguity of verbal theorizing by implementing simple models on a personal computer. But Watkins (1990)--noting that most theories are vague and that most theories invoke mediating processes (such as memory traces)--concludes that "the problem is mediationism" (p. 328). Broadbent thinks we need to tighten up our explanatory theories; Watkins thinks we need to get rid of them.

Obviously, I side with Broadbent on this issue, and not with Watkins. But let us look at the examples that Broadbent and Watkins support their positions with. Broadbent (1987) demonstrates that a simple computer simulation model can mimic several quite different patterns of reaction times commonly obtained in visual-search and memory-search experiments--patterns so different that they have been believed to reflect quite different sorts of underlying mechanisms. Watkins (1990) gives his own "cue overload principle" as an example of the sort of empirical laws that can be achieved without postulating mediating processes or states. (This principle says that the more items a cue subsumes, the less effectively it retrieves any one of them.) To show that mediating states are not essential to scientific explanation in general, Watkins cites Newton's law of universal attraction.

Newton, however, used a formal model to provide a deep explanation of many phenomena that, on the surface, appear quite diverse. As Reichenbach (1951) comments:

> The law of gravitation has the form of a rather simple mathematical equation. Logically speaking, it constitutes an hypothesis, which is not accessible to direct verification. It is established indirectly, since, as Newton showed, all the observational results summarized in Kepler's laws can be derived from it. And not only Kepler's results; from Newton's law, Galileo's law of falling bodies is likewise derivable, and so are many other observational facts, such as the phenomenon of the tides in their correlation to the positions of the moon. (p. 101)

Later, Reichenbach directly contrasts the basic approach advocated by Watkins (often called functionalism in psychology) with Newton's method:

> Whoever speaks of empirical science should not forget that observation and experiment have been capable of building up modern science only because they were combined with mathematical deduction. Newton's physics differs greatly from the picture of inductive science that had been drafted two generations earlier by Francis Bacon. A mere collection of observational facts, such as presented in Bacon's tables, would never have led a scientist to the discovery of the law of attraction. Mathematical deduction in combination with observation is the instrument that accounts for the success of modern science. (p. 103)

Certainly, Broadbent's modeling exercise fits Newton's hypothetico-deductive method better than Watkins's empirically induced cue-overload principle does.

It is from this perspective--that we should seek deep explanations of the lawful phenomena we observe--that I am encouraged by the current crop of memory models (e.g., Eich, 1982; Gillund & Shiffrin, 1984; Hintzman, 1986; Hintzman, 1988; Humphreys, et al., 1989; Murdock, 1982; Murdock, 1989; Raaijmakers & Shiffrin, 1980). Rather than focusing on specific tasks, in the style of the mathematical psychology of the 1960's, such models as Murdock's TODAM attempt to characterize the basic properties of a memory system that underlies performance in many tasks. Although existing efforts are only approximations to the ideal, the general approach can be characterized as one of postulating a kind of memorial deep-structure whose principles are manifested in various ways when the system is placed in different task environments. The focus of interest is on understanding how many apparently diverse empirical phenomena can arise from a small set of basic principles. As for the cue-overload principle, the current models all suggest that it may be just one of several implications of the fact that human memory is content-addressable. Through the use of formal models, such conjectures can be given rigorous test.

Author Note

This research was supported by grant BNS-87-11218, from the National Science Foundation. Correspondence should be addressed to Douglas L. Hintzman, Psychology Department, University of Oregon, 97403. Thanks are due to Johnathan Koehler for providing references on biases in human reasoning, and to Debbie Frisch and Tom Carr for comments on an earlier draft of this paper.

References

Anderson, J. R., & Reder, L. M. (1987). Effects of number of facts studied on recognition versus sensibility judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*, 355-367.

Beyth-Marom, R., & Fischhoff, B. (1983). Diagnosticity and pseudodiagnosticity. *Journal of Personality and Social Psychology*, *45*, 1185-1195.

Bjork, R. A. (1973). Why mathematical models? *American Psychologist*, *28*, 426-433.

Bjork, R. A. (1978). The updating of human memory. In G. H. Bower (Ed.), *The Psychology of Learning and Motivation, Vol. 12,* (pp. 235-259). New York: Academic Press.

Bower, G. H. (1961). Application of a model to paired-associate learning. *Psychometrika*, *26*, 255-280.

Broadbent, D. (1987). Simple models for experimental situations. In P. Morris (Ed.), *Modelling Cognition.* (pp. 169-185). London: John Wiley & Sons.

Bruner, J. S., Goodnow, J. J., & Austin, G. A. (1962). *A Study of Thinking*, New York: John Wiley & Sons.

Davis, P. J., & Hersh, R. (1981). *The Mathematical Experience* . Boston: Houghton Mifflin Co.

Doherty, M. E., Mynatt, C. R., Tweney, R. D., & Schiavo, M. D. (1979). Pseudodiagnosticity. *Acta Psychologica*, *43*, 111-121.

Eich, J. M. (1982). A composite holographic associative recall model. *Psychological Review*, *89*, 627-661.

Fischhoff, B. (1975). Hindsight ≠ foresight: The effects of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance*, *1*, 288-299.

Fischhoff, B., & Beyth, R. (1975). "I knew it would happen":--Remembered probabilities of once-future things. *Organizational Behavior and Human Performance*, *13*, 1-16.

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. Psychological Review, 91, 1-67.

Hasher, L., Goldstein, D., & Toppino, T. (1977). Frequency and the conference of referential validity. Journal of Verbal Learning and Verbal Behavior, 16, 107-112.

Hintzman, D. L. (1968). Explorations with a discrimination net model for paired-associate learning. Journal of Mathematical Psychology, 5, 123-162.

Hintzman, D. L. (1984). MINERVA 2: A simulation model of human memory. Behavior Research Methods, Instruments, & Computers, 16, 96-101.

Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. Psychological Review, 93, 411-428.

Hintzman, D. L. (1987). Recognition and recall in MINERVA 2: Analysis of the "recognition failure" paradigm. In P. E. Morris (Ed.), Modelling Cognition (pp. 215-229). London: John Wiley & Sons Ltd.

Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. Psychological Review, 95, 528-551.

Hintzman, D. L. (1990). Human learning and memory: Connections and dissociations. Annual Review of Psychology, 41, in press.

Hintzman, D. L., & Ludlam, G. (1980). Differential forgetting of prototypes and old instances: Simulation by an exemplar-based classification model. Memory & Cognition, 8, 378-382.

Humphreys, M. S., Bain, J. D., & Pike, R. (1989). Different ways to cue a coherent memory system: A theory for episodic, semantic, and procedural tasks. Psychological Review, 96, 208-233.

Janis, I. L., & Frick, F. (1943). The relationship between attitudes toward conclusions and errors in judging logical validity of syllogisms. Journal of Experimental Psychology, 33, 73-77.

Johnson, M. K., Bransford, J. D., & Solomon, S. (1973). Memory for tacit implications of sentences. Journal of Experimental Psychology, 98, 203-205.

Johnson-Laird, P. N. (1983). Mental Models. . Cambridge, MA: Harvard University Press.

Kline, M. (1985). Mathematics and the Search for Knowledge. . New York: Oxford University Press.

Koehler, J. J. (1989). The influence of prior beliefs on scientific judgments of evidence quality. (Manuscript submitted for publication.)

Kotovsky, K., Hayes, J. R., & Simon, H. A. (1985). Why are some problems hard? Evidence from Tower of Hanoi. Cognitive Psychology, 17, 248-294.

Leahey, T. H., & Harris, R. J. (1985). Human Learning . Englewood Cliffs, NJ: Prentice-Hall.

Lefford, A. (1946). The influence of emotional subject matter on logical reasoning. The Journal of General Psychology, 34, 127-151.

Lord, C. G., Ross, L. & Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. Journal of Personality and Social Psychology, 37, 2098-2109.

McClelland, J. L. (1979). On the time relations of mental processes: An examination of systems of processes in cascade. Psychological Review, 86 287-330.

McClelland, J. L. (1988). Connectionist models and psychological evidence. Journal of Memory and Language, 27, 107-123.

McClosky, M., & Zaragoza, M. S. (1985). Misleading postevent information and memory for events: Arguments and evidence against memory impairment hypotheses. Journal of Experimental Psychology: General, 114, 3-18.

Murdock, B. B., Jr. (1982). A theory for the storage and retrieval of item and associative information. Psychological Review, 89, 609-626.

Murdock, B. B. (1989). Learning in a distributed memory model. In C. Izawa (Ed.), Current issues in cognitive processes: The Tulane Symposium of Cognition. Hillsdale, N.J.: Erlbaum.

Murdock, B. B. Jr. (1960). The distinctiveness of stimuli. Psychological Review, 67, 16-31.

Mynatt, C. R., Doherty, M. E., & Tweney, R. D. (1977). Confirmation bias in a simulated research environment: An experimental study of scientific inference. Quarterly Journal of Experimental Psychology, 29, 85-95.

Nisbett, R. E., Fong, G. T., Lehman, D. R., & Cheng, P. W. (1987). Teaching reasoning. Science, 238, 625-631.

Nosofsky, R. M. (1988). Exemplar-based accounts of relations between classification, recognition, and typicality. Journal of Experimental Psychology: Learning, Memory, and Cognition, 14, 700-708.

Pagels, H. R. (1988). The Dreams of Reason . New York: Simon & Schuster.

Raaijmakers, J. G. W., & Shiffrin, R. M. (1980). SAM: A theory of probabilistic search of associative memory. In G. H. Bower (Ed.), The psychology of learning and motivation, Vol. 14. (pp. 207-262). New York: Academic Press.

Ratcliff, R., Clark, S. & Shiffrin, R. (1990). The list-strength effect: I Data and discussion. Journal of Experimental Psychology: Learning, Memory, and Cognition, 16, 163-178.

Reichenbach, H. (1951). The Rise of Scientific Philosophy . Berkeley, CA: University of California Press.

Suppes, P. (1984). Probabilistic Metaphysics, . Oxford: Basil Blackwell.

Townsend, J. T. (1990). Serial vs. parallel processing: Sometimes they look like Tweedledum and Tweedledee but they can (and should) be distinguished. Psychological Science, 1, 46-54.

Townsend, J. T., & Kadlec, H. (1989). Psychology and mathematics. In R. E. Mickens (Ed.), Mathematics and Science, Oxford: Oxford University Press.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. Cognitive Psychology, 5, 207-232.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. Science, 211, 453-458.

Wason, P. C., & Johnson-Laird, P. N. (1972). Psychology of Reasoning: Structure and Content. . Cambridge, MA: Harvard University Press.

Watkins, M. J. (1990). Mediationism and the obfuscation of memory. American Psychologist, 45, 328-335.