

- Nottebohm, F. (1977). Asymmetries in neural control of vocalization in the canary. In S. Harnad, R. W. Doty, L. Goldstein, J. Jaynes, & G. Krauthamer (Eds.), *Lateralization in the nervous system*. New York: Academic Press.
- Nottebohm, F. (1989). From bird song to neurogenesis. *Scientific American*, 260(2), 74-79.
- Passingham, R. E. (1982). *The human primate*. San Francisco, CA: W.H. Freeman.
- Ogden, J. A. (1988). Language and memory recovery after long recovery periods in left-hemispherectomized subjects. *Neuropsychologia*, 26, 645-659.
- Ogden, J. A. (1989). Visuospatial and other "right-hemispheric" functions after long recovery periods in left-hemispherectomized subjects. *Neuropsychologia*, 27, 765-776.
- Passingham, R. E. (1982). *The human primate*. San Francisco, CA: W. H. Freeman.
- Piattelli-Palmarini, M. (1980). *Language and learning: The debate between Jean Piaget and Noam Chomsky*. Cambridge, MA: Harvard University Press.
- Piattelli-Palmarini, M. (1989). Evolution, selection, and cognition: From "learning" to parameter setting in biology and the study of language *Cognition*, 31, 1-44.
- Purves, D., & Voyvodic, J. T. (1987). Imaging mammalian nerve cells and their connections over time in living animals. *Trends in Neuroscience*, 10, 398-404.
- Rakic, P. (1985). Limits of neurogenesis in primates. *Science*, 227, 1054-1056.
- Rumelhart, D. E., McClelland, J. L., & the PDP Research Group. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition. Vol. 1: Foundations*. Cambridge, MA: Bradford/MIT Press.
- Sarich, V. M., & Wilson, A. C. (1967). Immunological time scale for hominid evolution. *Science*, 158, 1200-1203.
- Teyler, T. J., Perkins, A. T., IV, & Harris, K. M. (1989). The development of long-term potentiation in hippocampus and neocortex. *Neuropsychologia*, 27, 31-39.
- Thatcher, R. W., Walker, R. A., & Giudice, S. (1987). Human cerebral hemispheres develop at different rates and ages. *Science*, 236, 1110-1113.
- Tsunoto, T., Hagihara, K., Sato, H., & Hata, Y. (1987). NMDA receptors in the visual cortex of young kittens are more effective than those of adult cats. *Nature*, 327, 513-514.
- Turner, R. W., Bainbridge, K. G., & Miller, J. J. (1982). Calcium-induced long-term potentiation in the hippocampus. *Neuroscience*, 7, 1411-1416.
- Werker, J. F. (1989). Becoming a native listener. *American Scientist*, 77, 54-59.
- Wiesel, T. N. (1982). Postnatal development of the visual cortex and the influence of the environment. *Nature*, 299, 583-591.

# 3 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.

## INTRODUCTION

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 by 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-termi-

nating 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, McCloskey 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. (See also Metcalfe, this volume.)

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 evaluations 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 reasons 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, Schrödinger, 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

