SESSION I PRESIDENTIAL ADDRESS

Russell Church, *Presider*

Where is the randomness for the human computer?

JOHN W. COTTON

University of California, Santa Barbara, California 93106

Psychologists traditionally have employed both statistical and process assumptions in models of human learning and performance. The corresponding tradition in the field of artificial intelligence is to minimize or eliminate the use of statistical assumptions. This article reviews some stochastic and nonstochastic models of human memory, probability learning, medical diagnosis, and concept identification. Some stochastic models are found to have a larger deterministic component than was previously realized; deterministic models applicable with random selection of stimuli can be represented in stochastic form. A policy of methodological determinism is recommended in which the model builder originally assumes that no random processes take place within the organism. If necessary, and as a last resort, such processes can be appended to an otherwise deterministic model for predictive convenience.

Today I want to set before you a theoretical problem in psychology; the question of how great a commitment we should make to the identification of specific causal factors in behavior, as opposed to accepting the milder discipline of identifying what is easy to explain causally and tentatively attributing the residual to random processes. You will recognize this problem as a case of understanding and guiding a paradigm shift (Kuhn, 1970) as psychology moves toward a computer simulation, artificial intelligence-based analysis of human intellectual activity. Much of psychology's success has been grounded in careful experimentation or observation accompanied by interpretations of behavioral differences among and within individual persons as attributable to random events in the environment and within the individual learner. However, as we have developed interests in computer simulation, we have encountered strong convictions on the part of our colleagues in the artificial intelligence community, convictions that random processes should rarely be attributed to individuals. I hope to analyze the arguments on both sides of this issue and to provide you

with some case studies showing the effects of the two approaches.

Another way to state my intentions is to comment directly on the title of this talk. As psychologists, we take for granted that there is variability in performance between and within individuals. Stochastic models may usefully describe these variations even if the underlying processes are deterministic, as in coin tossing. What I propose to ask today is to what degree apparently random variation in the behavior of a human being, viewed as a computer (Hunt, 1971), is attributable to external events rather than to internal ones.

We should remind ourselves that the question of the degree of precision with which causal factors can be identified is an old one that has been encountered in many areas of scholarship and science. Hume (1739/1962) defined a cause as having three properties: (1) spatial contiguity of cause and effect, (2) immediate temporal precedence of the effect by the cause, and (3) invariant occurrence of the effect once the cause has occurred. The third requirement, that the same effect always occurs whenever the same causal event has occurred, is at the heart of the controversy. Cook and Campbell (1979, pp. 14-18) refer to persons who make this third requirement as essentialists. I suppose that we all have sympathy, at least, for the essentialist position. But can we use it as a guide to psychological analysis or must we leave it behind as an article of faith that is cumbersome to carry to our various laboratories?

Turning now to explicitly psychological issues, I

I wish to thank the Committee on Research, University of California, Santa Barbara, for financial support needed in the preparation of this article. I also thank Lawrence Byrd, Alan Bundy, Kam Chow, Noel Fleming, Terry Smith, and Charles McClintock for their comments on the advisability of deterministic theorizing in psychology. However, none of these scholars should necessarily be thought to concur with the specific views expressed here.

quote from Bower and Hilgard's (1981, p. 274) chapter on Estes' stimulus sampling theory: "If behavior is causally determined, how do we account for all this variability? The deterministic approach assumes that the animals really differ in their genetic makeup and prior histories, and that these factors determine an animal's choices and sensitivity to reinforcement. The variable behavior by an animal across trials would be positing variation in the external stimulus situation, in the animal's momentary attention to certain stimuli, to fluctuations in his motivation, to variations in his processing of relevant memories, and so forth. The determinist argues that if all these effective causes were known to us at the moment of the animal's choice, then this choice would appear to be completely determined and, hence, predictable.

"A defensible alternative approach, however, assumes that because behavior is determined by so many variable and unmeasurable causes, the best prediction of behavior we can offer in practice is a probabilistic one. Rather than saying 'He will turn left this trial,' we say 'There's an 80 percent chance he will turn left.' That is like our saying of the toss of a fair coin that there is a 50 percent chance it will turn up heads. Thus, we can characterize our prediction of what is going to happen for a single subject by stating for each trial the probability that each of the several responses should occur."

In another analysis of this issue, Hintzman (1968, p. 125) refers to stochastic processes invoked in his SAL models as conveniences in the derivation of predictions and as "statements of ignorance, rather than assertions that learning is basically probabilistic." Falmagne (1974, pp. 148, 149, and 151), in commentary upon a series of chapters (Krantz et al., 1974) on mathematical psychology and cognitive phenomena, writes: "Whereas the stochastic tradition is one of conventionalism—models are seen as convenient approximations to data, and proving the equivalence of superficially different models is seen as a positive contribution—the simulation field appears to be guided by realist conceptions. Although existing models are recognized as imperfect, the ultimate aim appears to be to capture psychological reality Another, yet associated, difference involves the deterministic philosophy traditional in the computer simulation field. This is clearly a philosophically rather than a technically based discrepancy. Obviously, the deterministic vs. probabilistic character of a model is not an essential difference either formally (as the deterministic model is a limiting case of a probabilistic model) or technically (as random variables could be introduced at any decision node without altering the process in any essential way). However, when random elements are introduced in a simulation model, it is likely to be done with reluctance and apology

"The increasingly detailed level at which data are accounted for, as well as the language in which models are formulated, render the most recent concept-learning

studies stemming from the stochastic tradition quite close to the computer simulation approaches, as can be verified by comparing Millward and Wickens' [1974] chapter with Simon and Newell's [1974] account of concept learning. The essential difference between the two traditions appears to lie in their basic philosophical positions rather than in theoretical constructs and language. Yet, even in terms of philosophical attitudes, one could say that the work in the concept identification area has converged toward the simulation tradition, in that it appears to ultimately aim—explicitly or not—at a deterministic account of fine-grain data, with a corresponding de-emphasis of parsimony and analytically derived predictions."

A wholly deterministic view has been set forth in Winston's (1977, p. 254) text, in a section titled "Myth: Probabilistic Machinery Causes Inspiration and Explains Free Will": "People love explanations. Even when something defies analysis, default explanations fill the void, allowing uncomfortable puzzles to be settled with a minimum of fuss. Many ancient Greeks supported Socrates' opinion that deep, inexplicable thoughts come from the gods. Today's equivalent to those gods is the erratic, even probabilistic neuron "Powerful computer intelligence will rest on description representation, problem solving, and other powerful ideas, not on disorder."

From its beginnings with Logic Theorist (Newell, Shaw, & Simon, 1958; Newell & Simon, 1956), artificial intelligence work has favored the deterministic approach reflected in Winston's (1977) remarks. Simon's (1969) interpretation of the complex path of an ant crossing a beach is relevant here: Instead of assuming complicated internal processes in the ant, we may note the number and position of rocks barring the ant's path and interpret its behavior as controlled by goal directedness plus a simple detour response to each barrier encountered. One may argue that much apparently random behavior like that of the ant is externally rather than internally controlled.

Although deterministic assumptions have often been used in performance models of question answering, comprehension, and problem solving (Bundy & Welham, 1981; Newell & Simon, 1972; Norman & Rumelhart, 1975; Schank & Abelson, 1977; Winograd, 1972), much important artificial intelligence work has not been tested psychologically in the sense of seeing how closely human behavior matches the behavior predicted by relevant computer programs. Furthermore, variability may be less in question answering after learning is complete than in the period of acquiring such capabilities. Hence, psychologists may properly want to withhold judgment about Winston's (1977) strongly deterministic remarks until they have tried to see how they apply to specific problems of behavioral data interpretation, particularly in the field of learning. Before dealing with specific studies, however, let us remind

ourselves of our intellectual roots, for many of us received psychological training that makes it hard to look at behavior in the same way as a computer scientist does.

From the time of Galton (1822-1911) and Cattell (1860-1944), psychology has had a statistical approach to data. Test theory and factor analysis (Thurstone, 1947) made part of the apparently random behavior observed into a nuisance variable, something called error of measurement or error score, but made other random components the center of the interpretation of cognitive behavior, true score variance and factor scores being the reliable contribution of individual people's abilities to the total scores observed. The test theory and psychometric area of research continues to be very strong today, but our focus is upon a second field to which Thurstone made a contribution: His accumulation theory of learning (Thurstone, 1930) was one of the first mathematical models of a learning process and may be considered an early ancestor of the dominant tradition in mathematical learning theory, Estes' (1950, 1959) stimulus sampling theory (see also Estes & Suppes, 1974). As is well-known to us all, a variety of models by Estes and many other investigators fall within the domain of stimulus sampling theory; other modern mathematical theories of learning (e.g., Bush & Mosteller, 1955) sometimes lack surface features of stimulus sampling theory but make very similar predictions. All or almost all of these models imply that organisms use random processes such as sampling from externally presented or internally available stimulus elements in deciding the response to make in a given stimulus situation.

Despite the statistical side of early psychological research work, the seeds of determinism lay in some of that research. Guthrie's (1935, 1946) verbally rather than quantitatively stated theory of behavior was one of the inspirations of Estes' (1950) initial paper on stimulus sampling theory. The Guthrie model showed deterministic leanings in its assertion that if an exact stimulus situation was presented in which an organism had previously made a response and been removed from the situation, the response would almost invariably be made again. Later, we will see that an early chapter by Sternberg (1963) looked at deterministic approximations to stochastic learning models. At that point, I will go somewhat beyond Sternberg in suggesting that there is a good deal of determinism in many apparently stochastic models. And, at a later point, I will complete the circle by showing that deterministic models can simulate highly variable behavior and asserting that one deterministic model can even be so formulated as to be mathematically indistinguishable from well-known stochastic models. But, first, let us take a look at a few classic topics of cognitive psychology, such as human memory and probability learning.

COMPUTER MODELS OF HUMAN MEMORY

Recently, many of the most influential students of human memory have used computers either to build programs modeling aspects of human performance or as prototypes themselves of human memory processes. It is fair to say that stochastic processes are still involved in the most comprehensive of these theories (e.g., Anderson, 1976; Anderson & Bower, 1973). However, deterministic models are surprisingly effective in certain instances. The EPAM models of Feigenbaum (1961/1963) and of Simon and Feigenbaum (1964) predicted a number of phenomena such as the shape of the serial position curve and the effects of rate of stimulus presentation on learning. As you know, the central feature of EPAM was its discrimination network for controlling performance, with the network itself being the consequence of two learning processes: the building of images at its terminal nodes and the addition of new branches at those nodes. Figure 1 illustrates the predictive success of EPAM-III by showing a fit of some of Underwood's (1953) data on intralist similarity effects during paired associate learning, supporting a hypothesis that one-third of Underwood's subjects were coding the syllables visually and two-thirds were coding them aurally.

When Hintzman (1968) expanded the EPAM model to encompass further human memory phenomena, he took a stochastic direction, introducing three choice points for random decisions to be made: (1) whether or not to add a new test node on occurrence of an incorrect response, (2) whether or not to change the response after an error if a new node is not added, and (3) whether or not to add a new test node after a correct response.

Figure 1. Adequacy of the fit of EPAM-III predictions to observed learning difficulty (relative number of trials to criterion) as a function of stimulus similarity (Underwood, 1953). All data are for low response similarity. Numerical values were taken from Simon and Feigenbaum (1954, Table 1).

Use of these random processes facilitated Hintzman's predictions for a wide number of independent variables, such as number of response alternatives, spacing of response alternatives, spacing of item repetitions, and degree of original learning. Hintzman argued that the use of stochastic processes in his model was necessary in order to minimize the number of specific mechanisms assumed. Of 18 phenomena analyzed, 14 were simulated with reasonable success. As in the case of many competing deterministic models, little attention was given to the prediction of response variability or sequential statistics, except in Experiment 2 on number of response alternatives, in which the two-response condition simulation conformed well to fine-grain predictions of the one-element model of statistical learning theory.

Atkinson and Shiffrin (1968) developed an early quantitative model assuming that external input proceeded through a sensory register and on to a short-term store (STS) with a rehearsal buffer for maintaining the item in STS so that other STS operations, such as coding and identification of a matching representation in the long-term store (LTS), could permit item transfer to LTS. This model has been greatly expanded by subsequent research (e.g., Raaijmakers & Shiffrin, 1981; Schneider & Shiffrin, 1977; Shiffrin & Atkinson, 1969; Shiffrin & Schneider, 1977), and the sensory register has been combined with STS in later interpretations of the Atkinson-Shiffrin model (Shiffrin, 1977). For our present purposes, it is important to note that the model has been and continues to be stochastic rather than deterministic. For example, Experiment 8 of Atkinson and Shiffrin (1968) determined the probability of correct recall of the color of a playing card as a function of the length of the set displayed and of the serial position in which it had appeared. Of four parameters used to predict those percentages, one (δ , the probability of deleting the oldest item in the buffer was displayed) was clearly probabilistic and one or two others had probabilistic overtones.

Gilmartin, Newell, and Simon (1976) seemed to fault the Atkinson and Shiffrin (1968) model for its probabilistic nature, saying that the control processes in that model had not yet been fully classified. Gilmartin et al. and a companion report (Gilmartin, 1975) described a deterministic computer program called SHORT developed to simulate short-term memory processes, such as memory span for a list of digits. SHORT uses specific subprograms or strategies, such as "sense," "perceive," and "rehearse," to control behavior. At least one parameter, the 8 sec required to consolidate a new entry in the LTS, is estimated as a composite value from several previous studies. Other parameters are eight cells for STS and 4.0 single-syllable items or 2.9 two-syllable items rehearsed in 1 sec. However, curve-fitting parameters are not routinely estimated in individual simulations. SHORT is perfectly predictable in its behavior at Time t , but variability in performance from trial to trial and from person to person is generated in two ways:

(1) by letting the onset of each trial occur at a random time with respect to ongoing processes in SHORT and (2) by having stimuli for the different trials selected at random. The 13 simulations run by Gilmartin were relatively successful. However, Gilmartin et al. did not show observed values together with their plots of simulated results; the most quantitative comparison they present between theory and data is that mean forward serial recall for a visually presented list of 9 words was simulated at 4.56, 5.26, or 5.02 words, depending on the number of items assumed to fall in the first two rehearsal groups of STS, whereas Jahnke (1965) found mean recall of 4.6 words with lists of 6 and 10 words. SHORT seems to be a promising model, but it is still a mini-theory by comparison with its stochastic competitors.

Probability Learning

Probability learning studies require experimental subjects to guess which of two or more events such as A or B will occur on successive trials. After the guess for a trial has been made, feedback is given by displaying a certain letter or turning on a certain light associated with one of the possible events. Under the assumption of random noncontingent feedback with π being the probability of Event A on each trial, Estes and Straughan (1954) predicted from statistical learning theory that linear operators specific to each possible feedback would apply, leading to gradual growth in $p_n(A)$, the probability of guessing A on Trial n , from equiprobability for each possible response to π as an asymptotic value. On trials rewarding A,

$$p_n(A) = (1 - \Theta)p_{n-1}(A) + \Theta, \quad (1)$$

where Θ is the probability of sampling a given stimulus element on a trial. On trials rewarding B,

$$p_n(A) = (1 - \Theta)p_{n-1}(A). \quad (2)$$

Good fits of observed data with this theory often resulted. However, Myers (1976) has pointed to a variety of empirical weaknesses in this interpretation of probability learning. In the present context, the principal difficulty may be said to be extensive evidence from Myers (1970) and others showing that individuals pay attention to patterns of two or more previous events, such as AB or AABB. A variety of models were developed to account for this finding, with the most successful approach being the generalization model of Gambino and Myers (1967), in which linear operators depended upon the continuation or ending of a run of identical events and changes in the subjects' expectancies about a current run length generalized to other run lengths. Myers (1976) also outlined the beginnings of an information processing analysis of this phenomenon of responding differentially as a function of the immediately preceding reinforcement pattern.

Feldman (1962) proposed a quite complicated

deterministic model of probability learning. The later Feldman and Hanna (1966) model is simpler, more successful, and somewhat reminiscent of EPAM. It has both stochastic and deterministic features, combined in such a way as to merit being called a mixed model.

Figure 2 shows a discrimination net for the Feldman and Hanna (1966) stimulus structure. Consider, for example, a situation in which the immediately previous trials were B followed by A. This situation is represented on the net by sort of a backward path: We pursue the left-hand branch from the top (Event $t-1$ is A) and the right-hand branch below it (Event $t-2$ is B). We are led to a rectangular box labeled $\Pr(A) = 9/16$. This tells us that in 9/16 of the cases following a BA pattern the next stimulus will be an A.

Figure 2 can also be viewed as the decision network grown by the learner during the course of the experimental session. Like Estes and Straughan's (1954) model, it implies matching of the asymptotic $p(A)$ to π , the overall input probability of A. But the Feldman and Hanna (1966) model went further to match the input probabilities at the tips of the Figure 2 network to the conditional response probabilities $p(A|S)$, where S is the stimulus frequency for a branch. How effective was this model over a 1,000-trial experiment with 20 persons? On the average, 17% of the model's predictions of A or B responses were in error. This result compares favorably with a 21% error rate with $\alpha = .995$ in the Bush and Mosteller (1955) analogue of the Estes and Straughan model. Feldman and Hanna reported this error rate to be lower than that of several other models and sub-models as well. The true accuracy of the Feldman and Hanna model may be slightly underestimated, since its predictions were attempted for all trials even though part of a session consisted of learning the discrimination net on which the predictions were based.

How did this learning take place? The subjects were

Figure 2. Discrimination net for a probability learning task with fixed run probabilities. Note—This is Figure 2 from Feldman and Hanna (1966). (Copyright 1966 by Academic Press. Reprinted by permission.)

assumed to be informal statisticians who gathered information about $\Pr(A|AS)$, the probability of an A stimulus after a string beginning with A followed by a substring S . They compared this probability with their information about $\Pr(A|BS)$, the probability of an A stimulus after an equal-length, otherwise identical, string beginning with B. The longest string for which these two probabilities were significantly different was built into the decision network mentioned earlier. As training progressed, the decision tree grew larger and more like Figure 2 for the event structure. When no significant basis for choice existed on a given trial, the learner apparently responded randomly with $p_n(A) = .50$.

What can we say about the degree to which the Feldman and Hanna (1966) model is stochastic or deterministic? The nature of the probability learning task seems well suited for random behavior because the stimuli have such limited predictability. Accordingly, it is not surprising to find probabilistic nodes at the bottom of the behavioral analogue of Figure 2. But note that any variability in acquisition of the decision network is completely stimulus controlled. By this, I mean that the learner grows his or her decision network deterministically on the basis of statistical significance tests with no randomness introduced by the learner. This growth process is not intuitively satisfying and may be empirically implausible on the basis of more recent evidence on human responses in the face of uncertain stimuli (Slovic, Fischhoff, & Lichtenstein, 1977). Nonetheless, this is a moderately successful combined or mixed model with deterministic growth of its decision tree and stochastic control of responses generated at its nodes.

A strong determinist might take no offense at this model, noting that random processes are not invoked until all useful knowledge of the stimulus input sequence has been processed; even here, a deterministic pseudorandom-number generator could be called by the individual. One must remember, of course, that a person with a goal of maximizing the percentage of correct choices and with enough strategic knowledge would predict the more likely event at a node 100% of the time rather than doing the probability matching commonly obtained in this type of experiment.

Estes' (1976a, 1976b) most recent analysis of probability learning considered quite a different paradigm from that of earlier studies: He focused upon the effects of cycles of many observation trials showing paired "election candidates" winning or losing plus several test trials in each cycle in which subjects had to predict the results of additional contests. These experiments showed that the effects of total number of wins for each of two candidates not always pitted against each other could outweigh the relative frequency of their wins in earlier head-to-head contests. For example, this principle implies that Candidate A might have beaten Candidate B in 58% of their matched contests; but, if Candidate A had few wins against other candidates and Candidate B had many wins against other

candidates, Candidate B would be more frequently predicted as the winner in one-to-one choices by the experimental subject. Information about overall attractiveness of the candidates has triumphed over information about their relative attractiveness. Estes explained his findings with a scanning model: Persons are presumed to use a probe of a coded winning event, for example, plus any currently available contextual cues to seek a matching memory trace as a guide to response on a test trial. If that match is for Outcome A, then the individual chooses A.

Let me now say something a bit shocking.¹ This scanning model, despite its dependence on randomly selected internal traces, looks to me as if it were the random tag on a deterministic model, making it a mixed model like the one just discussed. I go further to say that all noncontingent cases of the Bush and Mosteller (1955) linear model and Estes and Straughan (1954) model are also mixed models, at least at the level of linear models rather than of sampling hypothesized internal and external stimulus elements. I defend this claim for the Estes (1967a, 1967b) scanning models as follows: When the only coded traces are of winning events, a test of Candidate A vs. Candidate B yields the following value for the probability of choosing A: $p(A) = W_A / (W_A + W_B)$, where W_A is the number of previous wins with A and W_B is the number of previous wins with B, based on all contests, not just those pairing A and B. The reason this version of the scanning model is a mixed model is that W_A and W_B are wholly determined by the experiment. Both in this model and in the other noncontingent models recently discussed, a response probability may be viewed as one of an infinite number of states the person may enter. Hence, what I have been calling a mixed model may be thought of as having deterministically entered states and randomly generated responses, given those states.

Lest you suspect that I will try to find some way to call all stochastic models deterministic, I hasten to say that all-or-none Markov models still look stochastic to me because state entry is not predictable from stimulus history. I have already promised to show some relations between such models and deterministic models. But, for now, let me say that one of the hard tasks of the determinist is to explain the superiority of all-or-none models over linear models in certain situations. For example, Bower (1961) found that the one-element all-or-none model for paired associated learning correctly predicted the same mean number of errors but a larger (observed) variance of errors than did the corresponding linear model. From the present perspective, this finding could be scored as a point for stochastic rather than deterministic theorizing.

This discussion of mixed models bears some resemblance to an early presentation of deterministic approximations in psychology. Sternberg (1963, p. 39) wrote: "The models we have been dealing with are, in a sense, doubly stochastic . . . For a single subject, both the

sequence of responses and the sequence of p-values are governed by probability laws." In describing a deterministic approximation for an urn model, he gave the following as one interpretation of the result (Sternberg, 1963, p. 42): "The approximate process defines a determined sequence of approximate probabilities for a subject. Like the original process the approximation is stochastic, but on only one 'level': the response sequence is governed by probability laws but the response probability sequence is not. According to this interpretation, the approximate model is not deterministic, but it is 'more deterministic' than the original urn scheme."

What is being said here is almost exactly the same, but for cases in which no deterministic approximation is required: With the linear model and noncontingent reinforcement, the process defines a determined sequence of probabilities for a subject, conditional upon the stimulus sequence assigned to that subject. For the subject, the response sequence is governed by probability laws, but the response probability sequence, such as $p(A) = .5, .55, .595, .5355, \dots$, for a stimulus sequence beginning with two A events followed by a B, assuming an initial response probability of .5 and a Θ of .1, is deterministic and is not governed by internal probability rules. Any randomness in stimulus values is by the choice of the experimenter, rather than because of the nature of the task of the learner.

Medical Diagnosis

An interesting comparison of stochastic and deterministic approaches to medical diagnosis was performed by Fox (1980), who experimented with information processing of relations between five symptoms and five diseases. Eighteen medical school students each made eventual diagnoses of 75 instances of disease from the list of five (tonsillitis, laryngitis, meningitis, hepatitis, and mumps) under consideration. Each final diagnosis followed a sequence involving an original computer presentation of a positive symptom (fever, vomiting, headache, earache, or difficulty in swallowing) followed by alternating requests and fulfillment of requests for information about the presence or absence of an individual symptom for that clinical case, with a tentative or final diagnosis by the student and the next set being possible at any time. Between each set of 24 diagnostic cases, there were 25 memory trials in which students were asked whether each possible symptom was in fact statistically associated with each possible disease. Response latencies from this memory task provided data on information available for a deterministic cognitive production system model of question-asking and diagnosis steps in the diagnosis task. A competing Bayesian production system made predictions for the same task.

What was the relative effectiveness of these two models? Overall diagnosis was slightly better for the deterministic cognitive model (77% correct) compared with the Bayesian model (72% with the most favorable

parameter values for that model). The two models were roughly comparable in their prediction of the first question by a student conditional upon a certain presenting symptom. However, the deterministic cognitive model was much superior in predicting the next question following a presenting symptom, question, and feedback about a possible second symptom.

The superiority of Fox's (1980) deterministic production system may be largely due to its cognitive nature, its use of memory data for which the Bayesian model has no application. Observed inconsistencies of performance within individuals would be problems for deterministic theorists to explain, but nonetheless, the data are helpful to those hoping to make psychology into a deterministic science.

SOME DETERMINISTIC MODELS OF CONCEPT LEARNING

Psychological models of concept identification (e.g., Bower & Trabasso, 1964) often stemmed from stimulus sampling theory and thus were stochastic. However, Hunt (1962) emphasized deterministic information processing in concept learning, Gregg and Simon (1967) and Restle (1962) allowed stochastic events only in the random selection of hypotheses to test, and Indow and Suzuki (1972, 1973) compared the predictive efficiency of certain stochastic and deterministic models of concept identification. Most of the mathematical psychology models of concept learning have been restricted to tasks in which the value of a single stimulus dimension such as color is a necessary and sufficient predictor of the correct response.

In developing a computer program to learn structural concepts such as an arch, Winston (1975) began an artificial intelligence tradition of deterministic modeling of concept learning. This tradition has continued with Michalski's (1975, 1976) AQVAL/1 program for meeting a simplicity criterion in the classification of physical tables, of schematic faces, and of nonsense figures into predefined categories and his related work (Michalski, 1980) on nonstatistical clustering without such predefinitions. Another important example is Wysotzki, Kolbe, and Selbig's (1981) use of a relational algebra model for structured objects as the starting point for programming learning of such concepts as the classifi-

cation of 89 chemicals into four groups of biological agents.

In the remainder of this section, I will report some work done by Chow and Cotton (Note 1) in developing and testing some deterministic, computer-based models of concept identification. These programs, written in the PROLOG language (Roussel, 1975; Clocksin & Mellish, 1981, Note 2) are slight variants of a family of programs developed by Cotton, Byrd, and Bundy (Note 3) for learning to solve algebra problems of the form $X + 3 = 7$ for X . Despite being more specialized, these programs have similarities to the Simon and Lea (1974) general rule inducer: There is a space of instances ("source" and "target," in our terminology) and a space of rules that could be induced for generating the target from the source. Each program searches the rule space for a rule appropriate to a given instance and uses that rule in an attempt to predict the target in new instances.

Let us consider a deterministic strategy for generating one-dimensional solutions to a concept identification task. Can such a model generate within-learner and between-learner variability that matches observed behavior? In the process, is a stochastic model the result of our deterministic assumptions? Before answering these questions, we describe a task closely coordinated to our computer program input format. Let a problem stimulus be a row of three digits, each a "0" or a "1." Let the correct response be "1" when the leftmost problem digit is a "1." Otherwise, the correct answer is "0." The learner first sees an example problem, such as "111," together with an answer display of "1." In later trials, he or she will see a problem, respond with an answer, and receive feedback on each trial.

What will the learner believe after the example trial? The model we consider first says that the learner begins at the right of the stimulus display and considers first the alternative hypotheses, "Match the rightmost digit" and "Reverse the rightmost digit." In view of the information in the example, the former hypothesis will be chosen. Table 1 shows this result for the example trial. Note that the notation $..1 \rightarrow 1$ is used to indicate that the first and second digits in the stimulus can be ignored, but that a "1" for the third digit implies a "1" for the answer. The remainder of Table 1 will be interpreted as we learn further assumptions of the model. Assume that any hypothesis developed by a subject will

Table 1
An Example Trial and Three Problem Trials Encountered by CONCEPT, Such That All Responses Shown Are Incorrect but All Subsequent Responses Will Be Correct

	Example	Problem 1	Problem 2	Problem 3	Problem 4
Source and Target	111 1	001 0	010 0	000 0	010 0
Old Hypothesis	None	$..1 \rightarrow 1$	$..1 \rightarrow 0$	$..1 \rightarrow 0$	$..1 \rightarrow 1$
Response	None	1	1	1	1
New Hypothesis	$..1 \rightarrow 1$	$..1 \rightarrow 0$	$..1 \rightarrow 0$	$..1 \rightarrow 1$	$1.. \rightarrow 1$

Note—The new hypothesis for Problem 4 is the solution.

be used on the next problem presented. If that hypothesis leads to the correct response, it will be retained for the next trial. If not, the learner will check the rightmost digit of the current stimulus for a previously unused hypothesis consistent with the current feedback. So Problem 1 leads to a new hypothesis opposite to the original one but still based on the third dimension.

On later trials, the same procedure of retaining successful hypotheses and giving up unsuccessful ones (win-stay, lose-shift strategy) is assumed to be followed. An assumption of complete memory of unsuccessful rules prevents trying a rule again after it has failed. Accordingly, later trials use hypotheses based on the central or left dimensions of the stimulus. The specific problems shown in Table 1 have been so ordered as to force an error on every problem and thus to move the learner through every possible hypothesis in a minimum number of trials. For this order of searching for hypotheses, the model formulated says that everyone receiving the example and problem sequence in Table 1 will be correct on all subsequent trials, regardless of the problem presented. All learners given some other specific sequence of example and problems are predicted to give identical response patterns to each other but not necessarily identical patterns to the sequence of Table 1.

We may call this model the stimulus sequence (SS) model, because it makes an explicit prediction of behavior in a set of trials with any specific stimulus sequence. As stated, the model could be accurate for a few individuals but is almost surely wrong for any unselected group of subjects in concept identification experiments. In its computer form, this model is called CONCEPT_RULE_IND, a concept program with rule memory for individuals with Permutation 6 in the arrangement of the three dimensions to be considered. There are five other permutations available through this program; they correspond to five other SS models, so that the likelihood is increased that some SS model might fit any given person.

Let us now assume something new about a concept identification experiment, rather than about the learners. If each of the eight possible combinations of three-dimensional stimuli just discussed is equally likely on any trial, with the stimulus presented on any trial being independent of those on all other trials, the SS model already discussed implies the state diagram for a Markov chain shown in Figure 3. Despite the deterministic nature of the SS model, adding the stimulus generation assumption causes the progression from hypotheses associated with the rightmost dimension to a hypothesis associated with the leftmost dimension to be probabilistic, as shown in Figure 3. This figure represents what may be called a random stimulus sequence (RSS) model, with the same plausibility or lack of plausibility as the SS model but the potential for making predictions of response patterns averaged across the population of possible stimulus sequences.

Figure 3 shows five labels of numbers of errors. This suggests the probability that the seven states of Fig-

Figure 3. Transition state diagram for the RSS deterministic concept model with random selection of stimuli and a specific hypothesis order. (A vertical bar separating two rules means that the left-hand rule is now held as a successor to the right-hand rule.)

Figure 4. Transition state diagram for the RSS-U deterministic model with random selection of stimuli and of ordering of hypotheses to be tested. (S4 means the state in which four errors remain to be made; other states are similarly labeled.)

ure 3 could be reduced to five, with States 0-4 being defined by the number of errors yet to be made before attaining the correct hypothesis. Figure 4 displays this revised representation of state transitions within the RSS model.

Not only is Figure 4 a simplification of the Markov process for the RSS model, but it also represents each of the five other possible RSS models with different permutations of dimension orders in the hypothesis stack. Accordingly, we speak of it as representing an RSS-U model (i.e., an RSS model with an unrestricted hypothesis order). The only difference between different permutation or hypothesis orders is in the entry point into the Markov chain of Figure 4. The most natural permutation method would yield two orders having four errors, two having two errors, and two having no errors remaining at the beginning of a session. However, if we assume that half the persons beginning with the rightmost dimension will test both hypotheses for that dimension and half will test only one hypothesis before moving to hypotheses based on the middle dimension, this will lead to four errors for the former group and only three errors for the other group. A similar argument for people starting with hypotheses based on the middle dimension will lead to one group of learners with two errors and another with only one error. This new assumption allows us to state the following initial vector for the RSS-U model:

$$P_1 = [1/5 \ 1/5 \ 1/5 \ 1/5 \ 1/5], \quad (3)$$

rather than having all the density of the initial vector in a single state, as would be true for the RSS state for a given permutation. An obvious transition matrix corresponds to Figure 4:

$$T = \begin{matrix} & \begin{matrix} 0 & 1 & 2 & 3 & 4 \end{matrix} & \text{errors} \\ \begin{matrix} 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{matrix} & \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix} \end{matrix} \quad (4)$$

Given the initial vector and transition matrix for the RSS-U model, methods summarized by Millward (1969) can be used to generate a variety of quantitative predictions of concept identification performance. Mathematical rather than computer program predictions are made with this model. The 50 members of the variable group of Cotton's (1971) Experiment 2 did receive 24 trials on a three-dimensional affirmative concept task in which each possible stimulus had an equal and independent chance of selection on each trial for a given person. Accordingly, that group seems an appropriate source of data for testing the RSS-U model. It should be mentioned that the stimuli of this study may be described abstractly as in Figure 3 but are actually single letters "a" and "b," varying in three dimensions, upright vs. sideways (Dimension 1), uppercase vs. lowercase (Dimension 2), and "aye" vs. "bee" (Dimension 3). How well does the theory meet this test?

Only 38 of the 50 students in the variable group met a criterion of 10 successive correct responses. All but 1 of the 12 nonsolvers made five or more errors, thus contradicting the prediction of a maximum for four errors. A further contradiction of the model (and indeed of many stochastic models as well) was the existence of 10 "backsliders" among the solvers, meaning that these 10 persons met the criterion of learning just mentioned but subsequently made one or more errors. A fourth contradiction of the model was that three of the solvers made from five to eight presolution errors, rather than the maximum of four that was predicted.

Further evidence comes from the analysis of hypotheses held during the experiment. In contrast to Levine's (1975) use of blank trials in such tracking, we identified hypotheses by comparing stimulus values and responses as a function of the type of feedback ("correct" or "incorrect") given. A win-stay, lose-shift strategy with the shift conforming to local consistency requirements has been already noted to follow from CONCEPT_RULE_IND. Thus it applies to the RSS-U model as well. Under the presumption that this strategy holds, most trials have identifiable hypotheses; when the strategy could not have been operating, the stimulus-response-feedback pattern shows the failure. We find that only 12 of the 50 students always were consistent with this implication of the RSS-U model.

Now consider traditional measures of goodness of fit in mathematical models of learning. We restrict this

analysis to presolution data for the 38 solvers and use Trial 1 as an example trial, ignoring each response given on it. Chow and Cotton (Note 1) found that the RSS-U model gave reasonably good predictions of mean number of errors (2.00 compared with 2.34), its standard deviation (1.41 compared with 1.66), mean trial of last error (4.00 compared with 4.95), and its standard deviation (3.46 compared with 3.59). The average number of error runs of a given length were also predicted fairly well. Chow and Cotton also discussed an RSS-U nine-state model that uses the observed distribution of number of errors as an initial vector and a nine-state transition matrix that is otherwise much like the one in Equation 4. The principal virtue of the nine-state model is that it allows for the occurrence of up to eight errors by some persons, presumably as a result of allowing certain irrelevant hypotheses to appear more than once in the hypothesis stack. Chow and Cotton concluded that these two deterministic models were fully as successful as two versions of the Bower and Trabasso (1964) model in predicting statistical properties of the data set for the 38 solvers in the variable group of the Cotton (1971) experiment. Chow and Cotton also found that a certain infinite-state deterministic model is mathematically equivalent to the stochastic Bower and Trabasso model with $q = .5$, where q is the probability of a chance error.

A difficulty of the RSS-U models is that, while they appear to go beyond traditional stochastic models in predicting behavior as a function of the structure and arrangement of the hypothesis stack, we have no a priori way to predict these arrangements for individual persons and are forced to use aggregate data from persons presumed to have a variety of such arrangements. Possibly, we can learn more about hypothesis (or rule) orderings by examining trial-by-trial data from individuals in the constant group of Cotton's (1971) Experiment 2, a group in which all 50 persons received the same sequence of 24 trials of stimuli and associated feedback. Here, similarities in performance by different learners can be taken as reflecting similarities in the order in which corresponding hypotheses were stacked and tested.

What do we learn from the constant group? We find that 22 of 40 solvers showed behavior completely consistent with the win-stay, lose-shift local consistency strategy, as did 1 of the 10 nonsolvers. Eighteen of the 22 consistent solvers showed particularly interesting patterns: Ignoring any errors on Trial 1 since it is theoretically an example trial, there were two solvers with zero errors, seven with errors only on Trial 4, and nine with errors only on Trials 3 and 4. Accordingly, we can conclude that 18 of the 22 most consistent performers in the constant group fall into three deterministic states comparable to the two-error, one-error, and zero-error states of the RSS-U model.

One reason to hesitate about the adequacy of the RSS-U model is that only 3 of the 22 consistent solvers ever showed both hypotheses for any dimension being

Table 2
Number of Solvers in the Constant Group of Cotton (1971,
Experiment 2) Conforming to Each Dimension
Ordering in CONCEPT-EXIND

Ordering Number	Errors Predicted on Trials	Number of Solvers
1 or 2		2
3	3	0
4	3, 4	7
5 or 6	4	9
Error Pattern Not Predicted		4

Note—CONCEPT-1XIND (one-trial example memory) makes identical predictions but adds errors on Trials 8 and 10 for both Orderings 4 and 6.

used in succession, as implied by that model. Accordingly, it seems desirable to discuss some slightly different deterministic models at this time. CONCEPT-1XIND is a PROLOG program without rule memory but with one-trial memory such that a consistency check (Trabasso & Bower, 1966) between stimuli and feedback on an error trial and its immediate predecessor control the selection of the hypothesis for the next trial. CONCEPT-EXIND is a very similar program with a consistency check for an error trial and all previous trials. However, two-trial memory is all that is actually needed for the predictions to be discussed next.

Table 2 shows that five of the six permutations of the three dimensions of this experiment can be used with the all-trials memory model, CONCEPT-EXIND, to generate error trial patterns that were actually observed. However, no individual learned with only a Trial 3 error, as implied by the program with Permutation 3. This suggests that this ordering, which put capital vs. lowercase letters first and the dimension for specific letters ("aye" vs. "bee") last, was not exhibited in our sample.

There are many loose ends in the deterministic interpretation of this experiment. For example, how well could these last two programs describe individual solvers in the variable group? Can the persons who do not conform to local consistency assumptions be interpreted as behaving deterministically nonetheless? How should backsliders' behavior be described? Can we tell whether a person following a win-stay, lose-shift strategy is selecting the hypothesis following an error deterministically rather than randomly, and so on? But it is time to sum up, and so I drop such questions for now.

CONCLUSION

Let us return to our first question: Where is the randomness for the human computer? I believe the case has been made that much intraperson variability can be attributed to stimulus variability without a need to appeal to random processes within the individual. It is also plausible, but perhaps a bit devoid of predictive

power, to attribute much interperson variability to differences in the order in which variables or dimensions are considered. I leave to you an overall assessment of strict determinism as a guiding principle for psychological research. Instead, let me make two points.

First, we have seen a good deal of convergence between stochastic and deterministic modeling. Some stochastic models have strongly deterministic core assumptions, and some deterministic models can generate stochastic counterparts quite easily. My second point is easily stated but requires elaboration: I am ready to advocate what might be called methodological determinism. By this, I mean that our preference should be to state and test very specific deterministic models because of the possibility of describing individual persons' moment-by-moment behavior with such models. This leaves the question of ultimate feasibility of the deterministic approach open. As a practical measure, I would agree with several psychologists quoted earlier in the belief that stochastic assumptions may currently be necessary to describe certain data sets adequately. Perhaps, though, these stochastic assumptions should be added at the last minute in the modeling process, rather than being introduced in the initial interpretation of a certain learning situation.

Except in an earlier mention of Hume's (1739/1962) views, I have resisted a strong temptation to consider our problems with determinism from the point of view of philosophy or of another science. Now I want to succumb to that temptation again in order to tell you where Einstein (1950, pp. 109-110) stood on this matter: "The aim of the [quantum] theory is to determine the probability of the results of measurement upon a system at a given time. On the other hand, it makes no attempt to give a mathematical representation of what is actually present or goes on in space and time It must be admitted that the new theoretical conception owes its origin not to any flight of fancy but to the compelling force of the facts of experience. All attempts to represent the particle and wave features displayed in the phenomena of light and matter, by direct course to a space-time model, have so far ended in failure. And Heisenberg has convincingly shown, from an empirical point of view, any decision as to a rigorously deterministic structure of nature is definitely ruled out, because of the atomistic structure of our experimental apparatus. Thus, it is probably out of the question that any future knowledge can *compel* [italics added] physics again to relinquish our present statistical theoretical foundation in favor of a deterministic one which would deal directly with physical reality."

Einstein went on to say that deterministic and statistical models will probably coexist, with the choice between them being made on the grounds of their relative simplicity. Despite his preference for deterministic models, he stated that he had no adequate competitor for the quantum theory. Yet he was unwilling to give up

his search for that alternative interpretation of existing knowledge.

What is the relevance of these remarks to us as psychologists? Although I view a deterministic interpretation of behavior as more promising than I did when I began this address a year ago, I believe that we have less compelling reasons for making a deterministic choice than Einstein did for making an unwanted stochastic choice. In any case, an essential feature of science is the freedom of investigators to think for themselves. So I close with one more sentence from Einstein (1950, p. 110): "It is open to every man to choose the direction of his striving; and also every man may draw comfort from Lessing's fine saying, that the search for truth is more precious than its possession."

REFERENCE NOTES

1. Chow, K. P., & Cotton, J. W. *A deterministic model of concept identification*. Paper presented at the Mathematical Psychology meetings, Santa Barbara, California, August 22-24, 1981.
2. Clocksin, W. F., & Mellish, C. S. *The UNIX PROLOG system (Version NU7 with top level)* (Software Report 5, 2nd ed.). Edinburgh: University of Edinburgh, Department of Artificial Intelligence, September 1980.
3. Cotton, J. W., Byrd, L., & Bundy, A. *How can algebra steps be learned by students with only arithmetic skills?* (Informal working paper). Edinburgh: University of Edinburgh, Department of Artificial Intelligence, September 1980.
4. Estes, W. K. Personal communication, November 12, 1981.

REFERENCES

- ANDERSON, J. R. *Language, memory, and thought*. Hillsdale, N.J.: Erlbaum, 1976.
- ANDERSON, J. R., & BOWER, G. H. *Human associative memory*. Washington, D.C.: Winston, 1973.
- ATKINSON, R. C., & SHIFFRIN, R. M. Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation* (Vol. 2). New York: Academic Press, 1968.
- BOWER, G. H. Application of a model to paired-associate learning. *Psychometrika*, 1961, **26**, 255-280.
- BOWER, G. H., & HILGARD, E. R. *Theories of learning* (5th ed.). Englewood Cliffs, N.J.: Prentice-Hall, 1981.
- BOWER, G. H., & TRABASSO, T. Concept identification. In R. C. Atkinson (Ed.), *Studies in mathematical psychology*. Stanford: Stanford University Press, 1964.
- BUNDY, A., & WELHAM, B. Using meta-level inference for selective application of multiple rewrite rule sets in algebraic manipulation. *Artificial Intelligence*, 1981, **16**, 189-212.
- BUSH, R. R., & MOSTELLER, F. *Stochastic models for learning*. New York: Wiley, 1955.
- CLOCKLIN, W. F., & MELLISH, C. S. *Programming in PROLOG*. Berlin: Springer-Verlag, 1981.
- COOK, T. D., & CAMPBELL, D. T. *Quasi-experimental designs*. Chicago: Rand McNally, 1979.
- COTTON, J. W. A sequence-specific concept identification model: Infrastructure for the Bower and Trabasso theory. *Journal of Mathematical Psychology*, 1971, **8**, 333-369.
- EINSTEIN, A. *Out of my later years*. New York: Philosophical Library, 1950.
- ESTES, W. K. Toward a statistical theory of learning. *Psychological Review*, 1950, **57**, 94-107.
- ESTES, W. K. The statistical approach to learning theory. In S. Koch (Ed.), *Psychology: A study of a science* (Vol. 2). New York: McGraw-Hill, 1959.
- ESTES, W. K. The cognitive side of probability learning. *Psychological Review*, 1976, **83**, 37-64. (a)
- ESTES, W. K. Some functions of memory in probability learning and choice behavior. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 10). New York: Academic Press, 1976. (b)
- ESTES, W. K., & STRAUGHAN, J. H. Analysis of a verbal conditioning situation in terms of statistical learning theory. *Journal of Experimental Psychology*, 1954, **47**, 225-234.
- ESTES, W. K., & SUPPES, P. Foundations of stimulus sampling theory. In D. H. Krantz, R. C. Atkinson, R. D. Luce, & P. Suppes (Eds.), *Contemporary developments in mathematical learning theory* (Vol. 1). *Learning, memory, and thinking*. San Francisco: Freeman, 1974.
- FALMAGNE, R. J. Mathematical psychology and cognitive phenomena: Comments on preceding chapters. In D. H. Krantz, R. C. Atkinson, R. D. Luce, & P. Suppes (Eds.), *Contemporary developments in mathematical psychology* (Vol. 1). *Learning, memory, and thinking*. San Francisco: Freeman, 1974.
- FEIGENBAUM, E. A. The simulation of verbal learning behavior. In E. A. Feigenbaum & J. Feldman (Eds.), *Computers and thought*. New York: McGraw-Hill, 1963. (Reprinted from *Proceedings of the Western Joint Computer Conference*, 1961, **19**, 131-132.)
- FELDMAN, J. Computer simulation of cognitive processes. In H. Borko (Ed.), *Computer applications in the behavioral sciences*. Englewood Cliffs, N.J.: Prentice-Hall, 1962.
- FELDMAN, J., & HANNA, J. F. The structure of responses to a sequence of binary events. *Journal of Mathematical Psychology*, 1966, **3**, 371-387.
- FOX, J. Making decisions under the influence of memory. *Psychological Review*, 1980, **87**, 190-211.
- GAMBINO, B., & MYERS, J. L. Role of event runs in probability learning. *Psychological Review*, 1967, **74**, 410-419.
- GILMARTIN, K. J. An information processing model of short term memory (Doctoral dissertation, Carnegie-Mellon University, 1975). *Dissertation Abstracts International*, 1975, **35B**, 5664-B. (University Microfilms No. 75-9082)
- GILMARTIN, K. J., NEWELL, A., & SIMON, H. A. A program modeling short-term memory under strategy control. In C. N. Cofer (Ed.), *The structure of human memory*. San Francisco: Freeman, 1976.
- GREGG, L. W., & SIMON, H. A. Process models and stochastic theories of simple concept formation. *Journal of Mathematical Psychology*, 1967, **4**, 246-276.
- GUTHRIE, E. R. *The psychology of learning*. New York: Harper & Row, 1935.
- GUTHRIE, E. R. Psychological facts and psychological theory. *Psychological Bulletin*, 1946, **43**, 1-20.
- HINTZMAN, D. L. Exploration with a discrimination net model for paired-associate learning. *Journal of Mathematical Psychology*, 1968, **5**, 123-162.
- HUME, D. *A treatise of human nature. Book I: Of the understanding*. Cleveland: World, 1962. (Originally published, 1739.)
- HUNT, E. B. *Concept learning: An information processing problem*. New York: Wiley, 1962.
- HUNT, E. B. What kind of computer is man? *Cognitive Psychology*, 1971, **2**, 57-98.
- INDOW, T., & SUZUKI, S. Strategies in concept identification: Stochastic model and computer simulation. I. *Japanese Psychological Research*, 1972, **14**, 168-175.
- INDOW, T., & SUZUKI, S. Strategies in concept identification: Stochastic model and computer simulation. II. *Japanese Psychological Research*, 1973, **15**, 1-9.
- JAHNKE, J. C. Primacy and recency effects in serial-position curves of immediate recall. *Journal of Experimental Psychology*, 1965, **70**, 130-132.
- KRANTZ, D. H., ATKINSON, R. C., LUCE, R. D., & SUPPES, P. (Eds.). *Contemporary developments in mathematical psychology* (Vol. 1). *Learning, memory, and thinking*. San Francisco: Freeman, 1974.

- KUHN, T. S. *The structure of scientific revolutions* (2nd. ed., enlarged). Chicago: University of Chicago Press, 1970.
- LEVINE, M. *A cognitive theory of learning. Research on hypothesis testing*. Hillsdale, N. J: Erlbaum, 1975.
- MICHALSKI, R. S. Variable-valued logic and its applications to pattern recognition and machine learning. In D. Rine (Ed.), *Computer science and multiple-valued logic*. Amsterdam: North-Holland, 1975.
- MICHALSKI, R. S. Learning by inductive inference. In J. C. Simon (Ed.), *Computer oriented learning processes*. Leyden: Noordhoff, 1976.
- MICHALSKI, R. S. Knowledge acquisition through conceptual clustering: A theoretical framework and an algorithm for partitioning data into conjunctive concepts. *International Journal of Policy Analysis and Information Systems*, 1980, 4, 219-244.
- MILLWARD, R. B. Deviations of learning statistics from absorbing chains. *Psychometrika*, 1969, 34, 215-232.
- MILLWARD, R. B., & WICKENS, T. D. Concept-identification models. In D. H. Krantz, R. C. Atkinson, R. D. Luce, & P. Suppes (Eds.), *Contemporary developments in mathematical psychology* (Vol. 1). *Learning, memory, and thinking*. San Francisco: Freeman, 1974.
- MYERS, J. L. Sequential choice behavior. In G. H. Bower & J. T. Spence (Eds.), *The psychology of learning and motivation* (Vol. 4). New York: Academic Press, 1970.
- MYERS, J. L. Probability learning and sequence learning. In W. K. Estes (Ed.), *Handbook of learning and cognitive processes* (Vol. 3). Hillsdale, N.J: Erlbaum, 1976.
- NEWELL, A., SHAW, J. C., & SIMON, H. A. Elements of a theory of human problem solving. *Psychological Review*, 1958, 65, 151-166.
- NEWELL, A., & SIMON, H. A. The logic theory machine. *IRE Transactions on Information Theory*, 1956, IT-2, 61-69.
- NEWELL, A., & SIMON, H. A. *Human problem solving*. Englewood Cliffs, N.J: Prentice-Hall, 1972.
- NORMAN, D. A., & RUMELHART, D. E. *Explorations in cognition*. San Francisco: Freeman, 1975.
- RAAIJMAKERS, J. G. W., & SHIFFRIN, R. M. Search of associative memory. *Psychological Review*, 1981, 88, 93-134.
- RESTLE, F. The selection of strategies in cue learning. *Psychological Review*, 1962, 69, 329-343.
- ROUSSEL, P. *PROLOG: Manuel de référence et d'utilisation*. Luminy: Université d'Aix-Marseille, Group d'Intelligence Artificielle, 1975.
- SCHANK, R., & ABELSON, R. P. *Scripts, plans, goals, and understanding*. Hillsdale, N.J: Erlbaum, 1977.
- SCHNEIDER, W., & SHIFFRIN, R. M. Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review*, 1977, 84, 1-66.
- SHIFFRIN, R. M. Commentary on "Human memory: A proposed system and its control processes." In G. Bower (Ed.), *Human memory. Basic processes*. New York: Academic Press, 1977.
- SHIFFRIN, R. M., & ATKINSON, R. C. Storage and retrieval processes in long term memory. *Psychological Review*, 1969, 76, 179-193.
- SHIFFRIN, R. M., & SCHNEIDER, W. Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review*, 1977, 84, 127-190.
- SIMON, H. A. *The sciences of the artificial*. Cambridge: M.I.T. Press, 1969.
- SIMON, H. A., & LEA, G. Problem solving and rule induction: A unified view. In L. W. Gregg (Ed.), *Knowledge and cognition*. New York: Erlbaum, 1974.
- SIMON, H. A., & FEIGENBAUM, E. A. An information processing theory of some effects of similarity, familiarization, and meaningfulness in verbal learning. *Journal of Verbal Learning and Verbal Behavior*, 1964, 3, 385-396.
- SIMON, H. A., & NEWELL, A. Thinking processes. In D. H. Krantz, R. C. Atkinson, R. D. Luce, & P. Suppes (Eds.), *Contemporary developments in mathematical psychology* (Vol. 1). *Learning, memory, and thinking*. San Francisco: Freeman, 1974.
- SLOVIC, P., FISCHHOFF, B., & LICHTENSTEIN, S. Behavioral decision theory. *Annual Review of Psychology*, 1977, 28, 1-39.
- STERNBERG, S. Stochastic learning theory. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 2). New York: Wiley, 1963.
- THURSTONE, L. L. The learning function. *Journal of General Psychology*, 1930, 3, 469-493.
- THURSTONE, L. L. *Multiple factor analysis*. Chicago: University of Chicago Press, 1947.
- TRABASSO, T., & BOWER, G. Presolution dimensional shifts in concept identification: A test of the sampling with replacement axiom in all-or-none models. *Journal of Mathematical Psychology*, 1966, 3, 163-173.
- UNDERWOOD, B. J. Studies of distributed practice: VIII. Learning and retention of paired nonsense syllables as a function of intralist similarity. *Journal of Experimental Psychology*, 1953, 45, 133-142.
- WINOGRAD, T. Understanding natural language. *Cognitive Psychology*, 1972, 3, 1-191.
- WINSTON, P. H. Learning structural descriptions from examples. In P. H. Winston (Ed.), *The psychology of computer vision*. New York: McGraw-Hill, 1975.
- WINSTON, P. H. *Artificial intelligence*. Reading, Mass: Addison-Wesley, 1977.
- WYSOTZKI, F., KOLBE, W., & SELBIG, J. Concept learning by structured examples—An algebraic approach. In A. Drinan (Ed.), *Proceedings of the Seventh International Joint Conference on Artificial Intelligence*. Vancouver, B.C: 1981 International Joint Conference on Artificial Intelligence, 1981.

NOTE

1. I should not be so surprised. Estes (Note 4) points out that this emphasis on deterministic aspects of his theory is apparent in the three deterministic learning axioms of Estes and Suppes (1974, p. 171). See also the discussion of the noncontingent Markov reinforcement schedule constant presentation set (Estes & Suppes, 1974, p. 177).