MATHEMATICAL LEARNING THEORY AND THE NEW "MENTAL FORESTRY"[1]

196

JAMES G. GREENO AND ROBERT A. BJORK
Human Performance Center, The University of Michigan, Ann Arbor, Michigan

In the decade or so that constitutes the period of our review,[2] progress toward understanding the processes involved in learning can only be described in terms that sound extravagant. It is our task to consider only one part of this remarkable progress, but in order to put our review in perspective, we begin with a brief sketch of our impression of the current condition of the psychology of learning and cognition.

A little more than 10 years ago, experimental psychologists in the field of learning seemed concerned primarily with changes in the frequency of certain response classes over trials as a function of experimental conditions. Since that time a new psychology of cognition has grown up, at least to the point of adolescence, if not young maturity. While it is important to remember that occasional far-sighted theorists and commentators since James (115) and before have seen clearly that cognition is a complex and worthy subject, it is nonetheless true that the emergent techniques and interests in the study of cognitive processes such as attention, encoding, search strategies, rehearsal processes, and understanding constitute a kind of revolution.[3]

[1] Preparation of this report was supported in part by the Air Force Office of Scientific Research under Contract No. F44620-72-C-003 with the Human Performance Center, Department of Psychology, University of Michigan; and in part by National Science Foundation Grant GB-31045.

[2] Mathematical learning theory has not been the topic of an earlier contribution to the *Annual Review*. We date our review mainly from Estes' review of learning theory in 1962 (65), although we have omitted some material dealt with by Millward in 1964 (24), and our review overlaps to some extent Hunt's review of computer simulation in 1968 (108). Our literature review ends with the calendar year 1971.

[3] Though not in the sense of Kuhn (130), whose concept of a revolution seems to us to be quite irrelevant to the analysis of scientific progress. We consider the new developments of the 1960s as completely continuous with the development of mathematical learning theory in the 1950s by many workers, including Estes (59, 62), Bush & Mosteller (37, 38), and Suppes & Atkinson (212), and although we did not participate in those developments, it seems to us that they were quite continuous with earlier well-known work such as that of Hull (106). Virtually all of the work reviewed here has been contributed by scientists trained by those who built psychology during what we and others would call the prerevolutionary period of cognitive psychology, or by those prerevolutionary scientists themselves. In our judgment, what we are reporting is *not* a paradigm shift, whatever that might be.

This revolution has influenced the study of learning very substantially. In particular, learning has been subjected to an analysis of processes such as attention, storage, and retrieval that underlie the overall changes of state that are labeled learning. Major innovations have occurred at levels involving general conceptualization, theoretical method, and experimental technique. Workers in the field are all aware of the new experimental paradigms that have been devised to isolate and examine particular processes, such as sensory storage of briefly exposed visual information (2, 71, 206), attentional and short-term storage processes in audition (34), short-term retention of verbal items (35, 173, 174, 201), encoding processes and the role of similarity (43, 179), processes of retrieval from both short- and long-term memory (42, 133, 148, 209), and the learning of concepts and rules (99, 190), to mention just a few.

Along with these developments in experimental methods, several formal developments have provided increased power for theoretical work. In his review of literature prior to 1962, Estes (65, p. 111) correctly foresaw the importance of developments in mathematics by which "psychologists have had placed in their hands the simple but powerful methods of finite Markov chains," which have provided the formal basis of much substantive analysis. Of similar importance have been continued developments in artificial intelligence, stimulating numerous efforts to understand psychological processes with the aid of computer simulation (110).

The most important theoretical development, in our opinion, has been the evolution of concepts for describing and analyzing psychological processes and structures. A notable feature of many theoretical contributions has been the assumption of considerable structural complexity. The structural properties of systems for processing information and memorizing have been considered in several theories (9, 161). Theorists have assumed a hierarchical process of feature extraction, with elementary properties analyzed initially, and more complex properties or names of stimuli processed or stored at higher levels (137, 197, 199). Models of problem solving have incorporated multilevel functioning, with recursive methods and executive control of processes (58, 184). Tree structures have been proposed to represent what a subject learns when a list of verbal items is memorized (76, 101), and when a classificatory concept has been induced during concept learning (108, 112, 229). And the problem of serial order has been approached with increased sophistication, with representations of sequential concepts as hierarchies of rules or transformations (189, 190, 204, 233).

A decade ago it was still appropriate to characterize the psychology of learning as the study of "mental chemistry," that is, the investigation of processes by which elements combine into more complex compounds. We sense that the study of learning is in a transition stage in which our concerns are moving toward an interest in the processes by which trees and other cognitive structures are acquired and modified. Perhaps "mental forestry" suggests a better analogy for characterizing the current transition stage, if not the final version of the psychology of learning that is emerging. At least, if we are still mental chemists, we are

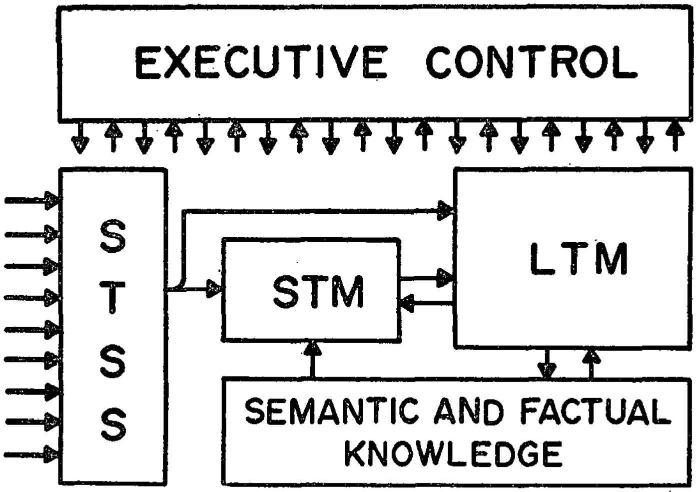

FIGURE 1. Divisions of information-processing function in the human memory system.

dealing with structures comparable to those of organic rather than inorganic compounds.

In the new psychology of cognition that has developed in the last decade or so, it is customary to represent some main divisions of function in a flow diagram, and Figure 1 shows the main components of what we take to be the current consensus. Information enters the system through sensory registers, where it is held in short-term sensory storage (STSS), but is lost very quickly—usually within fractions of a second—unless it is processed further into the system. Most theorists assume that further processing consists of attending to the sensory information in STSS, which has the effect of transferring information to short-term memory (STM). Short-term memory has a relatively small capacity of a few items or chunks of information, and items are apparently stored most frequently if not always in a form appropriate for rehearsal, with phonemic features playing a prominent role. Items are typically held in STM for times on the order of seconds, and retrieval of information from STM is rapid and reliable.

By long-term memory (LTM) we mean to denote a system of relatively large capacity, at least up to dozens of items. Retrieval from LTM is problematic, and it appears that much information apparently lost actually remains in the system but becomes inaccessible. The time scale we have in mind regarding LTM is on the order of a few minutes up to several hours. Examples of learning that involve LTM include memorizing a list of words during an experiment, or retaining information about experimental procedures while reading the results section of a scientific article. By semantic and factual knowledge, we refer to the person's store of knowledge about concepts and past events that is virtually permanent.

Terminology is troublesome here, since several authors have referred to this kind of permanent store as long-term memory, with considerable justification. But we have chosen the usage described here in order to make maximal contact with the existing literature.

Our reference to "executive control" in Figure 1 recognizes the importance of attention and decision processes in the storage, rehearsal, and retrieval of information. A subject may decide to ignore an item in STSS rather than enter it in STM (9). Contact with the lexical contents of semantic and factual knowledge involves a decision rule for determining whether enough information has been perceived to justify assigning a name to the stimulus (137, 197). There are mechanisms for selecting items held in STM for rehearsal (9, 17, 183), and for selecting attributes of stimuli for special attention and processing (40, 185, 239). Developments of organized groupings of items in LTM and strategies of searching in LTM for items (124, 202) are further functions carried out under executive control. When information is retrieved, a decision has to be made whether the information is adequate to justify a response (16, 29, 122, 161).

The learning theories that are the subject of this review deal with the process of acquiring information and remembering it. Using a criterion that is fairly standard, albeit arbitrary, we consider learning as a process resulting in retention of information for at least a few minutes. Thus, in relation to Figure 1, we consider theories dealing with the process of storing and retaining information in LTM. We omit consideration of STSS entirely, and our concern with STM is limited to processes of short-term retention that influence storage of information in LTM. Modifications in the structure of a person's semantic and factual knowledge certainly constitute learning, but rigorous theoretical analyses of such processes have not yet appeared. Another limitation on this review is imposed by the fact that nearly all quantitative theorizing during the last decade has dealt with the storage and retention of verbal information; we therefore have little to say about either animal learning or about the learning of skills. Animal learning has received limited attention by quantitatively oriented theorists during the last decade, and although quantitative analyses have been carried out with respect to the properties of skilled performance, we are aware of very little quantitative theorizing on the acquisition of skills, with Timpe's work (223–225) being the main exception that we know of.

On the question of what constitutes a "mathematical" learning theory, we are neither certain of our criteria, nor do we think, in spite of our efforts, that we have been completely consistent. In general, we have assumed that mathematical theories consist of hypotheses stated in a relatively formal way together with implications drawn from those hypotheses by means of mathematical reasoning. Nearly all of the quantitative theorizing covered by our review employs either probability theory or computer programs as the mathematical basis. But the relative importance of mathematical work in a contribution to theory varies a great deal. We are aware that other reviewers would probably apply quite different criteria for inclusion in this area. Since we are primarily interested in learning theory rather than mathematics, we think we have been quite conservative in excluding "nonmathematical" contributions, but it seems certain that others would include some of the contributions that we have failed to mention.

The remainder of this review presents our impression of the current state of theoretical understanding of learning processes as represented in relatively rigorous (i.e. "mathematical") theories. Our discussion in the next section is organized around the question: What is learned? Our discussion is focused on the problem of representing what is stored during learning, but we also comment briefly on the influence of forgetting and retrieval processes on acquisition and performance. In the third section we review theoretical analyses of specific varieties of learning, and in the final section we close by reviewing some analyses of formal properties and statistical methods.

WHAT IS LEARNED?

As an outcome of learning, information is stored in memory that was not there before, or some structure of knowledge is modified. Theoretical representations of what is stored during learning seem to be of two kinds: either information stored in memory is assumed to consist of representations of the items studied, or the contents of memory are assumed to be a set of rules for generating a response or a series of responses. The difference may be more apparent than real. We suppose that theorists who refer explicitly only to storage of item representations are being elliptical, since it is obvious that there must be some procedure for translating the stored information into responses. On the other hand, theories assuming that rules for response generation can be learned without any concomitant learning of item information tend to overlook the evidence that subjects can often respond correctly on tests that require the retrieval from memory of information about specific items.

Item Representations

Storage.—Theories that assume items rather than rules are stored vary on two dimensions. One dimension is whether items are assumed to exist in memory in a unitary way or are assumed to have a more complex representation, usually involving feature or attribute descriptions. A second dimension is whether the representation in memory is all-or-none or varies in strength along some discrete or continuous scale.

In the simplest combination of assumptions, an item is stored as a unit, and storage is an all-or-none process. This assumption is made in the all-or-none Markov models in which the learning of an item consists of a discrete change in state. The simplest case involves a two-state system in which the only distinction made is whether an item is learned or unlearned (27, 64). To analyze certain kinds of transient effects, a third state has been introduced corresponding to STM. An item may be stored in STM temporarily after study, with some probability of returning to the unlearned state. Entry into the learned state corresponds to achieving a record of the item in LTM (7, 86, 121, 235). Unitary, all-or-none storage is also assumed in analyses based on queuing theory (28, 183), where STM is assumed to consist of a queue of items waiting to be processed, and storage in LTM results when an item is processed, which occurs when the item reaches the first position in the queue. In his fixed-point model, Murdock (153, 154) also assumes all-or-none storage, but whether an item is retrievable or not after a given retention interval is determined by a fluctuation process assumed to reflect the interaction of forgetting and reminiscence.

Another class of theories assumes that storage involves the representation of an item in a unitary way, but that the strength of the representation is a variable. In one version, used by Bernbach (17, 18), storage consists of creating copies of an item, so an item may be represented by any number of copies in memory. In Bernbach's theory there is no separate system of STM, but the probability of adding copies by rehearsal decreases with time after an item's presentations, due to the increasing chance that all the copies of the item have been lost. Another theory that assumes unitary storage with varying strength is Wickelgren's (237) trace theory. Wickelgren assumes that the representation of an item consists of several traces, each varying in strength as a function of time since the item was studied. The several components of the memory system (Wickelgren assumed there are four) are represented by traces with different rates of consolidation and decay. Another variation was given by Atkinson & Shiffrin (9), who assumed that occupancy of STM is all-or-none, with a finite upper limit on the number of items that can be in STM at any time. Storage in LTM is assumed to involve representation of unitary items, but the strength of the representation is a continuous variable referred to as the amount of information in LTM for an item, with the amount of information transferred to LTM being proportional to the amount of time that the item was in STM. The idea that recognition depends on a decision process based on the familiarity or strength of an item's trace also uses the idea of a unitary variable-strength representation (16, 50, 100, 122, 146, 172, 238).

The second main alternative assumption about storage is that the representation of an item is not unitary but consists of a representation of some of its features. The simplest analyses using this idea are the N-element models of stimulus sampling theory (70, 214) where it is assumed that a stimulus situation consists of many elements or presents different potential perceptual patterns. At a specified stage of learning some proportion of the elements or patterns are associated with a given response in the subject's memory; at the limit, if a single response has always been paired with the stimulus, all the elements or patterns of that stimulus are represented in memory in association with the response. The assumptions of the mixed pattern-components model (8, 81, 83) are that all the components of a stimulus are connected with a response in memory, and if a subset of one or more of the original components appears as a test stimulus, the response will be performed. Thus the mixed model includes the implicit assumption that stimulus features or components are represented in memory. In a similar way, Bower's (29) multicomponent theory of the memory trace assumes that a set or vector of stimulus properties is stored in memory. Bower's analysis deals with the process of losing elements from the stored vector through decay or interference during a retention interval and thus differs from the analysis given in stimulus sampling theory, which deals with the process of building a set of stored components during training.

In a sense, the fluctuation model formulated by Estes (60, 61) provides a theory of memory storage that generalizes the two ideas given in ordinary stimulus sampling theory and in Bower's multicomponent analysis, since the fluctuation model considers the process of storing a set of component associations for a stimulus as well as the loss of those components over a retention interval. In the fluctuation model, Estes assumed that a subset of the potential elements for a stimulus are available for association with response in any situation, but over a retention interval the elements in the available subset are exchanged with elements that were unavailable originally, thus causing elements associated with the response in memory to become unavailable at a later test. Elements are not permanently lost in the fluctuation model; they may fluctuate back to the available set at a later time.

In some theories dealing with stimulus features, properties of certain features have been specified. In a theory given by Restle (187) representation of an item initially stored in memory may not be distinctive enough to permit discrimination from similar items. The representation needed includes the feature that distinguishes between similar stimuli or responses. Another theory specifying the nature of features represented in memory was given by Laughery (137), who specified a set of visual and acoustic features for letters and numerals and assumed that these are stored as part of one's permanent knowledge. Laughery's theory, like Bernbach's described above, does not distinguish between STM and LTM as separate systems. In Laughery's analysis, acoustic features of presented items are represented in memory and are updated and strengthened by rehearsal. Another theory that assumes a representation of features in the subject's permanent knowledge was given by Norman & Rumelhart (161). Their theory of storage uses the idea of contextual association; they assume that when an item is recognized in study, its features are tagged in memory showing the context in which it was presented. The information stored in memory according to Norman & Rumelhart's theory consists of associations between features of items and contextual tags. Association between features or properties and responses is the assumed nature of information storage in classical discrimination learning theory of the Hull-spence variety (207, 241). Other kinds of information assumed to be stored about an item include the frequency of its occurrence (20) and the time of its occurrence (202, 203), and the relative position of the item on some quantitative scale such as size (30).

Forgetting.—Forgetting is not a concern of this review except to the extent that it influences acquisition processes. Several different types of interactions between acquisition and forgetting are found among different theory types. In those theories assuming a unidimensional memory trace that varies in strength or number of stored elements, it is natural to assume that forgetting consists of a loss in strength or number of stored elements. After a series of learning trials, performance is determined by the net effect of the acquisition of strength on the learning trials and the loss in strength between learning trials. This interpretation of the interaction between forgetting and acquisition, although quite appealing intuitively, has the considerable defect of being nearly always completely wrong in its predictions of the effects of the temporal spacing of learning trails. Given any reasonable learning operator, such an interaction leads to the prediction that learning is optimized when the spacing between trials is minimized [for a discussion of this issue, see Bjork (23)].

In models in which what is stored can be in more than one memory state, such as STM and LTM, or available and unavailable, the interaction between forgetting and acquisition is more complex. Under some circumstances the interaction can be such that long-term learning profits from short-term forgetting. This kind of interaction can occur in models of the rehearsal-consolidation type (9, 132), of the stimulus fluctuation type (60), and of the multistate Markov type (86). In these theories, increased forgetting between any two learning trials reduces performance at that point in the learning process, but long-term learning may profit from either increased consolidation of long-term memory as the temporal spacing between two learning trials is increased, or from improved long-term acquisition on the second of the two learning trials. Long-term acquisition might be facilitated on the second learning trial in two ways. First, the number of formerly unconditioned and unavailable stimulus elements that become available and conditionable through fluctuation could increase with the temporal spacing between trials. On the other hand, the probability of transition from a forgotten state to long-term memory could exceed the probability of transition from short-term memory to long-term memory, and a longer interval between two learning trials would decrease the likelihood that an item would remain in short-term memory at the time of the second learning trial.

Retrieval.—In the foregoing theories of representational storage, retrieval processes sometimes do and sometimes do not play a major role in determining performance. When an item is assumed to be stored as a unit, there is not much basis for a theory of retrieval, although retrieval probability might be assumed to vary with number of copies stored, or, under the assumption that an item be stored as a unit in one or more different memory states, retrieval of an item might be assumed (89, 243) to vary with its state in memory.

When an item is assumed to be stored in terms of its features or trace strength, there is a natural basis for specifying retrieval mechanisms. Given that a feature list in memory incompletely specifies an item, different assumptions about the retrieval process yield different predictions of both frequency and types of errors on tests of recall and recognition (29, 137, 161). If a memory trace is assumed to vary in strength on a continuum, recall and recognition may be assumed to depend on decision mechanisms, such as that embodied in the theory of signal detectability (238, for example).

When judgments of list membership, recency, or frequency of occurrence are required, retrieval mechanisms play a critical role. There is no reason to expect that the retrieval mechanisms involved in such judgments tend to be simple; it is almost certain that comparisons among items, judgments of strength, relative position in structural organization, and encoded tags and labels are all involved.

Comments on alternative assumptions.—A review of alternative theories inevitably raises the question of which is correct. On the dimension of unitary representation versus representation of features the question is easy to answer. A theory that assumes representation of components or features of stimuli is more general than a theory assuming that each stimulus record is unitary, and several well-known facts about memory indicate that the more detailed assumption is needed in many situations. Any process of selection or variability of encoding must operate on a basis of multicomponent representation of some kind, and the operation of selective and variable processes in encoding seems quite well established. On the other hand, assumption of unitary representation is a useful special case of the more general idea, and can be used to represent the process of learning in many situations where the complicating factors of selection and variable encoding can be neglected.

The issue regarding all-or-none storage versus variable strength in the memory representation is considerably harder to evaluate. Even when unitary representation is assumed, it turns out, as Restle (188) showed, that all the properties of observable performance produced by all-or-none learning can be mimicked by a system in which learning is gradual, performance depends on a threshold, and an appropriate distribution of individual differences exists. This means that a definitive choice in favor or all-or-none learning cannot be made on empirical grounds, since the all-or-none hypothesis is empirically equivalent to a special case of gradual learning. If it is assumed that representation involves features of items, the issue becomes even harder to decide, since even if all-or-none storage of individual features is assumed, the possibility of partial representation implies that learning of items may not be all-or-none in nature. The question of deciding whether representations of individual features are stored in an all-or-none fashion involves the same formal difficulties as apply to that question regarding individual items, except that the decision has to involve several theoretical entities instead of just one.

While a definitive empirical case for the assumption of all-or-none storage seems unattainable, there appear to be strong theoretical and methodological reasons supporting the use of that idea. The main advantage of the all-or-none idea is that it is testable in Popper's (178) sense and therefore can serve as the basis of theoretical advances in which changes in the theory are guided directly by empirical evidence. Thus it continues to seem preferable to view cases in which storage of information appears to occur in a gradual fashion as cases in need of more detailed analysis, as Estes (66) suggested in 1964. This attitude has led to the analysis of several tasks as combinations of all-or-none processes. On the other hand, analyses based on the weaker assumption of gradual changes in strength of representation undoubtedly will continue to provide useful information, and results obtained under the two kinds of assumptions probably can be related to each other in future analyses.

Generative Representations

Storage.—In many theories the information stored in memory is in the form of rules for selecting or generating responses, rather than representation of events. The case studied most intensively involves rules for classifying stimuli in concept identification, but considerable analysis has also been given to the structure of sequential concepts, where the stored rule is a means of generating a sequence of responses.

In the case of classification, the outcome of learning is a rule or strategy that assigns a response to each stimulus in the set used in the experiment. There are two questions about what kind of information is stored in memory. First, what is the form in which the rule or strategy is represented? This is the direct question of what is learned in concept identification. The other question is, what kind of information is held in memory during learning? This second question involves issues about the retrieval of information from the store of permanent knowledge as well as the kind of information held in STM for use in processing information.

Regarding the first problem as to the way classification rules are represented, there have been two main assumptions. One assumption is that a concept is stored in the form of a list of properties, along with the name of the concept (116, 184). A second assumption is that the relevant properties of the concept are stored as a rule for making a series of tests and a decision (108, 109). For example, if a subject has learned that the concept GAX refers to the category of large, blue stimuli, the list representation could be in the form "GAX: large, blue," combined with the response rule that all entries on a property list should be matched by a stimulus for the name to be assigned (184). On the other hand, a decision tree for this concept might be:

Test color; if blue, proceed; if not blue, respond "not GAX."

Test size; if large, respond "GAX;" if not, respond "not GAX." The difference is obviously one of notation. The list of properties and an appropriate response rule lead to the same performance as the decision tree. In fact, any decision tree for a concept can be represented as a matrix in which each row is a list of properties (112). In the simplest case of concept identification, involving a single relevant attribute, some theories refer to selection of an attribute and association with a response as separate processes (32), while others refer to the selection of a strategy, which specifies a response for each stimulus in the set being used (185). The difference does not seem to be a substantive one. In more complex situations that require learning to classify patterns that have variable features, theorists have assumed that the importance given to the different features is modified on the basis of information received during learning. Thus the outcome of learning not only involves selection of attributes for testing, but also involves setting weights on the various tests for deciding to which category a presented pattern belongs (151, 200, 230), or in evaluating one's position in a game (198).

The issue as to the kind of information held in memory during processing *does* involve assumptions that differ in a substantive way. In several theories, the subject remembers a set of stimuli and their response assignments, and hypotheses about the concept are selected on the basis of scanning the remembered set of stimuli (45, 111, 116, 152, 227). In other theories, information is held in STM in the form of a list of hypotheses that are being considered (79, 118, 157, 185, 228) and there may also be memory for hypotheses that have been eliminated from past samples (10, 57, 96, 239). Chumbley (40) obtained experimental evidence on the question favoring the idea that it is hypotheses that are stored, rather than representations of stimuli.

Regarding the question of accessing information from permanent storage, nearly all theories have assumed a constant set of descriptors (for the stimulus-representation models) or possible hypotheses (for the hypothesis-list models) used by the subject during learning. When information is taken from permanent storage (usually when the current list of hypotheses is exhausted, or when an error in classification is made) the selection from the set in permanent storage is usually assumed to be random, with fixed probabilities for the various possible hypotheses. One interesting exception is the hypothesis of Falmagne (74, 75), who assumed that the sampling probability of an hypothesis is increased on trials when that hypothesis is consistent with information given, and decreased when it is inconsistent.

The assumption that what is learned is a rule for generating responses, rather than a representation of events, has been postulated in several theories about the learning of lists of verbal materials. One straightforward case involves lists of paired associates that include sets of similar stimuli paired with a single response. These tasks permit the subject to group items and use a rule of the form, "Give response R to stimuli with property P," in a fashion similar to standard concept identification. Theories dealing with this rule-learning aspect of associative learning have been developed. Batchelder (11, 12) developed the idea of a classification rule explicitly in his theory; other investigators have dealt with the problem by postulating various mechanisms of transfer between similar items represented in memory (8, 81–83, 92, 186, 187). Another possibility regarding list learning is that the subject's knowledge about the entire list is acquired in the form of a discrimination net. This idea, proposed by Feigenbaum (76) and used in an extensive study by Hintzman (101), proposes that information about a list is stored in the form of a decision hierarchy with each node of the tree being a test on some attribute of items in the list. The selection of response depends on the outcome of the series of decisions, just as in the analyses of concept identification described above.

Another important use of hierarchical representation of the outcome of learning has been in analysis of serial learning. Feigenbaum & Simon (78) used the idea of a discrimination net to represent the sequential learning that occurs when a person learns a list of words. Analyses of the acquisition of sequential concepts when subjects are shown series of numbers, letters, or patterns of switch settings have been given in a number of investigations (95, 189, 190, 204, 233). The various analyses share important features. A process of encoding relationships between elements is assumed, often involving the formation of units based on short subsequences. Then rules for combining the subunits are acquired, resulting in a hierarchical structure of rules for generating the sequence. Evidence obtained by Bjork (22) suggests the pleasant possibility that in at least some nontrivial cases, storage of components of a hierarchical structure is all-or-none in nature.

Forgetting.—In some generative representations, the structure that must be acquired is complex enough that forgetting can be assumed to influence both performance and the ongoing acquisition of the representation. Very often this is not the case; many of the typical concepts, rules, or structures involved in experiments are simple enough that they do not pose memory problems—for example, a concept such as "Give response R to stimuli that are red" is not difficult to remember. In the case of certain sequential concepts (22), however, and in the case where what is acquired is a tree structure or discrimination net (101), retaining the representation in memory is not trivial and may influence heavily the acquisition of the representation.

Even in those cases where the generative representation to be learned poses no memory problems in itself, the retention of the information content contained in a series of learning trials may constitute a significant memory load. Acquisition in such situations may, therefore, depend on memory in an indirect but critical way; the decisions, tests, and other operations involved in the learning of a generative representation can be effective only to the extent that the item information necessary for learning to take place is retained. The importance of memory for specific items is sometimes nullified procedurally by providing subjects with displays that relieve them of retaining such information (204).

Retrieval.—When what is stored is generative in form, retrieval from memory does not tend to be a problem. Given that a response rule has been learned, retrieving the response is typically straightforward, with two possible exceptions. One exception is in the learning of sequential concepts, in which case generating the series of responses required may be subject to failures of retrieval from memory. A second and more interesting case is the learning of lists of verbal materials by means of the storage of tree structures. In such cases retrieval becomes very important, especially in the case where the structure is incomplete. Both frequency and types of errors are influenced heavily by what is assumed about the tests and decisions involved in the retrieval process.

ANALYSES OF SPECIFIC LEARNING TASKS

The discussion in the preceding section deals with general principles that are postulated in current mathematical theories of learning. Any application of those principles takes the features of a specific learning task into account and specifies the processes of information storage, retention, and retrieval that occur in the situation being analyzed. Of course, the various experimental situations that have been developed for the study of learning provide information about different aspects of the learning process. The theoretical framework that we have adopted, described at the beginning of this review, has consequences for the interpretation we give of results obtained in various experimental settings. Within that framework we now review published analyses of various tasks, and while this kind of review inevitably takes on something of the character of a catalogue, we hope to indicate relationships between these specific analyses and the substantive issues discussed earlier.

Learning to Recognize

Discrete state models.—When a list of items is shown and later tested for recognition, information about learning is obtained with minimal requirements on the subject for retrieval of information. Kintsch & Morris (125) analyzed multitrial recognition learning of trigrams and obtained data agreeing with the all-or-none learning model. Kintsch (121) and Olson (171) investigated effects of varying intervals between presentations of items during recognition, and obtained results compatible with the all-or-none model, elaborated by the addition of a state representing short-term memory. Although Olson (171) found some evidence that items could be lost from LTM, it is a good approximation to assume that when subjects study items in recognition learning, each trial provides an opportunity for a stable, distinctive representation to be stored in LTM, and failing that, a representation is held in STM for a time.

Analyses based on decision theory.—While performance in multitrial recognition learning has the all-or-none property, an important factor is that subjects give just "yes" or "no" responses in that experiment. When subjects also give confidence ratings about their responses, a more complicated process is involved. Ideas taken from the theory of signal detectability have been used, assuming that the strengths of memory traces or feelings of familiarity vary among items stored in memory, and judgment of confidence indicates the amount of this strength or familiarity. The strengths of new or distractor items have a distribution analogous to the distribution of likelihood ratios on noise trails in detection. In analyses by Parks (172) and by Wickelgren & Norman (238) strengths of presented items are greater, on the average, and the amount by which the mean strength of presented items exceeds the mean of the distractor distribution depends on the time since presentation.

Freund, Loftus & Atkinson (80) analyzed recognition performance on number-letter pairs, using Atkinson & Shiffrin's (9) model of the storage of information about items in memory. The model includes an assumption that the amount of information stored is proportional to the time an item resides in STM, and the amount lost is proportional to the time elapsed between storage and test. Two alternative assumptions about retrieval were compared. It appeared preferable to assume that the amount of information determined a value of d' for a decision process like that used in the theory of signal detectability rather than to assume all-or-none retrieval of items with the probability of retrieval determined by the amount of information retained. Donaldson & Glathe (50) also discussed application of detectability theory to analysis of recognition and recall, especially regarding interpretation of the value of d' obtained in recall performance.

In analyses by Bernbach (16, 19) and Kintsch (122) a single distribution of strengths associated with items with stored representations is assumed. An item that has been presented has some probability of being stored, and as other items are presented and tested there is a possibility that the stored representation is lost. Thus in this analysis some of the presented items have strength taken from the distribution of stored items, and others come from the same distribution as new items. Kintsch's (122) analysis adds the feature that if an item's stored representation is sufficiently strong, it will not drop down into the initial distribution—in other words, it enters the learned state of the Markov model of learning.

Analyses assuming storage of features.—Further analyses of recognition have been provided, based on the idea that an item's representation in memory is a partial list of features. In his presentation of the multicomponent model, Bower (29) gave analyses of several kinds of recognition experiments, using the idea that some features of an item's representation are lost during a retention interval. Response on a recognition test depends on the number of features of a test item that match features retained in memory, relative to a criterion that the subject has, or compared with the number of matching features of other items presented. A related analysis given by Norman & Rumelhart (161) assumes that storage consists of tagging features of the studied item with context information. During a retention interval in which other items are presented, some of the tags of an item's features are tagged by new contexts and this produces uncertainty about the item when it is tested. In a test of recognition, the features of the tested item are scanned, and response depends on the number of features found with appropriate context tags.

Simulations of recognition.—Computer simulations of pattern recognition have provided important contributions to theory about the processes involved in learning to recognize. Hunt's (110) review presents a recent summary of the main ideas. Two hypotheses are particularly interesting for the theory of learning. One is that learning to recognize a character involves storing the features of each character presented during learning and computing the probability that a given feature comes from each character in the practice set. These probabilities are used when a test character is presented to make a decision about which character is being shown (200). A second idea is that characteristics consisting of local patterns are stored as a feature list for each character that can be recognized. Recognition involves an effort to match the stored characteristics with features of the presented character, and experience with known characters leads to adjustment in the weights given to the various characteristics in the decision about what character is being shown (230). General discussions of pattern recognizing systems and their learning capabilities have been given by Nilsson (159) and by Minsky & Papert (151).

Learning to Retrieve Lists

Free-recall memorizing.—Analyses of performance in free-recall memorizing by Kintsch & Morris (125) and by Waugh & Smith (236) have shown that memorizing items for recall is not all-or-none in nature, and that a two-stage Markov model gives an acceptable account of the data. After recognition pretraining, recall memorizing was an all-or-none process, suggesting that the whole process of memorizing for recall involves a first stage of storing a representation of the item, followed by a stage of learning to retrieve the item from memory.

The idea that stored representations involve features of items has been used in analyzing recall. In Bower's (29) model it is assumed that each item in memory is recalled if enough of its features are retained to satisfy a recall criterion. Missing features are filled in by the subject. A similar analysis of recall was given by Norman & Rumelhart (161), who assumed that in a recall test context information is made available and memory is accessed to find features of items studied in the context. If a sufficient number of features are retrieved to spedify an item uniquely, the subject gives that response. Otherwise, no response is given or the subject guesses.

Other discussions of recall by Kintsch (124) and Shiffrin (202) use the idea that items are stored in memory in a format involving features, including information about the time of study, and these features serve as a basis for search and retrieval of items. Cowan (47) gave a rigorous analysis of retrieval in the case where items are taken from two categories, based on the idea of differing associative strengths between items. And Albert (1) has analyzed interresponse times in the output of a learned list in terms of a linear death process, which he applied in a continuous form and also as a discrete-time urn model.

The role of short-term memory.—The role played by STM in memorizing for recall has been investigated in several studies. A major achievement of these analyses has been the development of a rigorous explanation of the serial position effect. Primacy is explained by the occurrence of extra processing of items studied at the beginning of the list, when STM is relatively uncrowded. Recency is explained by the fact that items studied at the end of the list are likely to still be in STM at the time of the test. Particular models embodying this general explanation differ to some extent. Waugh & Norman (235), Atkinson & Shiffrin (9), Norman & Rumelhart (161), Bower (28), and Reitman (183) assumed that STM is a discrete stage of processing, while Bernbach (17, 18) and Laughery (137) postulated a single memory system with a rehearsal process that has the property of making rehearsal more likely for recently presented items. Another difference is that Norman & Rumelhart (161) and Laughery (137) assumed that items are represented in STM as feature lists, thereby providing an explanation for effects of similarity, especially regarding intrusions. An alternative analysis of the serial position effect, based on concepts of retroactive and proactive inhibition, was given by Kuno (131). And Thomas (222) analyzed serial position effects with concepts from information theory.

Serial recall.—When the subject's task includes recalling items in the order of their presentation, theoretical analysis requires assumptions about the way in which information about order is stored and retrieved. In Laughery's (137) model, the data structure representing each item includes a substructure that holds the name of the following item. Thus order information is represented as a feature, capable of being lost from memory in the same way as other features of the item. Feigenbaum & Simon's (78) analysis of serial recall, based on the idea of a discrimination net, assumes that when a series of tests has been carried out, terminating at a node containing a response image, information that is stored along with the response image specifies the series of tests to be conducted to obtain the next item. And Lander (134, 135) has analyzed interassociations in serial learning from the standpoint of McGeoch's remote associates hypothesis.

Serial pattern learning.—We have all been aware, at least since Lashley's (136) convincing argument, that performance of many serially ordered tasks depends on an organization more sophisticated than item-to-item linkages. Vitz & Todd (233) have analyzed learning of binary sequences, concluding that events are encoded as runs of homogeneous events, and the structure that is acquired is a hierarchy involving runs of runs at each level. Restle (189) and Restle & Brown (190) have analyzed serial pattern learning of sequences of the numbers 1–6. Subsequences of short runs (e.g. 2345) and trills (e.g. 2323) are frequent bases of encoding, and sequences with hierarchical structures of these subunits are easily learned. Hierarchical structures are formed when movement from one subunit to the next can be accomplished using a transformation such as repetition, transposition (repeat the subunit with each element moved up or down by some fixed amount), or reflection (repeat the subunit, but with each element replaced by the element equally far from the opposite end of the array).

Simon & Kotovsky (204) analyzed the induction of a pattern from part of a letter sequence, postulating that subjects induce rules for generating the sequence based on relationships of identity and adjacency in the alphabet between entries in the sequence that are adjacent, or separated by one, or by two, and so on. Gregg's (95) analysis of sequential concept acquisition using patterns of switch settings provided information about the ways in which relations between patterns are encoded, as well as generating rules varying in complexity. Gyr, Brown & Cafagna (97) investigated models for inducing sequential patterns differing in the level of abstraction of hypotheses considered by the subject. Bjork's (22) analysis of the learning of arithmetic sequences assumes that rules induced by the subject may be constant (add or subtract *n*) or advancing (add or subtract one more than last time), and each component rule in the sequence is added to cognitive structure in an all-or-none fashion.

LEARNING TO RETRIEVE ITEMS ON CUE

After a paired associate is memorized, the subject is able to retrieve and perform the response term of the pair when the stimulus term is presented. Recent studies, following on prior analyses showing that in many experimental procedures memorizing of paired-associates is approximately an all-or-none process (see Estes 66), have provided more detailed analyses of the process of memorizing, or have considered situations in which the learning process has more than one stage.

The role of short-term memory.—Several analyses using the concept of rehearsal in STM have provided information about the way in which subjects adjust rehearsal procedures in relation to aspects of the learning situation. In continuous memorizing when the response for a stimulus is changed, the new item replaces the old one in STM, providing little interference with rehearsal of other items, although when an item with a new stimulus replaces the tested item it has approximately the same interfering effect as a new item (6). With a procedure requiring overt rehearsal, subjects enter all items in STM, and effectiveness of rehearsal is improved in the sense that the rate of information transfer to LTM is increased (9). A particularly interesting result was obtained in an experiment by Loftus reported by Atkinson & Wickens (10). Procedures were compared in which subjects either had to recall a letter response or had the easier task of selecting one of two alternative letters presented to them. The analysis indicated that with the two-choice procedure, subjects held only one item in STM at a time, entered nearly all presented items into STM, and rehearsed relatively efficiently; while with the recall procedure, subjects held about three items in STM at a time, about one-half of the items presented were not entered in STM, and rehearsal was less efficient.

As an approximation, STM can be represented as a Markov state, with an item assumed to occupy that state as long as it is in STM (7). This idea has been used in analyzing the greater difficulty of learning paired-associate items in longer lists (39), in analyzing perseverative errors in associative learning (14) and verbal discrimination (36), and in investigating the effect of spacing between the presentations of an item (23, 86). To analyze the effects of test trials given at varying intervals after study, Young (243) has postulated a system with two levels of STM, one of which involves probabilistic retrieval, and Izawa (114) has used the assumption that test trials may potentiate the effect of later study trials by causing unavailable stimulus elements to become effective.

Assumption of two learning stages.—In especially systematic and thorough empirical comparisons of alternative assumptions about paired-associate memorizing by Atkinson & Crothers (7) and by Cotton et al (46), results have shown that learning in some situations produces at least three levels of performance. Data in the form of sequences of errors and correct responses are unable, however, to support decisions between models differing in the subtler details of the way in which the performance levels are achieved. Suppes, Groen & Schlag-Rey (215) found that a model postulating at least two stages of learning was also required to analyze the latency of response during paired-associate memorizing. And

Kintsch (120) showed that a two-stage analysis could be used to analyze some of the cases in which Rock's (193) replacement procedure fails to yield results expected from the all-or-none assumption.

Analyses involving two postulated stages of learning are especially useful if there are hypotheses and supporting empirical results that specify the nature of the stages. One class of analyses has considered stages of learning relating to responses. Crothers (48) analyzed learning of associations with compound responses; the stages of learning corresponded to associating the response components with the stimulus. Bower & Theios (31) analyzed learning after change of response for a paired associate, where the first stage was unlearning the first association, made incorrect by the response change. Millward (149) and Nahinsky (156) examined a learning system in which the subject learns not to give certain responses that are incorrect, thereby improving performance prior to learning the correct association. Wolford (240) has assumed that forward and backward associations are stored, each in an all-or-none fashion. This idea provides an explanation of recognition and recall of either stimuli or responses.

Another class of hypotheses has attributed multiple stages in learning to requirements for discrimination between items in retrieving stored associations. Restle (187) proposed a two-stage theory in which the first stage is storage of an engram or representation of an association and the second stage is the formation of a distinctive trace, based on discovery of a discriminative feature. Polson, Restle & Polson (175) used the theory for analyzing a situation where confusions occurred between specific pairs of items, so that errors in the second stage were identifiable as responses to items similar to the one tested. Analyses using the mixed pattern-components model (8) have converged to a similar interpretation. Friedman & Gelfand (81) analyzed performance on tests after study of stimuli having shared components and different responses. The main hypotheses involve retention of association between stimulus patterns and their responses. Friedman, Trabasso & Mosberg (83) concluded that the learning of an item has two stages, and in the intermediate state there is uncertainty in the retrieval process if there are other items with some of the item's stimulus components. The completion of learning involves storing a representation of the association in which the stimulus is an integrated pattern, thus discriminating it from other similar stimuli.

In comparing recognition and recall performance after study of paired associates, Estes & DaPolito (69) obtained results consistent with Kintsch & Morris' (125) idea that one stage of learning allows the subject to recognize the item, but that a second stage may be needed to permit retrieval. Results obtained by Humphreys & Greeno (107) led to a further hypothesis that the first stage is storage of a representation of the stimulus-response pair as a kind of Gestalt unit, and the item is made reliably retrievable in the second stage. While this idea is consistent with certain findings obtained when results of negative transfer experiments are analyzed (89, 90), there are also experimental results supporting the alternative view of Underwood & Schulz (232) that the two main stages of associative learning are response acquisition and acquisition of stimulus-response connections (see Underwood 231).

Analyses of discrimination.—Several analyses, some mentioned above, have recognized stimulus discrimination as an important factor in paired-associate memorizing. Some investigators have proposed specific hypotheses about the process of discrimination learning. Bower (30) analyzed learning of paired associates whose stimuli are ordered on some linear dimension such as size. He showed that serial position effects that are obtained can be explained by assuming that the effective stimulus is a quantity corresponding to the magnitude of the nominal stimulus relative to the adaptation level determined by the set of stimuli. Bower also considered implications for learning and transfer of the idea that ordered stimuli are represented in a cognitive structure that has the general properties of a linear ordering.

Analyses of paired-associate memorizing based on the concept of a discrimination net have been given by Feigenbaum (76) and by Hintzman (101). Hintzman's analysis was particularly thorough and systematic, and had the especially useful feature of exploring the limits of the concept of stimulus discrimination in analyzing variables that influence difficulty of paired-associate memorizing. For example, it appears that appropriate assumptions about stimulus discrimination are sufficient to explain a variety of effects involving the number of response alternatives, spacing between presentations of an item, and interactions between amount of practice and interference of both the proactive and retroactive varieties But effects of list length on learning difficulty and some salient facts about negative transfer appear to require explanation involving other kinds of mechanisms.

Stimulus-Response Conditioning

While studies of classical and instrumental conditioning have not enjoyed recently the popularity they had in years past, investigations of considerable significance and interest are still being conducted. Research based on quantitative models has been carried out on eyeblink conditioning, escape and avoidance conditioning, instrumental conditioning, and discriminative conditioning.

Eyeblink conditioning.—Eyeblink conditioning in rabbits was studied by Theios & Brelsford (220), who found that response sequences had the properties implied by a two-stage Markov model of the learning process. In eyeblink conditioning with human subjects, some results agree with the Markov model proposed by Bower & Theios (31), but Prokasy and his co-workers have found a slightly more complex form of the learning process—one in which performance remains at its initial level for some number of trials, then improves by an incremental process toward asymptotic performance. A model originally introduced by Norman (162) has parameters that give four measures of performance and learning: the initial level of performance, the number of trials before performance begins to change, the rate of change in performance once it starts changing, and the asymptotic level of performance. The model has been applied to analyze effects of the intensity of an unconditioned stimulus (180), as well as effects of interstimulus interval and the difference in performance between subjects classified as voluntary and involuntary responders (181). The model has also been useful in separating the learning and performance effects of certain drugs on conditioned responding in rats (234).

Shock-escape and avoidance conditioning.—The two-stage Markov model has also been used in analyzing shock-escape training (31, 216), and has provided an analysis of the effects of overtraining and successive reversals on reversal learning in a shock-escape T-maze (217). The process of conditioning an avoidance response also has been studied with the two-stage Markov model (31). Theios & Brelsford (221) found evidence that the first stage of learning is the conditioning of the instrumental response of running when frightened, while the second stage involves storing a permanent record of the association between the conditioned stimulus and the emotional arousal response. A series of experiments conducted by Brelsford (33) tested specific implications of Theios & Brelsford's interpretation of avoidance conditioning, and provides excellent examples of the way in which specific quantitative theorizing can generate innovative experimental procedures as a means of providing strong tests of hypotheses. Theios (219) has provided an informative review of analyses of aversive conditioning carried out in his laboratories.

Gibbon (85) has investigated free-operant avoidance in some detail, with particular emphasis on how asymptotic responding is sustained. In his view, asymptotic behavior can be treated as a combination of reconditioning and extinction, and he derives predictions of both interresponse and intershock intervals on the basis of a finite state (semi-Markov chain) interpretation of behavior at asymptote.

Instrumental conditioning.—Norman (163) analyzed instrumental conditioning as reflected by changes in interresponse time distributions as a function of schedules in which reinforcement probability is contingent on interresponse time. In his analysis of asymptotic interresponse time distributions under random ratio, variable interval, and DRL-like schedules, Norman assumed that nonreinforcement of a response increases the likelihood of long interresponse times, and that the reinforcement of a response characterized by a particular interresponse time t both increases the likelihood of short interresponse times and increases the likelihood of interresponse times about equal to t.

Discriminative conditioning.—Nearly all recent theoretical work on instrumental conditioning has been carried out with respect to situations in which discrimination plays a major role. Several approaches have been used in analyses of discrimination learning.

Lovejoy (140, 141) gave a thorough discussion of experimental findings concerned with attention in discrimination learning, and used these as a guide in developing assumptions of a theory. The theory, in which reinforcement influences the tendency to attend to the various stimulus dimensions and the strength of response associated with individual stimulus properties, was shown to be consistent with many facts in the complex and often puzzling literature of discriminative instrumental conditioning.

Other investigators have applied assumptions of discrete changes in state in analyzing discrimination learning. Clayton (41) analyzed the learning of a simple spatial discrimination in terms of the three-state version of all-or none learning stated by Greeno & Steiner (94). He examined in some detail the influence of several experimental variables, such as magnitude of reward and correction or noncorrection, on the transition probabilities in the model. In general, Clayton found support for the notion that such simple discriminations are learned in an all-or-none fashion, and he was able to demonstrate that events following a correct response are the primary determiners of the probability of transition to the learned state.

Another analysis using discrete changes in state was given by Lee (138), who assumed that each discrete stimulus used or each point on a stimulus continuum is in one of two states, conditioned to one of the response alternatives, and stimuli similar to one given on a trial have relatively high probability of changing state because of reinforcement given on the trial. Analyses have also been given in which the subject is assumed to go through discrete changes in state regarding the discrimination learning problem. Massaro (145) postulated a three-state system with the subject having either an appropriate or inappropriate strategy or being unconditioned. In a theory by Lynn (143), used to analyze discrimination by monkeys, it was assumed that the subject either has learned the correct response for a problem or is in the unlearned state, but throughout the problem there is a probability of not attending appropriately to the stimuli.

Analyses based on the idea that learning involves gradual quantitative change have been given by Holman (103, 105), who has been concerned with the relative effects of reward and nonreward on simple discriminative learning in rats. In his view, in order to isolate the functional form of the learning process one needs to untangle the effects of reward and nonreward, which are intertwined under normal circumstances. Holman has found (105) that when response probability is low events that increase response probability (reward) are more effective in changing behavior than are events that decrease response probability (nonreward), and when response probability is high the opposite is true. Holman (103) interpreted spontaneous alternation in the T-maze running of rats in terms of the differential effects of reward and nonreward, and he derived predictions of alternation behavior on the basis of that interpretation.

Durup (52–55) has also developed techniques for measuring changes of different kinds that occur during discrimination learning. The basis of the analysis is a model of choice behavior in which choice occurs in several stages, making the response a result of a random walk process (26, 63), and the model permits separation of learning to approach a positive alternative from learning to avoid a negative alternative. A particularly interesting development is Durup's interpretation of the stages of choice in the random walk model in relation to theoretical processes of attention, memory, and decision (56), thus connecting the random walk model with theoretical analyses like Lovejoy's (140, 141).

Classical continuity theory has been used by Spiker (207) as the basis for an extension of Hull-Spence discrimination learning theory. And Wolford & Bower

(241) have shown that some empirical results thought to oppose continuity theory are compatible with it.

A final example of theoretical work on discriminative conditioning is the work of Boneau & Cole (25), who used the ideas of statistical decision theory to analyze asymptotic performance in discriminative conditioning. They assumed that stimuli have variable effects, so that a given effective stimulus is associated with reinforcement probabilistically, and decisions whether to respond depend on the value of the reinforcement, the cost of responding, and the remembered frequency of reinforcement on past occasions of the effective stimulus.

Acquiring Information for Choice and Decision

The extensive theoretical work on sequential choice behavior during the last decade or so is a topic worthy in and of itself for an *Annual Review* chapter. The situations studied have involved discrete responses and response continuums, learnable and nonlearnable sequences, contingent and noncontingent outcome schedules, and variations in reward and reward schedules. Fortunately, there exist three excellent review articles that cover essentially all of the quantitative theorizing in this domain. In his article on probability learning in 1964, Estes (67) reviewed developments up to that point. Myers' systematic article in 1970 (155) gives an overview of theoretical work on sequential choice behavior and covers his own extensive work on the topic in some detail. Finally, Estes' recent theoretical review (68) of probability learning provides an integrative review of theoretical developments up to the same point in time as our review.

We will not review in sketchy form here the material reviewed so well in the three articles mentioned above. Rather, we refer the reader to those articles, and confine ourselves to a remark that the theoretical developments in the area of sequential choice behavior are an important component of the emerging psychology of learning and cognition that is pointed to in the introduction to this chapter. As Myers (155) emphasized, probability learning was interpreted almost exclusively in a conditioning framework until the middle 1960s. Since that time, an ever-increasing number of theories of sequential choice behavior emphasize cognitive processes such as encoding, chunking, hypothesis testing, and so forth. Thus, during the last decade, theorists have shifted their attention from the way in which response probabilities change with reward or nonreward to the way in which information processing mechanisms determine, on the basis of the past sequence of events, the current choice or series of choices.

Learning Classification Rules

Single dimensional rules.—In experimental tasks requiring subjects to classify stimuli into categories, learning is mainly the acquisition of a rule relating properties of stimuli to response. A substantial number of studies have been carried out in relation to Restile's (185) rigorous development of the idea that rule learning involves selection from a set of possibilities. Bower & Trabasso (32) derived many theorems about statistical properties of data and carried out a number of experiments giving startlingly good empirical support for the model, especially regarding the counterintuitive assumption that sampling from the hypothesis set occurs with replacement. Further work by Trabasso & Bower (228) provided thorough analysis of the implications of the idea of learning by selection for situations involving transfer of training, variation in cue salience and number of irrelevant cues, and overtraining. A clarification of a basic mathematical property of the model has been given by Fisher (79). Cotton (45) has given a substantive extension of the model dealing with dependence of response prior to solving the problem on trial-to-trial relationships between stimuli.

Falmagne (72) presented an analysis based on assuming that performance improves gradually over trials according to a linear operator, but at some random trial the solution is found and the correct response is known for all stimuli. In another analysis, Falmagne (74) assumed that sampling weights of hypotheses are altered by confirmations and disconfirmations that occur. Her analysis also provides a way of estimating the probability of resampling after errors and correct responses, and the frequency of a passive state in which subjects respond on the basis of no hypothesis about the classification rule. The statistical methods were used to test hypotheses about effects of the intertrial interval used in concept identification (75).

Several analyses have been given dealing with the role of STM in the process of learning a classification rule in a concept identification experiment. Trabasso & Bower (227) showed that results obtained when the relevant dimension shifts after every other error refute the assumption of sampling with no memory, and proposed an assumption that subjects remember stimulus characteristics on successive trials and then for three or four trials avoid sampling cues that are inconsistently paired with responses. Gregg & Simon (96) described four alternative theories, varying in depth of memory storage during concept identification; their estimates of the memory parameter indicated relatively weak use of memory by subjects.

Two theories have been proposed asuming that the role of memory in concept identification primarily involves retention of information about hypotheses that the subject is considering as possible solutions. Chumbley (40) analyzed solution of a four-dimension problem, assuming that the subject considers all four possible dimensions, eliminating hypotheses on the basis of stimulus information, but sometimes losing track after errors. The frequency of losing track of the hypothesis set was found to increase with shorter intertrial intervals and with a requirement of dealing with three concurrent problems rather than a single problem. Wickens & Millward (239) analyzed solution of a 12-dimension problem by practiced subjects. They concluded that subjects have STM capacities varying from one to four for hypotheses under active consideration, and when all hypotheses in STM have been eliminated those possiblities are kept out of the set available for sampling. On the other hand, there is a limitation on the number of eliminated hypotheses that the subject can remember, so that some eliminated hypotheses eventually return to the set of potential solutions.

Multidimensional rules.—Analyses have also been given for acquisition of conceptual rules more complex than those involving a single stimulus dimension. The simplest case involves learning a four-way classification consisting of a two-by-two table of binary dimensions. Trabasso & Bower (226) analyzed this case as independent selection of the two relevant stimulus attributes. Nahinsky (157) analyzed learning of conjunctive classification, assuming that acquisition occurs on the basis of sampling from the set of two-cue combinations, with focusing occurring on both positive and negative instances, and retention of the first stimulus-response pair shown in the experiment. Another analysis of concept learning based on the idea of hypothesis sampling was given by Joyner (118), who considered solution of a problem in which three players individually give numbers, with the goal of having the sum of their numbers equal a target number. In this case the elements of the hypothesis set reflect the subjects' knowledge of basic arithmetic operations.

As an alternative to the assumption of selection from a set of hypotheses, an assumption that positive and negative instances are stored and then scanned to find defining properties also has been used in analyses of concept identification. Hunt & Hovland (111) gave an analysis using this idea, including the assumption that a limited set of dimensions is used in the scanning process at any one time. Johnson (116) presented a more elaborate system, with a mechanism for selecting among subclasses of stimulus attributes for consideration, and special subroutines for attempting to find conjunctive and disjunctive solution concepts. In a somewhat more formal analysis, Mott & Ross (152) described an algorithm for generating a list of possible concepts from a set of positive and negative instances based on the logical theory of implicants of an incomplete truth function.

Hunt and his co-workers have given analyses of the process of learning rules for classification, using the powerful conceptual method of representing the acquired concept as a decision tree (108). Hunt, Marin & Stone (112) tested systems varying in their sophistication in choosing stimulus items for test as well as in the amount of memory that was given to the system. Studies of the efficiency of the systems were carried out, as well as comparisons of simulated concept learning with performance by human subjects. Hunt (109) examined further variations regarding the method of selecting attributes for test and for assigning a label when an instance is encountered that is not specified on the basis of earlier learning.

Classification in associative learning.—Some attention has been given to the process of acquiring a classification rule during the course of paired-associate memorizing, when relationships among items can be used to simplify the task by grouping related items that have the same response. The situations studied have used the conditions of short lists and few response alternatives so the process of associative learning is approximately all-or-none. Then the transfer of association to new instances appears also to be an all-or-none process (92). The process of acquiring a category rule has been analyzed by Batchelder (11) as an all-or-none process, and data appear to favor that idea over the mixed pattern-components model of stimulus sampling theory (12).

FORMAL ASPECTS OF MATHEMATICAL MODELS

In this final section we will first present a brief classification of types of mathematical learning theory, based on their mathematical properties rather than their substantive assumptions. Then we will review contributions to theory of a primarily methodological nature, concerned with formal properties of models, including issues involved in evaluation.

Types of Theory

A major dimension of variation among mathematical theories of learning involves the degree to which specific mental processes are postulated. At one end of this continuum are quantitative descriptions of performance, where the only theoretical entity is response probability. Experimental events (trials) have effects on response probability represented by operators, and effects of experimental variables are represented by changes in the values of parameters of the transformational operators. Theories that specify characteristics of psychological process and structure in more detail provide fuller explanation of performance, but more importantly give more exact understanding of the system studied. We differ from some commentators (96) who seem to see discrete categories of theory, such as "stochastic models," that are different in kind from "process models" or other similar dichotomies. Our view is that various theories about any process will vary both in their degrees of specificity and in the kinds of specific process assumed. Furthermore, the amount of specificity built into a theory at any stage of knowledge represents an important dimension of research strategy. And since research strategies apparently pay off more or less randomly, it is almost surely to the benefit of science that different investigators tend to choose different strategies, so the scientific community has what amounts to a portfolio of strategies in order to benefit from whatever advantages the various approaches may have for different problems at various stages of progress.

While theories differ in specificity in a more or less continuous fashion, there are some benchmarks. One step toward specificity is an assumption of a state space, with different levels of performance associated with the various states, and learning represented as transition among the states. More specificity is achieved in a stochastic model in which the states are interpreted as conditions of cognitive status, such as those given in Figure 1 of this article, or as stages of learning or problem solving involving specific cognitive achievement such as storage of representations. Even more specific assumptions are made in some theories, often in connection with the need to specify processes sufficiently so that an operating computer program can be written in an available programming language. It is reasonable to suppose that some of the processes described summarily in an information-processing theory, such as scanning a list and comparing with a test item, might be described in some detail in a theory dealing with processes in a more molecular way.

In our view, there is no argument against the goal of describing cognitive processes in more specific ways, as well as having descriptions that apply to a wide variety of situations in which learning and other cognitive processes occur. At the same time, significant questions remain as to optimal strategy to be used at any given stage of theoretical progress and empirical knowledge about the processes that occur in accomplishing various tasks. Given the empirical knowledge that is available regarding a process, some possible statements of hypotheses are so general that their likelihood of falsification in data is negligible, and they probably have poor prospects for providing a basis for significant new knowledge and understanding. At the same time, other possible hypotheses involve so many detailed assumptions that go beyond available empirical evidence that they are sure to be refuted in their first confrontation with data. While it can be instructive to see a combination of assumptions that are mutually consistent and capable of giving successful performance on a complex task, it remains the task of science to develop testable hypotheses that can be challenged meaningfully in experiments, and an overabundance of detail in a theory can often preclude the use of experimental findings in guiding modifications of theory toward more accurate description. These cautions notwithstanding, we suspect that the current state of knowledge about learning is such that quite a variety of theoretical approaches are likely to lead to meaningful advances in knowledge and understanding, and we are encouraged by the relatively broad band of degrees of specificity in current theoretical work.

Mathematical and Statistical Studies

A considerable number of investigations have occurred where mathematical models are studied from the point of view of their general properties or statistical considerations involved in their application. Norman & Yellott (170) investigated properties of models that lead to probability matching, and convergence and limit properties that follow from several general assumptions about learning operators have been studied by Norman and others (104, 147, 164–169, 242). There have also been analyses of families of learning operators characterized by properties such as commutativity (142, 144), and generalizations of discrete models, such as those deriving from stimulus sampling theory, as continuous-time processes (4, 51, 213).

Suppes (210) studied asymptotic properties of learning systems formulated on the basis of stimulus-response principles, and showed that they can have limiting properties that make them isomorphic to finite automata with many of the characteristics needed to describe language and other complex behavior. Discussion of the status of this result by Arbib (3) and Suppes (211) has emphasized the importance of determining the number of states needed to produce performance that people show when they use language.

Both the linear model and the all-or-none model have been the topic of mathematical studies. Some analyses have generalized the linear model (21, 113, 165), and several mathematical studies of the all-or-none model have been carried out. Two aspects of the all-or-none assumption have been distinguished, with several investigators noting that stationarity of performance before learning and constancy of the probability of learning involve separate assumptions (73, 177, 188, 196). Rouanet et al (196) have given careful discussion of the way in which individual differences can influence tests of the all-or-none model. An especially promising development has been the use of the all-or-none model (119, 205) and other more general models (49) in the analysis of item-presentation sequences in order to develop presentation strategies that optimize learning efficiency.

A considerable literature is developing dealing with technical matters such as parameter estimation and evaluation of learning models. Bernbach (15) and Millward (150) have given general methods of deriving empirical predictions from Markov models. General discussions of estimation have been given by Holland and Kraemer (102, 126–128). A thorough discussion of all-or-none parameters by Polson (176) deals with estimation as well as tests of hypotheses. Theios (218) has given a method of estimating performance parameters and trials of transition appropriate for experiments having many responses in each subject's sequence of performance. Regarding evaluation of models, Rouanet (194) has given an analysis based on relationships between models that deal with varying degrees of fine-grainedness in the partition of experimental events, and Regnier & Rouanet (182) have shown that different interpretations of all-or-none learning can be distinguished empirically by moving to a finer partition of events in which a subject's complete performance is considered. Questions involving statistical tests of goodness-of-fit continue to be raised (96), and Kraemer (129) and Hanna (98) have proposed alternatives to the standard and still much-used indices such as goodness-of-fit chi square.

Investigations have been made of the kinds of inference that can be justified on the basis of experimental data. Richard (192) and Jonkheere (117) have considered examples in which different psychological processes can lead to identical empirical predictions and thus not to distinguishable in data. A specific form of of this problem arises when a mathematical model has more parameters than can be identified on the basis of data from specific kinds of experiments, and several models have been analyzed regarding the identifiability of parameters (7, 86, 88, 91, 93, 94, 208).

TEXTBOOKS

Texts by Galanter (84) and Greeno (87) have been written for beginning undergraduates with emphasis on systematic theory and inclusion of mathematical models of learning. For more advanced undergraduates or beginning graduate students, Atkinson, Bower & Crothers' (5) text presents the basic concepts of stimulus sampling theory, includes application to a number of problems, and gives considerable attention to techniques of use. A book of problems by Batchelder, Bjork & Yellott (13) is available. Rouanet's (195) text also introduces basic concepts of stimulus sampling theory, and provides an excellent introduction to the theory of conditioning on a continuum of responses. Restle & Greeno's

(191) text gives three chapters to mathematical learning theory, as well as four chapters dealing with mathematical and statistical issues that are especially relevant to learning models. Coombs, Dawes & Tversky (44) have a chapter on learning, and they present general discussion relating to the use of mathematical theory in psychology. Reitman's (184) text presents an introduction to the use of list processing notation and computer simulation in psychological theory, including some problems in learning. Other books of particular interest are Feigenbaum & Feldman's (77) collection of papers in computer simulation and artificial intelligence, Neimark & Estes' (158) collection of papers on stimulus sampling theory, and Norman's (160) collection of theoretical papers on human memory. A newly available book by Levine & Burke (139) on model techniques for learning theories presents basic mathematical material in probability theory, matrix algebra, and Markov chains, along with techniques involved in geometric series, difference equations, and identifiability of parameters, all of which give a very helpful source of mathematical background useful in learning theory. A most pleasing development in the field of text materials is Kintsch's (123) general text in learning and memory, in which results based on stochastic models and simulation are considered in detail as part of the general literature on learning, rather than as a segregated subject.

LITERATURE CITED

1. Albert, D. 1968. Freies Reproduzieren von Wortreihen als stochastische Entleerung eines Speichers. *Z. Exp. Angew. Psychol.* 15:564–81
2. Alpern, M. 1953. Metacontrast. *J. Opt. Soc. Am.* 43:648–57
3. Arbib, M. A. 1969. Memory limitations of stimulus-response models. *Psychol. Rev.* 76:507–14
4. Arnold, B. C. 1967. Response distribution for the continuous-time N-element pattern model. *J. Math. Psychol.* 4:489–500
5. Atkinson, R. C., Bower, G. H., Crothers, E. J. 1965. *Introduction to Mathematical Learning Theory.* New York: Wiley. 429 pp.
6. Atkinson, R. C., Brelsford, J. W., Shiffrin, R. M. 1967. Multiprocess models for memory with applications to a continuous presentation task. *J. Math. Psychol.* 4:277–300
7. Atkinson, R. C., Crothers, E. J. 1964. A comparison of paired-associate learning models having different acquisition and retention axioms. *J. Math. Psychol.* 1:285–315
8. Atkinson, R. C., Estes, W. K. 1963. *Handbook of Mathematical Psychology*, ed. R. D. Luce, R. R. Bush, E. Galanter, 121–268. New York: Wiley. 606 pp.
9. Atkinson, R. C., Shiffrin, R. M. 1968. *The Psychology of Learning and Motivation*, ed. K. W. Spence, J. T. Spence, 2:90–197. New York: Academic. 255 pp.
10. Atkinson, R. C., Wickens, T. D. 1971. *The Nature of Reinforcement*, ed. R. Glaser, 66–120. New York: Academic. 379 pp.
11. Batchelder, W. H. 1970. An all-or-none theory for learning on both the paired-associate and concept levels. *J. Math. Psychol.* 7:97–117
12. Ibid 1971. A theoretical and empirical comparison of the all-or-none multilevel theory and the mixed model. 8:82–108
13. Batchelder, W. H., Bjork, R. A., Yellott, J. I. Jr. 1966. *Problems in Mathematical Learning Theory.* New York: Wiley. 184 pp.
14. Bernbach, H. A. 1965. A forgetting model for paired-associate learning. *J. Math. Psychol.* 2:128–44
15. Bernbach, H. A. 1966. Derivation of learning process statistics for a general Markov model. *Psychometrika* 31:225–34
16. Bernbach, H. A. 1967. Decision processes in memory. *Psychol. Rev.* 74:462–80
17. Bernbach, H. A. 1969. *The Psychology of Learning and Motivation*, ed. G. H. Bower, J. T. Spence, 3:201–40. New York: Academic. 398 pp.
18. Bernbach, H. A. 1970. See Ref. 160, 103–16
19. Bernbach, H. A. 1971. Strength theory and confidence ratings in recall. *Psychol. Rev.* 78:338–40
20. Binder, A., Estes, W. K. 1966. Transfer of response in visual recognition situations as a function of frequency variables. *Psychol. Monogr.* 80:Whole no. 631
21. Bishir, J., Drewes, D. W. 1969. A generalized linear learning model. *J. Math. Psychol.* 6:121–38
22. Bjork, R. A. 1968. All-or-none subprocesses in the learning of complex sequences. *J. Math. Psychol.* 5:182–95
23. Bjork, R. A. 1970. See Ref. 160, 307–32
24. Blough, D. S., Millward, R. B. 1965. Learning: Operant conditioning and verbal learning. *Ann. Rev. Psychol.* 16:63–94
25. Boneau, C. A., Cole, J. L. 1967. Decision theory, the pigeon, and the psychophysical function. *Psychol. Rev.* 74:123–35
26. Bower, G. H. 1959. *Studies in Mathematical Learning Theory*, ed. R. R. Bush, W. K. Estes, 109–24. Stanford Univ. Press. 432 pp.
27. Bower, G. H. 1961. Application of a model to paired-associate learning. *Psychometrika* 27:255–80
28. Bower, G. H. 1967. *Proceedings of the Second Conference on Learning, Remembering, and Forgetting*, ed. D. P. Kimble, 112–85. New York: NY Acad. Sci.
29. Bower, G. H. 1967. *The Psychology of Learning and Motivation*, ed. K. W. Spence, J. T. Spence, 1:230–327. New York: Academic. 381 pp.
30. Bower, G. H. 1971. *Adaptation Level Theory*, ed. M. H. Appley, 175–201. New York: Academic. 346 pp.
31. Bower, G. H., Theios, J. 1964. *Studies in Mathematical Psychol-*

ogy, ed. R. C. Atkinson, 1–31. Stanford Univ. Press. 414 pp.
32. Bower, G. H., Trabasso, T. R. 1964. See Ref. 31, 32–94
33. Brelsford, J. W. Jr. 1967. Experimental manipulation of state occupancy in a Markov model for avoidance conditioning. *J. Math. Psychol.* 4:21–47
34. Broadbent, D. E. 1958. *Perception and Communication*. London: Pergammon. 338 pp.
35. Brown, J. 1958. Some tests of the decay theory of immediate memory. *Quart. J. Exp. Psychol.* 10:12–21
36. Buchwald, A. M. 1969. Effects of "right" and "wrong" on subsequent behavior: A new interpretation. *Psychol. Rev.* 76:132–43
37. Bush, R. R., Mosteller, F. 1951. A model for stimulus generalization and discrimination. *Psychol. Rev.* 58:413–23
38. Bush, R. R., Mosteller, F. 1955. *Stochastic Models for Learning*. New York: Wiley. 365 pp.
39. Calfee, R. C., Atkinson, R. C. 1965. Paired-associate models and the effects of list length. *J. Math. Psychol.* 2:254–65
40. Chumbley, J. 1969. Hypothesis memory in concept learning. *J. Math. Psychol.* 6:528–40
41. Clayton, K. N. 1969. *Reinforcement and Behavior*, ed. J. T. Tapp, 96–119. New York: Academic. 429 pp.
42. Collins, A. M., Quillian, M. R. 1969. Retrieval time from semantic memory. *J. Verb. Learn. Verb. Behav.* 8: 240–47
43. Conrad, R. 1964. Acoustic confusion in immediate memory. *Brit. J. Psychol.* 55:75–84
44. Coombs, C. H., Dawes, R. M., Tversky, A. 1970. *Mathematical Psychology: An Elementary Introduction*. Englewood Cliffs, N. J.: Prentice-Hall. 419 pp.
45. Cotton, J. W. 1971. A sequence-specific concept identification model: Infra-structure for the Bower and Trabasso theory. *J. Math. Psychol.* 8:333–69
46. Cotton, J. W., Sloss, J. M., Taplin, J. R., Nielson, A. E. 1968. A comparison of a Hullian paired-associate learning model with other models appropriate to an R^NT^5 design with 1:1 pairing of stimuli and responses. *J. Math. Psychol.* 5:442–62
47. Cowan, T. M. 1966. A Markov model for order of emission in free recall. *J. Math. Psychol.* 2:470–83
48. Crothers, E. J. 1964. See Ref. 31, 95–115
49. Crothers, E. J. 1965. Learning model solution to a problem in constrained optimization. *J. Math. Psychol.* 2:19–25
50. Donaldson, W., Glathe, H. 1970. Signal-detection analysis of recall and recognition memory. *Can. J. Psychol.* 24:42–56
51. Donio, J. 1969. Stimulus sampling learning models described as continuous-time stochastic processes. *J. Math. Psychol.* 6:240–57
52. Durup, H. 1964. Un modèle stochastique du comportement conduisant à la réponse, chez un animal en situation d'apprentissage discriminatif. I. Apprentissage sans correction, première approximation. *Cah. Psychol.* 7:21–29
53. Ibid. II. Apprentissage sans correction, seconde approximation,57–59
54. Ibid 1965. III. Apprentissage avec correction. 8:89–99
55. Ibid 1966. IV. Rôle du modèle comme hypothèse nulle par rapport à des résultats expérimentaux. 9:65–73
56. Ibid. Les modèles à cheminement aléatoire et l'analyse du comportement de choix discriminatif, 135–42
57. Erickson, J. R. 1968. Hypothesis sampling in concept identification. *J. Exp. Psychol.* 76:12–18
58. Ernst, G. W., Newell, A. 1969. *GPS: A Case Study in Generality and Problem Solving*. New York: Academic
59. Estes, W. K. 1950. Toward a statistical theory of learning. *Psychol. Rev.* 57:94–107
60. Ibid 1955. Statistical theory of distributional phenomena in learning. 62:369–77
61. Ibid. Statistical theory of spontaneous recovery and regression, 145–54
62. Estes, W. K. 1959. *Psychology: A Study of a Science. General Systematic Formulations, Learning, and Special Processes*, ed. S. Koch, 2:380–491. New York: McGraw-Hill. 706 pp.
63. Estes, W. K. 1960. *Mathematical Methods in the Social Sciences*, ed.

K. J. Arrow, S. Karlin, P. Suppes, 265–76. Stanford Univ. Press
64. Estes, W. K. 1961. New developments in statistical behavior theory: Differential tests of axioms for associative learning. *Psychometrika* 26:73–84
65. Estes, W. K. 1962. Learning theory. *Ann. Rev. Psychol.* 13:107–44
66. Estes, W. K. 1964. All-or-none processes in learning and retention. *Am. Psychol.* 19:16–25
67. Estes, W. K. 1964. *Categories of Human Learning*, ed. A. W. Melton, 89–128. New York: Academic. 356 pp.
68. Estes, W. K. 1972. Research and theory on the learning of probabilities. *J. Am. Statist. Assoc.* 67:81–102
69. Estes, W. K., DaPolito, F. 1967. Independent variation of information storage and retrieval processes in paired-associate learning. *J. Exp. Psychol.* 75:18–26
70. Estes, W. K., Suppes, P. 1959. See Ref. 26, 137–79
71. Estes, W. K., Taylor, H. A. 1964. A detection method and probabilistic models for assessing information processing from brief visual displays. *Proc. Nat. Acad. Sci.* 52:446–54
72. Falmagne, R. 1965. Un modèle mixte pour la situation d'identification de concepts. *Bull. Etudes Rech. Psychol.* 14:219–41
73. Falmagne, R. 1969. Le caractère tout-ou-rien ou progressif de l'apprentissage de concepts: Discussion théorique et examen de résultats expérimentaux. *Année Psychol.* 69:143–60
74. Falmagne, R. 1970. Construction of a hypothesis model for concept identification. *J. Math. Psychol.* 7:60–96
75. Ibid 1971. Concept identification: Effect of the intertrial interval on the efficacy and consistency of information processing. 8:240–47
76. Feigenbaum, E. A. 1963. See Ref. 77, 297–309
77. Feigenbaum, E. A., Feldman, J., Eds. 1963. *Computers and Thought*. New York: McGraw-Hill. 535 pp.
78. Feigenbaum, E. A., Simon, H. A. 1962. A theory of the serial position effect. *Brit. J. Psychol.* 53:307–20
79. Fisher, M. A. 1971. A note on the generalization of some results in hypothesis sampling theory. *J. Math. Psychol.* 8:235–39
80. Freund, R. D., Loftus, G. R., Atkinson, R. C. 1969. Applications of multiprocess models for memory to continuous recognition tasks. *J. Math. Psychol.* 6:576–94
81. Friedman, M. P., Gelfand, H. 1964. Transfer effects in discrimination learning. *J. Math. Psychol.* 1:204–14
82. Friedman, M. P., Rollins, H., Padilla, G. 1968. The role of cue validity in stimulus compounding. *J. Math. Psychol.* 5:300–10
83. Friedman, M. P., Trabasso, T., Mosberg, L. 1967. Tests of a mixed model for paired-associates learning with overlapping stimuli. *J. Math. Psychol.* 4:316–34
84. Galanter, E. 1966. *Textbook of Elementary Psychology*. San Francisco: Holden-Day. 419 pp.
85. Gibbon, J. 1971. Scalar timing and semi-Markov chains in free-operant avoidance. *J. Math. Psychol.* 8:109–38
86. Greeno, J. G. 1967. Paired-associate learning with short-term retention: Mathematical analysis and data regarding identification of parameters. *J. Math. Psychol.* 4:430–72
87. Greeno, J. G. 1968. *Elementary Theoretical Psychology*. Reading, Mass.: Addison-Wesley, 324 pp.
88. Greeno, J. G. 1968. Identifiability and statistical properties of two-stage learning with no successes in the initial stage. *Psychometrika* 33:173–216
89. Greeno, J. G. 1970. See Ref. 160, 261–84
90. Greeno, J. G., James, C. T., DaPolito, F. J. 1971. Cognitive interpretation of negative transfer and forgetting of paired associates. *J. Verb. Learn. Verb. Behav.* 10:331–45
91. Greeno, J. G., Millward, R. B., Merryman, C. T. 1971. Matrix analysis of identifiability of some finite Markov chains. *Psychometrika* 36:389–408
92. Greeno, J. G., Scandura, J. M. 1966. All-or-none transfer based on verbally mediated concepts. *J. Math. Psychol.* 2:388–411
93. Greeno, J. G., Steiner, T. E. 1964. Markovian processes with identifiable states: General considerations

and application to all-or-none learning. *Psychometrika* 29:309–33
94. Ibid 1968. Comments on "Markovian processes with identifiable states: General considerations and application to all-or-none learning." 33:169–72
95. Gregg, L. W. 1967. *Concepts and the Structure of Memory*, ed. B. Kleinmuntz, 107–42. New York: Wiley. 286 pp.
96. Gregg, L. W., Simon, H. A. 1967. Process models and stochastic theories of simple concept formation. *J. Math. Psychol.* 4:246–76
97. Gyr, J. W., Brown, J. S., Cafagna, A. C. 1967. Quasiformal models of inductive behavior and their relation to Piaget's theory of cognitive stages. *Psychol. Rev.* 74:272–89
98. Hanna, J. F. 1969. Some information measures for testing stochastic models. *J. Math. Psychol.* 6:294–311
99. Haygood, R. C., Bourne, L. E. Jr. 1965. Attribute- and rule-learning aspects of conceptual behavior. *Psychol. Rev.* 72:175–95
100. Hinrichs, J. V. 1970. A two-process memory-strength theory for judgment of recency. *Psychol. Rev.* 77: 223–33
101. Hintzman, D. L. 1968. Explorations with a discrimination net model for paired-associate learning. *J. Math. Psychol.* 5:123–62
102. Holland, P. W. 1968. A remark on Bower's estimate for the one-element Markov learning model. *Psychometrika* 33:217–24
103. Holman, E. W. 1966. Tests for spontaneous alternation. *Psychol. Rev.* 73:427–36
104. Holman, E. W. 1969. Asymptotic properties of monotonic learning models. *J. Math. Psychol.* 6:456–69
105. Ibid. The relative effects of reward and nonreward in simple learning, 497–513
106. Hull, C. L. 1952. *A Behavior System.* New Haven: Yale Univ. Press
107. Humphreys, M., Greeno, J. G. 1970. Interpretation of the two-stage analysis of paired-associate memorizing. *J. Math. Psychol.* 7:275–92
108. Hunt, E. B. 1962. *Concept Learning: An Information Processing Problem.* New York: Wiley. 286 pp.
109. Hunt, E. B. 1967. See Ref. 95, 77–106
110. Hunt, E. 1968. Computer simulation: Artificial intelligence studies and their relevance to psychology. *Ann. Rev. Psychol.* 19:135–68
111. Hunt, E. B., Hovland, C. I. 1963. See Ref. 77, 310–28
112. Hunt, E. B., Marin, J., Stone, P. I. 1966. *Experiments in Induction.* New York: Academic. 247 pp.
113. Iosifescu, M., Theodorescu, R. 1965. On Bush-Mosteller stochastic models for learning. *J. Math. Psychol.* 2:196–203
114. Izawa, C. 1971. The test trial potentiating model. *J. Math. Psychol.* 8:200–24
115. James, W. 1890. *The Principles of Psychology*. New York: Dover
116. Johnson, E. S. 1964. An information-processing model of one kind of problem solving. *Psychol. Monogr.* 78:Whole no. 581
117. Jonkheere, A. R. 1967. *Les Modelès et la Formalisation du Comportement*, 405–11. Paris: Centre Nat. Rech. Sci. 424 pp.
118. Joyner, R. C. 1970. Computer simulation of individual concept learning in the three-person common target game. *J. Math. Psychol.* 7:478–514
119. Karush, W., Dear, R. E. 1966. Optimal stimulus presentation strategy for a stimulus sampling model of learning. *J. Math. Psychol.* 3:19–47
120. Kintsch, W. 1963. All-or-none learning and the role of repetition in paired-associate learning. *Science* 140:310–12
121. Kintsch, W. 1966. Recognition learning as a function of the length of the retention interval and changes in the retention interval. *J. Math. Psychol.* 2:412–33
122. Kintsch, W. 1967. Memory and decision aspects of recognition learning. *Psychol. Rev.* 74:496–504
123. Kintsch, W. 1970. *Learning, Memory, and Conceptual Processes.* New York: Wiley. 498 pp.
124. Kintsch, W. 1970. See Ref. 160, 333–74
125. Kintsch, W., Morris, C. J. 1965. Application of a Markov model to free recall and recognition. *J. Exp. Psychol.* 69:200–6
126. Kraemer, H. C. 1963. Comments on parameter estimation for Restle's theory of cue learning. *Psychometrika* 28:185–88
127. Ibid 1964. Estimation in the three-state Markov learning model. 29: 131–41

128. Kraemer, H. C. 1964. Point estimation in learning models. *J. Math. Psychol.* 1:28–53
129. Kraemer, H. C. 1965. The average error of a learning model, estimation and use in testing the fit of models. *Psychometrika* 30:343–52
130. Kuhn, T. S. 1962. *The Structure of Scientific Revolutions.* Univ. Chicago Press. 172 pp.
131. Kuno, U. 1965. A model for serial verbal learning. *Psychometrika* 30: 323–41
132. Landauer, T. K. 1969. Reinforcement as consolidation. *Psychol. Rev.* 76:82–96
133. Landauer, T. K., Freedman, J. L. 1968. Information retrieval from long-term memory: Category size and recognition time. *J. Verb. Learn. Verb. Behav.* 7:291–95
134. Lander, H. J. 1969. Beiträge zur Psychologie des Gedächtnisses VI; Über Strukturbildungsprozesse beim menschlichen Gedächtnis. II. Die Veränderung der intraserialen Assoziationen im Vergessensprozess. *Z. Psychol.* 176:1–21
135. Ibid. III. Ein stochastisches Operatormodell zur Beschreibung serialer Lernprozesse, 157–81
136. Lashley, K. S. 1951. *Cerebral Mechanisms in Behavior*, ed. L. A. Jeffress, 112–36. New York: Wiley
137. Laughery, K. R. 1969. See Ref. 17, 135–200
138. Lee, W. 1966. Conditioning parameter model for reinforcement generalization in probabilistic discrimination learning. *J. Math. Psychol.* 3:184–96
139. Levine, G., Burke, C. J. 1972. *Mathematical Model Techniques for Learning Theories.* New York: Academic. 288 pp.
140. Lovejoy, E. 1966. Analysis of the overlearning reversal effect. *Psychol. Rev.* 73:87–103
141. Lovejoy, E. 1968. *Attention in Discrimination Learning, a Point of View and a Theory.* San Francisco: Holden-Day. 135 pp.
142. Luce, R. D. 1964. See Ref. 31, 380–98
143. Lynn, E. 1969. A stimulus sampling model for simultaneous discrimination learning. *J. Math. Psychol.* 6:277–85
144. Marley, A. A. J. 1967. Abstract one-parameter families of commutative learning operators. *J. Math. Psychol.* 3:414–29
145. Massaro, D. W. 1969. A three state Markov model for discrimination learning. *J. Math. Psychol.* 6:62–80
146. Massaro, D. W. 1970. Perceptual processes and forgetting in memory tasks. *Psychol. Rev.* 77:557–67
147. McGregor, J. R., Zidek, J. V. 1965. A sequence of limiting distributions of response probabilities. *Psychometrika* 30:491–97
148. Meyer, D. E. 1970. On the representation and retrieval of stored semantic information. *Cogn. Psychol.* 1:242–99
149. Millward, R. B. 1964. An all-or-none model for noncorrection routines with elimination of incorrect responses. *J. Math. Psychol.* 1:392–404
150. Millward, R. B. 1969. Derivations of learning statistics from absorbing Markov chains. *Psychometrika* 34: 215–32
151. Minsky, M., Papert, S. 1969. *Perceptrons.* Cambridge, Mass.: MIT. 258 pp.
152. Mott, T. H. Jr., Ross, B. M. 1968. An application of truth functional logic to concept-attainment experiments. *J. Math. Psychol.* 5:470–95
153. Murdock, B. B. Jr. 1967. Theoretical note: A fixed-point model for short-term memory. *J. Math. Psychol.* 4:501–6
154. Murdock, B. B. Jr. 1970. See Ref. 160, 285–304
155. Myers, J. L. 1970. *The Psychology of Learning and Motivation, Advances in Research and Theory*, ed. G. H. Bower, 4:109–91. New York: Academic. 339 pp.
156. Nahinsky, I. D. 1967. Statistics and moments-parameter estimates for a duoprocess paired-associate learning model. *J. Math. Psychol.* 4:140–50
157. Nahinsky, I. D. 1970. A hypothesis sampling model for conjunctive concept identification. *J. Math. Psychol.* 7:293–316
158. Neimark, E. D., Estes, W. K. 1967. *Stimulus Sampling Theory.* San Francisco: Holden-Day. 681 pp.
159. Nilsson, N. J. 1965. *Learning Machines.* New York: McGraw-Hill
160. Norman, D. A. 1970. *Models of Human Memory.* New York: Academic. 537 pp.
161. Norman, D. A., Rumelhart, D. E. 1970. See Ref. 160, 21–66

162. Norman, M. F. 1964. See Ref. 31, 173–87
163. Norman, M. F. 1966. An approach to free-responding on schedules that prescribe reinforcement probability as a function of interresponse time. *J. Math. Psychol.* 3:235–68
164. Norman, M. F. 1966. Limiting distributions for some random walks arising in learning models. *Ann. Math. Statist.* 37:393–405
165. Norman, M. F. 1968. On the linear model with two absorbing barriers. *J. Math. Psychol.* 5:225–41
166. Ibid. Some convergence theorems for stochastic learning models with distance diminishing operators, 61–101
167. Ibid 1970. Limit theorems for additive learning models. 7:1–11
168. Ibid 1971. Slow learning with small drift in two-absorbing-barrier models. 8:1–21
169. Norman, M. F., Graham, N. V. 1968. A central limit theorem for families of stochastic processes indexed by a small average step size parameter, and some applications to learning models. *Psychometrika* 33:441–50
170. Norman, M. F., Yellott, J. I. Jr. 1966. Probability matching. *Psychometrika* 31:43–60
171. Olson, G. M. 1968. *Stanford Univ. Inst. Math. Stud. Soc. Sci. Tech. Rep. 131*
172. Parks, T. E. 1966. Signal-detectability theory of recognition-memory performance. *Psychol. Rev.* 73:44–58
173. Peterson, L. R., Hillner, K., Saltzman, D. 1962. Time between pairings and short-term retention. *J. Exp. Psychol.* 64:550–51
174. Peterson, L. R., Peterson, M. J. 1959. Short-term retention of individual verbal items. *J. Exp. Psychol.* 58:193–98
175. Polson, M. C., Restle, F., Polson, P. G. 1965. Association and discrimination in paired-associates learning. *J. Exp. Psychol.* 69:47–55
176. Polson, P. G. 1970. Statistical methods for a general theory of all-or-none learning. *Psychometrika* 35:51–72
177. Polson, P. G., Greeno, J. G. 1969. Nonstationary performance before all-or-none learning. *Psychol. Rev.* 76:227–31
178. Popper, K. R. 1959. *Logic of Scientific Discovery*. New York: Basic Books. 480 pp.
179. Posner, M. I., Mitchell, R. F. 1967. Chronometric analysis of classification. *Psychol. Rev.* 74:392–409
180. Prokasy, W. F., Harsanyi, M. A. 1968. Two-phase model for human classical conditioning. *J. Exp. Psychol.* 78:359–68
181. Prokasy, W. F., Kumpfer, K. 1969. One- and two-operator versions of a two-phase model applied to the performances of Vs and Cs in human eyelid conditioning. *J. Exp. Psychol.* 80:231–36
182. Regnier, S., Rouanet, H. 1968. *Calcul et Formalisation dans les Sciences de l'Homme*, 201–24. Paris: Centre Nat. Rech. Sci.
183. Reitman, J. S. 1970. See. Ref. 160, 117–48
184. Reitman, W. R. 1965. *Cognition and Thought*. New York: Wiley. 312 pp.
185. Restle, F. 1962. The selection of strategies in cue learning. *Psychol. Rev.* 69:329–43
186. Ibid 1964. A cognitive interpretation of intensity effects in stimulus generalization. 71:514–16b
187. Restle, F. 1964. See Ref. 31, 116–72
188. Restle, F. 1965. The significance of all-or-none learning. *Psychol. Bull.* 62:313–24
189. Restle, F. 1970. Theory of serial pattern learning: Structural trees. *Psychol. Rev.* 77:481–95
190. Restle, F., Brown, E. 1970. See Ref. 155, 249–332
191. Restle, F., Greeno, J. G. 1970. *Introduction to Mathematical Psychology*. Reading, Mass.: Addison-Wesley. 371 pp.
192. Richard, J. F. 1967. See Ref. 117, 135–49
193. Rock, I. 1957. The role of repetition in associative learning. *Am. J. Psychol.* 70:186–93
194. Rouanet, H. 1967. See Ref. 117, 113–28
195. Rouanet, H. 1967. *Les modèles stochastiques d'apprentissage*. Paris: Gauthier-Villars. 260 pp.
196. Rouanet, H., Ackerman, W., Iff, M., Bartz, A., Dupas, J. 1967. Épreuve de l'hypothèse d'acquisition par tout-ou-rien dans une expérience d'apprentissage de couples. *Année Psychol.* 67:409–41
197. Rumelhart, D. E. 1970. A multicomponent theory of the perception of briefly exposed visual dis-

plays. *J. Math. Psychol.* 7:191–218
198. Samuel, A. L. 1963. See Ref. 77, 71-108
199. Selfridge, O. G. 1959. *The Mechanization of Thought Processes*, ed. D. V. Blake, A. M. Uttley, 511–29. London: Her Majesty's Stationary Office
200. Selfridge, O. G., Neisser, U. 1963. See Ref. 77, 237–50
201. Shepard, R. N., Teghtsoonian, M. 1961. Retention of information under conditions approaching a steady state. *J. Exp. Psychol.* 62:302–9
202. Shiffrin, R. M. 1970. See Ref. 160, 375–447
203. Shiffrin, R. M., Atkinson, R. C. 1969. Storage and retrieval processes in long-term memory. *Psychol. Rev.* 76:179–93
204. Simon, H. A., Kotovsky, K. 1963. Human acquisition of concepts for sequential patterns. *Psychol. Rev.* 70:534–46
205. Smallwood, R. D. 1971. The analysis of economic teaching strategies for a simple learning model. *J. Math. Psychol.* 8:285–301
206. Sperling, G. 1960. The information available in brief visual presentation. *Psychol. Monogr.* 74:1–29
207. Spiker, C. C. 1970. An extension of Hull-Spence discrimination learning theory. *Psychol. Rev.* 77:496–515
208. Steiner, T., Greeno, J. G. 1969. An analysis of some conditions for representing N state Markov processes as general all-or-none models. *Psychometrika* 34:461–87
209. Sternberg, S. 1966. High-speed scanning in human memory. *Science* 153:652–54
210. Suppes, P. 1969. Stimulus-response theory of finite automata. *J. Math. Psychol.* 6:327–55
211. Suppes, P. 1969. Stimulus-response theory of automata and TOTE hierarchies: A reply to Arbib. *Psychol. Rev.* 76:507–10
212. Suppes, P., Atkinson, R. C. 1960. *Markov Learning Models for Multiperson Interactions*. Stanford Univ. Press. 296 pp.
213. Suppes, P., Donio, J. 1967. Foundations of stimulus-sampling theory for continuous-time processes. *J. Math. Psychol.* 4:202–25
214. Suppes, P., Ginsberg, R. 1963. A fundamental property of all-or-none models, binomial distribution of responses prior to conditioning, with application to concept formation in children. *Psychol. Rev.* 70:139–61
215. Suppes, P., Groen, G., Schlag-Rey, M. 1966. A model for response latency in paired-associate learning. *J. Math. Psychol.* 3:99–128
216. Theios, J. 1963. Simple conditioning as two-stage all-or-none learning. *Psychol. Rev.* 70:403–17
217. Theios, J. 1965. The mathematical structure of reversal learning in a shock-escape T-maze: Overtraining and successive reversals. *J. Math. Psychol.* 2:26–52
218. Theios, J. 1968. Finite integer models for learning in individual subjects. *Psychol. Rev.* 75:292–307
219. Theios, J. 1971. *Aversive Conditioning and Learning*, ed. F. R. Brush, 297–346. New York: Academic. 626 pp.
220. Theios, J., Brelsford, J. W. Jr. 1966. A Markov model for classical conditioning: Application to eyeblink conditioning in rabbits. *Psychol. Rev.* 73:393–408
221. Theios, J., Brelsford, J. W. Jr. 1966. Theoretical interpretations of a Markov model for avoidance conditioning. *J. Math. Psychol.* 3:140–62
222. Thomas, H. B. G. 1968. An information-theoretic model for the serial position effect. *Psychol. Rev.* 75:409–20
223. Timpe, K. P. 1965. Einige Probleme der systemtheoretischen Analyse sensomotorischer Lernvorgänge. *Z. Psychol.* 171:400–9
224. Timpe, K. P. 1966. Die Beschreibung einfacher Trackingleistungen mit Hilfe eines zeitoptimalen Porzesses. *Elek. Inform. Kyber.* 2:107–19
225. Timpe, K. P. 1971. *Visual Information Processing and Control of Motor Activity*, 523–25. Sophia: Bulgarian Acad. Sci.
226. Trabasso, T., Bower, G. 1964. Component learning in the four-category concept problem. *J. Math. Psychol.* 1:143–69
227. Ibid 1966. Presolution dimensional shifts in concept identification: A test of the sampling with replacement axiom in all-or-none models. 3:163–73
228. Trabasso, T., Bower, G. H. 1968. *Attention in Learning: Theory and Research*. New York: Wiley. 253 pp.

229. Trabasso, T., Rollins, H., Shaughnessy, E. 1971. Storage and verification stages in processing concepts. *Cogn. Psychol.* 2:239–89
230. Uhr, L., Vossler, C. 1963. See Ref. 77, 251–68
231. Underwood, B. J. 1964. See Ref. 67, 47–78
232. Underwood, B. J., Schulz, R. W. 1960. *Meaningfulness and Verbal Learning*. New York: Lippincott. 430 pp.
233. Vitz, P. C., Todd, T. C. 1967. A model of learning for simple repeating binary patterns. *Psychol. Rev.* 75:108–17
234. Warburton, D. M., Greeno, J. G. 1970. General shape function model of learning with applications in psychobiology. *Psychol. Rev.* 77:348–52
235. Waugh, N. C., Norman, D. A. 1965. Primary memory. *Psychol. Rev.* 72:89–104
236. Waugh, N. C., Smith, J. E. K. 1962. A stochastic model for free recall. *Psychometrika* 27:141–54
237. Wickelgren, W. A. 1970. See Ref. 160, 67–102
238. Wickelgren, W. A., Norman, D. A. 1966. Strength models and serial position in short-term recognition memory. *J. Math. Psychol.* 2:316–47
239. Wickens, T. D., Millward, R. B. 1971. Attribute elimination strategies for concept identification with practiced subjects. *J. Math. Psychol.* 8:453–80
240. Wolford, G. 1971. Function of distinct associations for paired-associate performance. *Psychol. Rev.* 78:303–13
241. Wolford, G., Bower, G. H. 1969. Continuity theory revisited: Rejected for the wrong reasons? *Psychol. Rev.* 76:515–18
242. Wolter, D. G., Earl, R. W. 1969. The asymptotic values of response probabilities in the Audley-Jonkheere learning model. *Psychometrika* 34:203–14
243. Young, J. L. 1971. Reinforcement-test intervals in paired-associate learning. *J. Math. Psychol.* 8:58–81

CONTENTS

DEVELOPMENTAL PSYCHOLOGY, *Eugene S. Gollin and Mark Moody* . . . 1
BASIC DRIVES, *Matthew J. Wayner and Robert J. Carey* 53
MATHEMATICAL LEARNING THEORY AND THE NEW "MENTAL FORESTRY," *James G. Greeno and Robert A. Bjork* 81
STUDENT DEVELOPMENT AND COUNSELING, *Harold B. Pepinsky and Naomi M. Meara* 117
VISUAL SENSITIVITY, *John Lott Brown* 151
THE SENSE OF SMELL, *Trygg Engen* 187
PERCEPTION, *Herschel W. Leibowitz and Lewis O. Harvey, Jr* 207
MEASUREMENT OF PERSONALITY TRAITS: THEORY AND TECHNIQUE, *Allen L. Edwards and Robert D. Abbott* 241
PHYSIOLOGICAL PSYCHOLOGY: SLEEP, *Harold L. Williams, Frank A. Holloway, and William J. Griffiths*. 279
INTERPERSONAL ATTRACTION, *Donn Byrne and William Griffitt* 317
THE STUDY OF SMALL GROUPS, *Robert Helmreich, Roger Bakeman, and Larry Scherwitz* 337
PSYCHOLOGY AND CULTURE, *Harry C. Triandis, Roy S. Malpass, and Andrew R. Davidson* 355
PERSONNEL ATTITUDES AND MOTIVATION, *John B. Miner and H. Peter Dachler*. 379
ENVIRONMENTAL PSYCHOLOGY, *Kenneth H. Craik* 403
SOCIAL AND COMMUNITY INTERVENTIONS, *Emory L. Cowen* 423
SCALING, *Norman Cliff* 473
REPRINT INFORMATION 507
INDEXES:
AUTHOR INDEX 509
SUBJECT INDEX 531
CUMULATIVE INDEX OF CONTRIBUTING AUTHORS, VOLUMES 20 TO 24 . 548
CUMULATIVE INDEX OF CHAPTER TITLES, VOLUMES 20 TO 24 549