What Is Learned in Sequential Learning? An Associative Model of Reward Magnitude Serial-Pattern Learning

Douglas G. Wallace and Stephen B. Fountain
Kent State University

A computational model of sequence learning is described that is based on pairwise associations and generalization. Simulations by the model predicted that rats should learn a long monotonic pattern of food quantities better than a nonmonotonic pattern, as predicted by rule-learning theory, and that they should learn a short nonmonotonic pattern with highly discriminable elements better than 1 with less discriminable elements, as predicted by interitem association theory. In 2 other studies, the model also simulated behavioral "rule generalization," "extrapolation," and associative transfer data motivated by both rule-learning and associative perspectives. Although these simulations do not rule out the possibility that rats can use rule induction to learn serial patterns, they show that a simple associative model can account for the classical behavioral studies implicating rule learning in reward magnitude serial-pattern learning.

One recurring question, almost a leitmotif in the study of comparative cognition, is how best to characterize complex animal behavior. Can a complex sequence of behavior through time, for example, be understood as a complex emanation of simple associative processes? Do putatively complex behaviors demand explanation in terms of higher order cognitive processes? Similarly, in the field of sequential learning, a fundamental question that is not yet fully answered is "What is learned in sequential learning?" In animal sequential-learning research, claims that animals chunk information and form hierarchical representations to facilitate sequential learning and memory (Dallal & Meck, 1990; Fountain, Henne, & Hulse, 1984; Macuda & Roberts, 1995; Roberts, 1979; Terrace, 1987) have inspired research designed to determine what processes mediate chunking and related phenomena. For example, serial-learning research has investigated a number of factors thought to affect how animals encode sequences of events (Capaldi, 2002; Capaldi, Verry, Nawrocki, & Miller, 1984; Fountain, 1990; Fountain et al., 1984; Fountain & Rowan, 1995a; Fountain, Rowan, & Benson, 1999; Fountain, Wallace, & Rowan, 2002; Swartz, Chen, & Terrace, 1991; Terrace, 1987, 1991, 2002; Terrace & Chen, 1991a, 1991b). Evidence has accumulated that performance in sequential-learning tasks may be mediated by discrimination-learning processes (e.g., Capaldi, 1985, 1994; Capaldi & Miller, 1988; Fountain, Benson, & Wallace, 2000; Stempowski, Carman, & Fountain, 1999), by a representation of the serial position of items (e.g., Burns, Dunkman, & Detloff, 1999; Chen, Swartz, & Terrace, 1997; Roitblat, Pologe, & Scopatz, 1983), or by a representation of pattern organization through some form of rule learning (Fountain et al., 1984; Fountain & Rowan, 1995a, 1995b).

Traditionally, associative models of sequential learning have assumed that the critical factor controlling behavior in sequential tasks is associations between events in a sequence. According to this view, the subject learns that one stimulus (an element of behavior or an ordinal position cue) predicts the next stimulus (or behavioral element) in the sequence (e.g., Capaldi & Molina, 1979; Jensen & Rohwer, 1965). Thus, according to these models, sequence learning can be construed as a form of discrimination learning, and factors such as stimulus discriminability and stimulus generalization should be important determinants of behavior. Serial-position models assume that sequence elements become associated with their serial position (Burns & Gordon, 1988; Burns, Hulbert, & Cribb, 1990; Chen et al., 1997; Roitblat et al., 1983). Rule-learning (RL) models, on the other hand, stress central organizational processes. In the domain of sequential learning, for example, RL models propose that rats learn abstract "rules" to represent the structure they find in sequences (e.g., Hulse, 1978; Restle & Brown, 1970b).

Computational modeling can sometimes provide unique evidence for or against the notion that simpler processes can explain complex behavior of a given sort. A properly stated computational model can provide predictions about the operation of simple processes in complex settings and can thus provide a stronger argument than heuristic models or behavior alone for or against rejecting simpler explanations in favor of symbolic constructs (cf. Church, 1997). In these experiments, computational modeling was used to evaluate the claim that rats use symbolic mechanisms such as rule learning when they learn to track patterned sequences of events.

Early Research Supporting a Rule-Learning View of Rat Serial Learning

Several critical predictions of the RL theory of rat sequential learning have been tested using variations of standard runway paradigms in which rats learned to *track* patterns by running fast or

slowly in anticipation of food quantities presented in sequences. Perhaps the most important prediction from this model is that rats should be sensitive to pattern structure. A second prediction is that rats should show evidence that the rules they learn are in fact abstract.

Rats Are Sensitive to the Structure of Reward Magnitude Patterns

Several studies provide evidence that rats are sensitive to the formal structure of the serial patterns they learn. For example, Hulse and Dorsky (1977, 1979) created patterns from a stimulus "alphabet" (Jones, 1974) of various quantities of food. The patterns they created varied in formal structure. A strongly monotonic sequence of 14–7–3–1–0 food pellets was formally simple in structure, requiring a single "less than" rule to describe the relationships of all successive pairs of quantities. A weakly monotonic 14–5–5–1–0 pattern and a nonmonotonic 14–1–3–7–0 pattern were progressively more complex, requiring combinations of "less than," "greater than," and "equal" rules to adequately describe the patterns. Rats learned to track the formally simple monotonic pattern most rapidly and to the most proficient level of tracking performance of the three patterns, followed by the weakly monotonic and the nonmonotonic patterns. Hulse and Dorsky (1977) ruled out a variety of pattern-learning strategies the rats might have adopted to learn these patterns in favor of the hypothesis that rats are sensitive to pattern structure. A number of other studies have shown that rats' pattern-tracking performance can be predicted by the complexity of pattern structure when patterns are sequences of food quantities (Fountain, Evensen, & Hulse, 1983), sequences of brain-stimulation reward quantities (Fountain & Annau, 1984; Fountain, Schenk, & Annau, 1985), and sequences of flashing lights (Fountain, 1990; Fountain, Raffaele, & Annau, 1986). These results with rats paralleled those obtained with human subjects using patterns comprised of letters and numbers. For example, Simon and Kotovsky (1963) showed that easier patterns had simpler formal descriptions. They concluded that an almost perfect prediction of the difficulty of a pattern could be obtained by simply counting the number of symbols required to describe the formal structure of the pattern.

Rats' Rules Are Abstract

Two different studies provide evidence that the rules rats used in serial-pattern learning are, in fact, abstract. In one study (Hulse & Dorsky, 1979), rats demonstrated the ability to generalize a rule from one set of patterns to a different pattern. Rats were trained with a set of sequences that were either consistently monotonically decreasing or randomly increasing and decreasing. Later, rats were transferred to one of two sequences comprised of new quantities, either monotonically decreasing (16–9–3–1–0) or nonmonotonic (16–1–3–9–0). The differential pattern tracking observed for the four groups during the transfer phase indicated that training produced positive and negative transfer as predicted by RL theory. Specifically, monotonic training produced better tracking in the transfer phase than random training when both received a monotonic pattern in the transfer phase. Monotonic training produced poorer tracking in the transfer phase than random training when both received a nonmonotonic pattern in the transfer phase. In a second study, Fountain and Hulse (1981) evaluated rats' ability to extrapolate patterns of varied structural complexity. Rats were better able to predict a new 0-pellet element added to a structurally simpler 14–7–3–1 pattern than 0 pellets added to a structurally complex 14–3–7–1 pattern. The authors reasoned that the monotonic pattern produced better anticipation of 0 than the nonmonotonic pattern because the monotonically decreasing 14–7–3–1 sequence was described by a single "less than" rule that would lead naturally to the added 0 after 1.

The foregoing rule-generalization and extrapolation results in rats parallel analogous results obtained with human subjects. Restle and Brown (1970c), for example, reported results of transfer studies indicating that human subjects could generalize rule structures from one pattern to another. Restle and Brown (1970a) also showed that when different groups of subjects learned different patterns of similar formal structure, subjects' error profiles (error rates and distribution of errors in the patterns) were remarkably similar between patterns. Finally, Restle and Brown (1970a) showed that subjects tend to extrapolate runs and trills and even higher order structures. Thus, rats and humans appear to have functionally similar, though not equivalent, rule-generalization and extrapolation capacities.

The results of rat rule-generalization and extrapolation experiments have been taken to be some of the strongest evidence that rats are able to encode some representation of the abstract relational rules that describe the patterns that they learn. Rats' rules appear to be abstract in the sense that they relate pattern elements without being tied to the particular identity of the elements they describe. Rule-generalization and extrapolation results thus support the idea that rats are able to encode some representation of the structure of the pattern independent of the items that actually make up the pattern.

Sequential Learning Viewed as Discrimination Learning

As an alternative to rule learning, a potential means of learning to track sequences is to learn a series of stimulus–response or stimulus–stimulus associations relating successive stimuli of the series. Sequential learning by this strategy is essentially discrimination learning (cf. Capaldi & Molina, 1979), and this view of sequential learning has a long history (cf. Hull, 1931; Hunter, 1920; Skinner, 1934). In other work on rat sequential learning, for example, when sequences are comprised of quantities of food, the difficulty of anticipating the successive quantities and the vigor of rats' response in anticipation of a given quantity can be influenced by a number of associative factors under appropriate conditions. These associative factors include generalization of "reward–signal capacity" between similar quantities of the sequence (cf. Capaldi & Molina, 1979) and "remote anticipations" of quantities to be received beyond the immediate trial (cf. Capaldi & Miller, 1988; Capaldi, Nawrocki, & Verry, 1983), to name but two (cf. Capaldi, 2002).

Most important to our discussion is Capaldi's proposal that generalized reward–signal capacity is an important factor in pattern tracking (Capaldi & Molina, 1979). Simply put, this view supposes that each food quantity in a sequence serves as a cue for the next. Acquisition depends in part on discriminability and salience of pattern elements, but tracking performance is also determined by generalization between food quantity signals. For

example, the cue generalization idea would describe performance on a 14–3–7–1–0 sequence of food quantities as follows: Rats would anticipate 0 on the basis of memory of 1 that came to signal it, but anticipation of 0 would be modulated by the fact that 1, by generalization to the similar item 3, would also lead to some anticipation of 7. Other interactions would also be expected, in which the amount of generalization would depend, as it does in traditional discrimination tasks, on the relative similarity of cues. This kind of generalization between signals would be expected on the basis of traditional notions of discrimination learning (Blough, 1975; Spence, 1936, 1937).

Evidence supporting this interitem association (IA) theory has been generated from various runway studies. First, Capaldi and Molina (1979) demonstrated that a structurally complex sequence constructed from highly discriminable food quantity elements (1–29–0) produced better pattern tracking relative to a structurally simple sequence constructed of less discriminable elements (20–10–0). Second, in response to Hulse and Dorsky's (1977) rule-learning studies, Haggbloom (1985) demonstrated that pattern tracking was disrupted on transfer only when associative cues were removed but was unaffected by manipulations that violated rule or serial-position information. Finally, Haggbloom and Brooks (1985) showed that discriminability of pattern elements, not pattern structure, was the best predictor of pattern extrapolation.

The foregoing studies directly challenged the RL theory. However, the parametric weakness of the idea that cue generalization is an important factor in pattern tracking and the cognitive zeitgeist of the times worked in favor of the RL theory. IA theory had difficulty describing and predicting tracking in long sequences like those used by Hulse comprised of multiple, interacting (i.e., generalizing) stimuli drawn from a dimension of unknown psychophysical properties (viz., the food quantity dimension). For example, Capaldi, Verry, and Davidson (1980) stated that the generalized reward–signal capacity received by 1 in sequences such as 14–7–3–1–0 or 14–5–5–1–0 was "completely indeterminate" (p. 583). Heuristic hypotheses such as the idea that cue generalization is an important factor in pattern tracking—useful as they may be—simply cannot be expected to make stern predictions in complex behavioral paradigms without additional formalization. However, it should be equally clear that whereas verifying such a heuristic hypothesis is not likely under such circumstances, conclusively ruling out such an idea is equally difficult. With the foregoing in mind, a simple computational model with limited assumptions and parameters was used to evaluate whether rats' apparent rule-learning behavior could be accounted for by commonly accepted associative principles of animal discrimination learning.

The Sequential Pairwise Associative Memory (SPAM) Model: A Computational Model of Serial-Pattern Learning

In considering how to begin to model serial-pattern learning processes, a variety of different computational approaches were considered, from simple stochastic models to connectionist models to production system approaches. Each of these varieties of computational models has been applied to sequential learning and memory problems at one time or another. For example, "random walk" models have been advanced by Roitblat (1984) and, re-cently, Capaldi (e.g., Neath & Capaldi, 1996); connectionist models have been advanced by Murdock (1995a), among others; and a forerunner of production system models was pioneered on serial-pattern learning problems studied in humans by Simon, Newell, and their associates (Newell & Simon, 1961; Simon & Kotovsky, 1963). The principal concern was that the model should have characteristics of simple associative systems. Specifically, the computational engine should associate specific items in "memory" and produce generalization phenomena. A connectionist analog was chosen because it has these properties, and stochastic models and production systems were rejected because they are not easily stated in terms consistent with basic principles of discrimination learning. Several varieties of connectionist and related models have the desired properties and thus also bear at least superficial resemblance to Capaldi's item-association and cue-generalization ideas. In particular, the models developed by Murdock and Metcalfe (TODAM and CHARM, respectively; Eich, 1982; Metcalfe, 1990; Murdock, 1982, 1983) have these properties. These models have the added advantage that both Murdock's and Metcalfe's models have also been used successfully to simulate a broad array of human associative-learning and memory phenomena (Metcalfe, 1990, 1993; Murdock, 1982, 1983), including some rote sequential-learning phenomena (Murdock, 1983, 1992, 1995a).

The sequential pairwise associative memory (SPAM) model developed for the present simulations used the learning rule described by Murdock (1982, 1983) and Metcalfe (Eich, 1982; Metcalfe, 1990); and subsequently used by Metcalfe (1990) in her simulations of memory "blending" through generalization in eyewitness misinformation effects. However, SPAM can be differentiated from Murdock's and Metcalfe's models because SPAM is much more primitive and has none of the more elaborate mechanisms for simulating remote associations or "novelty monitoring" found in more recent versions of TODAM and CHARM, respectively (Metcalfe, 1993; Murdock, 1992, 1993). That is, SPAM shares a common learning rule with TODAM and CHARM, but other features of the model, such as how it anticipates events through generalization, differ significantly from TODAM and CHARM.

When successive pairs of items are associated in SPAM, the result of the association process is added to a common trace. Because the trace is a composite of all prior associations, retrieval can result in generalization when the same or similar cue items were earlier associated with two or more different items. This is the basis of generalization in SPAM. For example, items are represented by vectors in SPAM. Items such as 14 pellets, 7 pellets, 3 pellets, 1 pellet, and 0 pellets are represented by vectors. Because these items represent quantities of food that presumably vary systematically in similarity to each other, vectors that vary systematically in similarity can represent them. Under these conditions, if the composite trace contains information about the association of 1 and 0 only, cueing the composite trace with 1 retrieves 0. On the other hand, if the composite trace contains the associations 3 with 7 and 1 with 0, cueing the trace with 1 retrieves both 0 (retrieved directly by 1) and, to a lesser degree, 7 (cued by generalization of 1 to 3 because 1 and 3 are similar). This generalization would result in an "anticipation" of a quantity greater than 0. This effect is reminiscent of generalization typically observed in discrimination learning, and it is also reminiscent of Capaldi's idea of cue generalization, in which the capacity to

signal reward or nonreward can generalize from one cue to other similar cues. Because SPAM is an associative model that has characteristics of simple associative systems, namely, pairwise associations and generalization, it was used to test whether an instance of this type of mechanism could account for the phenomena that various investigators have reported in their studies of rat reward magnitude (food quantity) pattern learning.

Experiment 1: SPAM Model Simulations of Pattern-Tracking Results From Hulse and Dorsky (1977) and Capaldi and Molina (1979)

RL theory cannot account for Capaldi and Molina's (1979) data, which indicate that short nonmonotonic patterns like 1–29–0 were learned faster than short monotonically decreasing patterns like 20–10–0. RL theory predicts that the structurally simpler monotonic sequence should be easier to learn. IA theory explained the foregoing outcome by noting that the nonmonotonic sequence was comprised of items that were more discriminable than those of the monotonic sequence. In contrast, proponents of RL theory believed that IA theory could not account for Hulse and Dorsky's (1977) data, which indicate that long monotonically decreasing patterns like 14–7–3–1–0 were learned faster than long nonmonotonic sequences like 14–1–3–7–0 (Fountain, et al., 1984, 1985; Hulse, 1980; Hulse & Dorsky, 1977, 1979; Roitblat, 1982; Roitblat et al., 1983). RL theory was invoked to do so.

In Experiment 1, SPAM, which works according to the principles of pairwise interitem association and generalization, was used in an attempt to simulate pattern tracking for sequences analogous to those studied by Hulse and Dorsky (1977) and Capaldi and Molina (1979). Simulations motivated by the design of the original study by Hulse and Dorsky examined tracking of long monotonic (14–7–3–1–0) and nonmonotonic (14–1–3–7–0) patterns. In contrast, simulations motivated by the design of the original study by Capaldi and Molina examined short patterns of food quantities with high (1–29–0) and low (20–10–0) element discriminability. Additional simulations were performed to examine the effects of manipulating SPAM's encoding and retention parameters.

Method

The SPAM program. The SPAM program was written in C (Turbo C, Borland International, Scotts Valley, CA) for PC-compatible hardware. As in the model described by Metcalfe (Eich, 1982; Metcalfe, 1990) and Murdock (1982, 1983, 1985), items to be stored in memory in SPAM were represented as 63-element random vectors, in which each element of the vector is termed a "feature" of the item. The features of item vectors were random normal deviates so that features were statistically independent.[1] Deviates were restricted to the range −2.0 to 2.0, and item vectors were normalized so that the dot product of each vector with itself was 1. The dot product of any two unrelated item vectors should approach 0. An item file of 100 unrelated item vectors of this sort was generated.

Because the simulations required using items that varied systematically in similarity to each other, item *dimensions* (sets of items with stepwise increments and decrements in similarity) were also set up from which the items to be used as "reward magnitudes" (food quantities) in the simulations were drawn. The item dimensions were constructed in the same manner as the one used by Metcalfe (1990) in her simulation of color-shift phenomena in memory (so-called "memory blends"). Each dimension was comprised of 11 items. To produce such a dimension, we first created a set of independent (unrelated) items, such as Item Vectors **20–30**, using our

standard item-generation routine. Item Vectors **21–29** were then replaced by Item Vector **20**. Then, the first six features of Items **21–29** were set equal to the same features of Item **30**. Next, the adjacent six features of Items **22–29** were set equal to those of Item **30**. This process continued until the ninth set of six features was replaced for Item **29** only. (Note that Metcalfe, 1990, achieved the same result by randomly choosing the six features—without replacement—to be modified on each step of this routine. Because features are independent, the result is the same using either method.) An item in such a dimension differs from its neighbors by approximately 10% per step away in the dimension, and item vectors at the extremes of the dimension (i.e., Items **20** and **30**, in this case) are unrelated.

One assumption of the model is that perceived food quantity is a log-linear dimension. This assumption, based on the view that Weber's Law should apply to the perception of food quantity as it does to the perception of quantity in other stimulus dimensions, is common to theories of generalization and traditional notions of discrimination learning (Blough, 1975; Spence, 1936, 1937). This assumption was also implicitly accepted by Roitblat (1982), who attempted to apply Blough's (1975) model of generalization to the serial-pattern learning phenomena reported by Hulse and Dorsky (1979). To create a log-linear food quantity dimension, food quantities were represented by vectors in an approximate log-linear mapping that expected larger just noticeable differences as food quantity increased. In the following simulations, stimulus dimension Vector **0** represented 0 food pellets, Vector **1** represented 1 food pellet, Vector **3** represented 3 pellets, Vector **5** represented 7 pellets, Vector **6** represented 10 pellets, Vector **7** represented 14 pellets, Vector **8** represented 20 pellets, and Vector **9** represented 29 pellets.

In SPAM, two item vectors such as "14 pellets" (represented by Vector **A**) and "7 pellets" (represented by Vector **B**) may be associated by the process of *convolution*, which is computed according to the following equation:

$$(\mathbf{A} * \mathbf{B})_m = \sum_{(ij) \in S(m)} a_i b_j,$$

where $\mathbf{A}$ and $\mathbf{B}$ are the item vectors $(a_{-31}, a_{-30}, \ldots, a_{30}, a_{31})$ and $(b_{-31}, b_{-30}, \ldots, b_{30}, b_{31})$, respectively, that are being associated, and $S(m) = \{(i, j) \mid -(n-1)/2 \le i, j \le (n-1)/2, \text{ and } i + j = m\}$. If $n = 3$, that is, if item vectors are three elements (features) in length, it is straightforward to compute that the convolution of $\mathbf{A}$ and $\mathbf{B}$, called $\mathbf{A}*\mathbf{B}$, will result in the 5-element vector, $\mathbf{A}*\mathbf{B} = \{a_{-1}b_{-1}, a_0b_{-1} + a_{-1}b_0, a_1b_{-1} + a_0b_0 + a_{-1}b_1, a_1b_0 + a_0b_1, a_1b_1\}$, following Metcalfe (1990). A numerical example of the convolution of two 5-feature item vectors is shown in Figure 1.

"Encoding" is accomplished by adding the resulting trace vector and item vectors that were components of the association to the composite trace vector, $\mathbf{T}$, which is the composite memory trace: $\mathbf{T}_n = \alpha \mathbf{T}_{n-1} + \gamma_1 \mathbf{A} + \gamma_2 \mathbf{B} + \omega(\mathbf{A} * \mathbf{B})$, where $\mathbf{T}_{n-1}$ and $\mathbf{T}_n$ represent the trace before and after encoding information for a new event, respectively; $\mathbf{A}$ and $\mathbf{B}$ are vectors representing the individual items contributing to the association; and $\mathbf{A}*\mathbf{B}$ is the convolution of the vectors representing 14 pellets and 7 pellets. Following Murdock (1983), the Greek letter α represents a retention parameter that determines how much of the prior learning will be retained. (Note that Murdock, 1983, called this parameter a "forgetting" parameter.) The symbols γ_1 and γ_2 are weights for the item vectors, and ω is an encoding parameter that represents the strength of the convolution vector (the pairwise association) that is to be added to the composite trace vector. Encoding new information is accomplished simply by adding vectors to the trace vector, $\mathbf{T}$. That is, the model assumes only one composite trace, Trace

---

Item Vector A

	a_{-2}	a_{-1}	a_0	a_1	a_2		
	.301	-1.205	.387	.904	-.387		
	-.169	.677	-.217	-.508	.217	-.562	b_{-2}
-.169	.452	-1.811	.582	1.359	-.582	1.503	b_{-1}
1.129	-.243	.972	-.312	-.730	.312	-.807	b_0
-2.271	-.103	.413	-.133	-.310	.133	-.343	b_1
.943	.135	-.542	.174	.407	-.174	.450	b_2
1.812	-1.987	.176	.540	-.174			

(Left column labeled **Trace Vector**; right column labeled **Item Vector B**)

Resulting Trace = {-.169, 1.129, -2.271, .943, 1.812, -1.987, .176, .540, -.174}
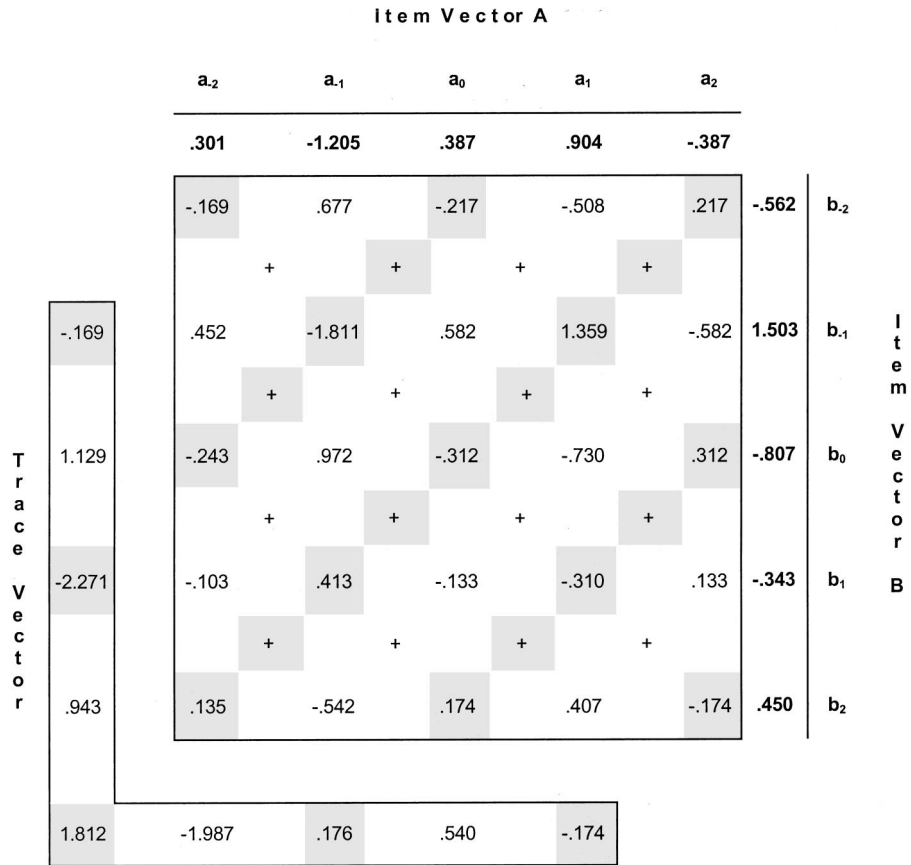
Figure 1. A numerical example of the convolution of two events, A and B, representing 2 food quantities received sequentially. Events A and B are represented by Item Vectors **A** and **B**. A and B are associated by multiplying each element of Item Vector **A** by each element of Item Vector **B** to produce a matrix of values, then summing along the diagonals to produce the resulting trace vector for this association. The resulting trace vector is then simply added to the composite memory trace vector containing all of the pairwise associations already accumulated through exposure to the sequence. Retrieval through correlation is a similar process, in which the matrix is generated by multiplying elements of the retrieval cue (one of the item vectors used in previous associations) by elements of the composite trace, then summing along diagonals to produce the retrieved trace. The retrieved trace is then compared with all possible items in the lexicon, and an anticipated item score is calculated.

Vector **T**, that contains all of the individual pairwise associations (convolutions) of items that compose a sequence of events.

"Retrieval" is accomplished by the process of *correlation*, which is represented by #, where a cue item is correlated with the memory trace. Correlation is defined as

$$R_m = \sum_{(ij)\in S(m)} a_i b_j,$$

where $S(m) = \{(i, j)| -(n-1)/2 \le i, j \le (n-1)/2,$ and $i - j = m\}$, following Metcalfe (1990). So, for example, **A** (the vector representing 14 pellets) might be correlated with the trace, **T**. This retrieval would be depicted as **A#T**, and it would recover a vector representation of the item previously associated with **A**, in this case, Vector **B** (which represents 7 pellets). The recovered item vector will be similar to the originally encoded item vector but will be somewhat degraded by noise inherent in the convolution and correlation processes. When the composite trace contains multiple convolutions with the same cue item (e.g., both **A*****B** and **A*****C**, where **C** represents 3 pellets), the result of correlation of **A** with the composite trace (i.e., **A#T**) will be a vector bearing information about each item originally associated with the cue, **A**. In this case, the recovered vector (**A#T**) would contain noisy representations of both Vector **B** (7 pellets) and Vector **C** (3 pellets) in composite form. To determine how a subject with such a composite memory trace would perform on a test, the recovered vector is compared with each possible item in the dimension (that is, with each possible food quantity in the dimension), and the most similar item is taken to be the item "anticipated" by the subject. For purposes of comparing item vectors, the similarity of two vectors is defined as the dot product of the vectors (cf. Metcalfe, 1990). In this example, similarity comparisons would be made by calculating the dot product of the recovered vector with each vector representing a food quantity in the dimension.

Procedure. Figure 2 provides a schematic of the general steps involved in an iteration of the computational model. For example, a simulation of
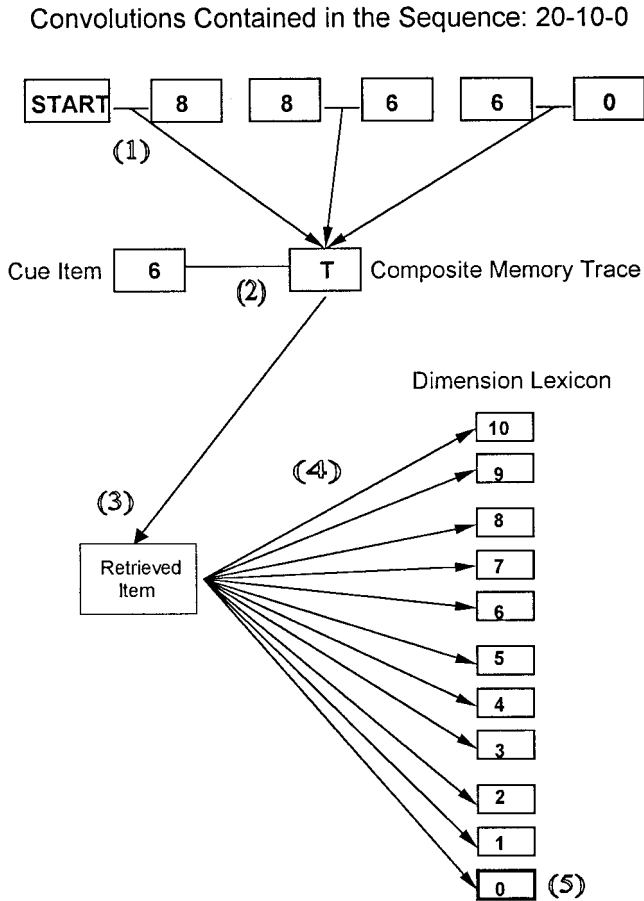
Convolutions Contained in the Sequence: 20-10-0

Figure 2. Flow diagram for one iteration of the sequential pairwise associative memory (SPAM) model for the sequence 20–10–0 (coded as Vector **8**–Vector **6**–Vector **0**), with 6 as a cue for the anticipated item. The schematic depicts the following steps: (1) association by convolution, (2) storage in a composite memory trace, (3) retrieval by correlation of a cue item with the composite memory trace, (4) determining dot products (similarity) of the retrieved item vector and each individual lexicon item vector, and (5) recognition by selecting the lexicon item with the highest similarity to the retrieved item.

performance for a 20–10–0 sequence would include three associations, hereafter called *convolutions*. The first convolution would be **Start*****8**, where "**Start**" is an independent item vector representing distinctive cues signaling the first trial of the sequence, and Item Vector **8** represents 20 pellets of food. Later convolutions would be **8*****6** and **6*****0**. The convolution process is labeled (*1*) in Figure 2. The item that was convolved with **0**, namely **6**, was later used to cue the composite memory trace, labeled (*2*) in Figure 2. This resulted in the retrieved item, labeled (*3*), that was compared with the items in the lexicon. The lexicon was *dimensional*, that is, the lexicon contained all the items of the stimulus dimension. The comparison process generated dot product values that reflected the degree of similarity between the retrieved item and each individual item in the lexicon, as indicated by (*4*) in Figure 2. The lexical item with the highest dot product value was chosen as the item best predicted by the cue for that iteration, shown as (*5*) in Figure 2. To obtain a good representation of the central tendency of the program's performance, 1,000 iterations were performed for each sequence simulation. It should be noted that in comparable studies with models having similar encoding and retrieval processes, namely TODAM and CHARM, 500 or 1,000 iterations per simulation are the norm (see, e.g., Eich, 1982, or Murdock, 1983).

Proportion responding to an item in the lexicon was calculated as the proportion of times out of 1,000 iterations that an item in the lexicon was chosen (as shown in Figure 3). The proportion responding scores for items in the lexicon were entered into the anticipated item equation: Anticipated Item = Σ[lexicon item * p (responding to lexicon item)], where "lexicon item" refers to the item vector number in the stimulus dimension. This equation weights the lexicon "quantity" by probability of response to produce weighted scores. Summing the weighted scores across all lexical items results in a value considered to reflect the overall anticipated item. The anticipated-item scores are considered analogous to running speeds elicited by the cue associated with the final item of the sequence. A smaller anticipated item represents slower running, and runway latency should be directly proportional to this score.

Table 1 shows sets of convolved items for the simulation series of Experiment 1. The amount of information contained in the trace was conserved across conditions. Irrelevant convolutions, that is, convolutions involving items independent of the stimulus dimension, were included as necessary to keep the amount of information in the composite memory trace constant at 8 convolutions total across sequence simulations, as shown in Table 1.

The four parameter conditions indicated in Table 1 reflect different assumptions about how memory for associated items may function. As one may recall, α is a retention parameter, γ_1 and γ_2 are weights for the item vectors, and ω is an encoding parameter. Item vector weights, γ_1 and γ_2, were set to zero throughout because adding information about individual items increases noise in the trace without affecting the general outcome of simulations. Parameter Condition A produced perfect encoding of new convolutions ($\omega = 1.0$), with perfect retention of past learning ($\alpha = 1.0$). Condition B produced perfect encoding of new convolutions ($\omega = 1.0$), but poorer retention ($\alpha = .5$) favored more recent convolutions as new con-
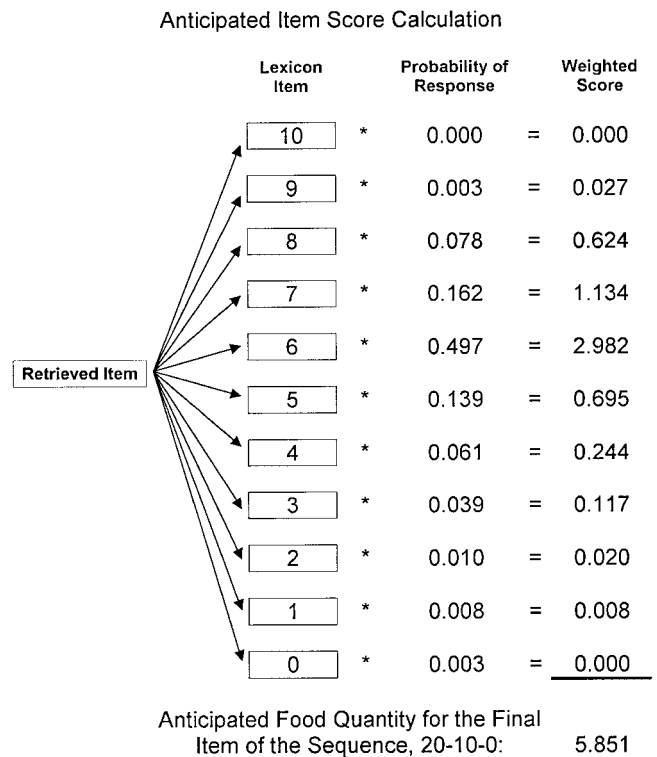
Anticipated Item Score Calculation

Lexicon Item		Probability of Response		Weighted Score
10	*	0.000	=	0.000
9	*	0.003	=	0.027
8	*	0.078	=	0.624
7	*	0.162	=	1.134
6	*	0.497	=	2.982
5	*	0.139	=	0.695
4	*	0.061	=	0.244
3	*	0.039	=	0.117
2	*	0.010	=	0.020
1	*	0.008	=	0.008
0	*	0.003	=	0.000

Anticipated Food Quantity for the Final
Item of the Sequence, 20-10-0:   5.851

Figure 3. Example calculation of an anticipated item score for Element 3 of the sequence 20–10–0 (coded as Vector **8**–Vector **6**–Vector **0**), where the probability of responding to an item of the lexicon was based on 1,000 iterations of the sequential pairwise associative memory (SPAM) model.

Table 1

SPAM Model Simulations of 3- and 5-Element Monotonic and Nonmonotonic Sequence Tracking in Experiment 1

Convolutions contained in the trace and cue items
for each simulated sequence

Convolutions contained in the short monotonic trace: 8–6–0	Convolutions contained in the short nonmonotonic trace: 1–9–0
Start*8	**Start*1**
8*6	**1*9**
6*0	**9*0**
+5 irrelevant convolutions	+5 irrelevant convolutions
Cue items: **Start**, **8**, and **6**	Cue items: **Start**, **1**, and **9**

Convolutions contained in the long monotonic trace: 7–5–3–1–0	Convolutions contained in the long monotonic trace: 7–1–3–5–0
Start*7	**Start*7**
7*5	**7*1**
5*3	**1*3**
3*1	**3*5**
1*0	**5*0**
+3 irrelevant convolutions	+3 irrelevant convolutions
Cue items: **Start**, **7**, **5**, **3**, and **1**	Cue items: **Start**, **7**, **1**, **3**, and **5**

Parametric manipulations performed on each sequence

A: $\alpha = 1.0$, $\omega = 1.0$	Perfect encoding and perfect retention
B: $\alpha = .5$, $\omega = 1.0$	Poor encoding and perfect retention
C: $\alpha = 1.0$, $\omega = .5$	Perfect encoding and poor retention
D: $\alpha = .5$, $\omega = .5$	Poor encoding and poor retention

Note. SPAM = sequential pairwise associative memory.

volutions were added to the trace. Condition C produced poorer encoding of new convolutions ($\omega = .5$) and perfect retention ($\alpha = 1.0$). Condition D produced both poorer encoding and retention ($\alpha = \omega = .5$).

Results and Discussion

Predictions from SPAM simulations parallel rat behavioral data observed in serial-pattern learning studies motivated by both RL and IA theories of sequential learning. Table 2 lists the anticipated food quantities for the final 0-pellet item in each sequence for Parameter Conditions A and B. In general, the final item of the short nonmonotonic sequence, **1**–**9**–**0** (the list of vectors representing the food quantity pattern 1–29–0), was tracked best relative to the other three sequences. The long monotonic sequence, **7**–**5**–**3**–**1**–**0** (the list of vectors representing the food quantity pattern 14–7–3–1–0), was the second best with regard to tracking the final item. The finding that the short nonmonotonic sequence produced better tracking for the final element than the short monotonic sequence qualitatively matches data reported by Capaldi and Molina (1979). In addition, although the difference was small in Parameter Condition A, the finding that the long monotonic sequence produced better tracking for the final element than the long nonmonotonic sequence matches data reported by Hulse and Dorsky (1977). This overall pattern of results was observed across both Parameter Conditions A and B, with a better match of simulation and behavioral results in Parameter Condition B, where the parameters produced perfect encoding of new convolutions ($\omega = 1.0$) and poorer retention of old, previously encoded convolutions ($\alpha = .5$). Parallel outcomes in Parameter Conditions A and

B indicate that manipulation of retention (the α parameter) in the model produces quantitative rather than qualitative differences in sequence tracking. It should be noted that SPAM is an end-state model that learns a scaled version of the final pattern of convolved vector features in a single pass. This means that SPAM predicts the relative levels of asymptotic performance rather than acquisition rates, and a small advantage of long monotonic over long nonmonotonic series in asymptotic anticipation of the final 0-pellet item is a close match to what Hulse and Dorsky (1977) actually reported.

Figures 4 and 5 show pattern profiles for short and long monotonic and nonmonotonic sequences for each parameter condition. The ordering of pattern difficulty observed in Table 2 was found for each parameter condition. Specifically, short sequences produced better tracking with a nonmonotonic item arrangement, whereas long sequence tracking was better with a monotonic arrangement of items. Quantitative predictions changed across parameter conditions, but changing α, that is, manipulating retention of old convolutions, was the only manipulation that affected the quantitative outcome. Reducing the retention of old convolutions by reducing alpha from 1.0 to .5, that is, creating a recency effect in the trace, produced a better quantitative fit of rat behavioral data for long monotonic and nonmonotonic sequence tracking. In contrast, manipulating ω in Parameter Conditions C and D produced pattern profiles identical to those of Parameter Conditions A and B, respectively. Thus, when old convolutions were equally retained, manipulating the strength of new convolutions at the time of encoding (the ω parameter) did not affect the simulation results.

To better understand how parameter conditions affected retrieval, the distribution of retrieved items was examined for items of each pattern under different parameter conditions. Figures 6 and 7 show the distribution of responding to items in the lexicon for the last 0 element of long and short monotonic and nonmonotonic sequences. For both patterns the distribution of responses changed when α, the retention parameter, was manipulated but not when ω, the encoding parameter, was manipulated. Changes in anticipated-item scores due to manipulating α, reported in Figures 4 and 5, were not only due to a shift in the peak of the distribution but also to changes in the shape of the distribution. This changed distribution of responses, which represents differential generalization to items in the lexicon, produced a better fit relative to the behavioral data with $\alpha = .5$. This suggests that α can be manipulated to

Table 2

Results of SPAM Model Simulations of 3- and 5-Element Sequence Tracking in Experiment 1

	Anticipated final food quantity	
Sequence	Parameter Condition A ($\alpha = 1.0$)	Parameter Condition B ($\alpha = .5$)
Short nonmonotonic: **1**–**9**–**0**	0.723	0.379
Long monotonic: **7**–**5**–**3**–**1**–**0**	2.561	1.184
Long nonmonotonic: **7**–**1**–**3**–**5**–**0**	2.583	2.422
Short monotonic: **8**–**6**–**0**	5.851	4.230

Note. Parameter conditions: $\alpha = 1.0$ or .5; $\gamma_1 = \gamma_2 = 0.0$; $\omega = 1.0$. SPAM = sequential pairwise associative memory.
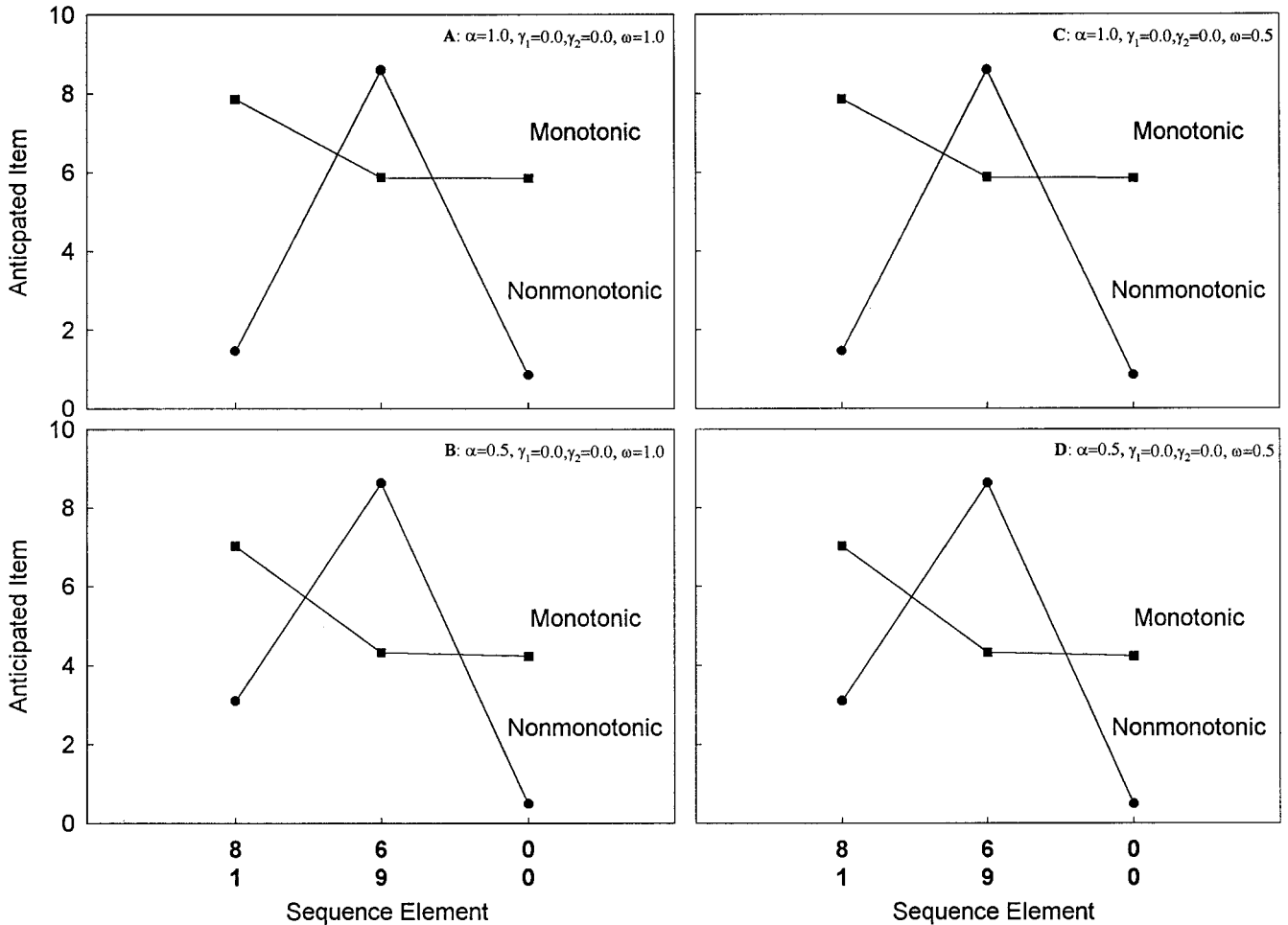
Figure 4. Anticipated item scores for each item of the monotonic 20–10–0 and nonmonotonic 1–29–0 patterns of Experiment 1 (coded as Vectors **8–6–0** and **1–9–0**, respectively) for each of the four parameter conditions. In Parameter Conditions A and B, encoding was maximized ($\omega = 1.0$) and retention was manipulated. In Parameter Conditions A and C, retention was maximized ($\alpha = 1.0$) and encoding was manipulated. Parameter Condition D completed the 2×2 design, with reduced retention ($\alpha = .5$) and reduced encoding ($\omega = .5$) relative to Parameter Condition A.

change effects quantitatively without changing the qualitative prediction that, in this case, short and long sequences should produce better tracking with nonmonotonic and monotonic item arrangements, respectively. Future work exploring the utility of manipulating α and ω to improve the quantitative predictions of SPAM may help us understand other factors that influence serial-pattern learning. For now it would seem that setting α to .5, thereby creating a recency effect, can produce a better quantitative fit of the data. It should be noted that the retention parameter, α, makes SPAM sequential because α values less than 1.0 differentially weaken earlier convolutions as later convolutions are added to the trace. Therefore, recently encountered information has a larger influence on what is anticipated than does information encountered earlier. Murdock (1995a) has stated that TODAM has always included such a parameter and offers a rationale for doing so in the domain of human memory. Additional research, computational and behavioral, is needed to examine the validity of doing so for

modeling rat serial-pattern learning. Considering that manipulations of the encoding parameter, ω, did not influence element anticipation, but manipulations of the retention parameter, α, did influence predicted performance, the following experiments restrict simulations to Parameter Conditions A ($\alpha = 1.0$, $\omega = 1.0$) and B ($\alpha = .5$, $\omega = 1.0$).

The foregoing results indicate that SPAM, a mathematical model with properties of simple associative learning (pairwise associations and generalization), can provide qualitative predictions about serial-pattern learning that match behavioral data. The results provide initial evidence that a simple associative account may be a reasonable challenge to the RL account of serial learning. However, to examine whether the results obtained were somehow an accident of the particular sequences chosen for study, Experiment 2 examined the extent to which the foregoing results might generalize to other sequences of varying length and item discriminability.
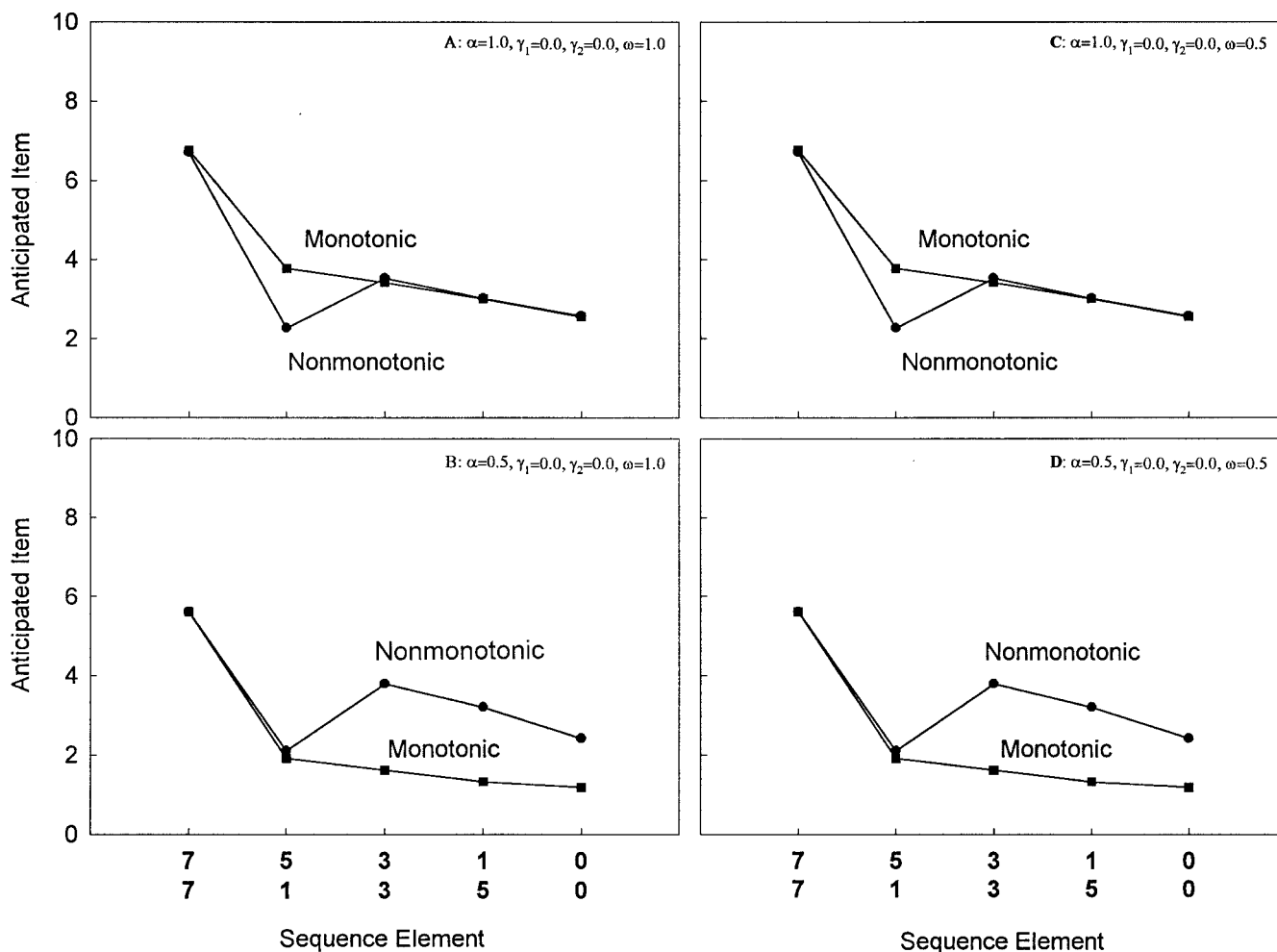
Figure 5. Anticipated item scores for each item in the monotonic 14–7–3–1–0 and nonmonotonic 14–1–3–7–0 patterns of Experiment 1 (coded as Vectors **7–5–3–1–0** and **7–1–3–5–0**, respectively) for each of the parameter conditions shown in Figure 4.

Experiment 2: SPAM Model Simulations of Pattern Tracking for 3-, 4-, 5-, 6-, and 7-Item Sequences

Experiment 2 tested the extent to which the results from Experiment 1 can be generalized to other sequences of varied length and element discriminability. These simulations were a small set of the total possible combinations of values for the variables sequence length, element discriminability, and item arrangement.

Method

The procedures of Experiment 1 were used to simulate pattern tracking for a variety of sequences (see Table 3). Experiment 2 used Parameter Conditions A ($\alpha = 1.0$, $\gamma_1 = \gamma_2 = 0.0$, $\omega = 1.0$) and B ($\alpha = .5$, $\gamma_1 = \gamma_2 = 0.0$, $\omega = 1.0$) for all simulations. As in Experiment 1, irrelevant convolutions were included as necessary to keep the amount of information in the composite memory trace constant at a total of 10 convolutions for all sequence simulations.

Results and Discussion

Table 3 contains the results of simulations for various sequences of 3 to 7 items under Parameter Conditions A and B. For Parameter

Condition A, sequences comprised of 3 or 4 items produced better tracking with a nonmonotonic arrangement of items. When sequence length was increased to 5 items, monotonic sequences rivaled pattern tracking of corresponding nonmonotonic sequences. Table 3 also contains results for sequences comprised of 6 or 7 items. The final item was consistently better predicted with a monotonic arrangement of items for these sequences. Under Parameter Condition B, sequences comprised of 3 items produced better tracking with a nonmonotonic arrangement of items. Sequences comprised of 4 items marked the transition from nonmonotonic sequences producing better tracking to monotonic sequences producing better tracking. Sequences with 5, 6, and 7 items consistently produced better tracking under monotonic sequence arrangements. As sequence length increased from 3 to 7 items, then, there appears to have been a change in which item arrangements most consistently produced the best pattern tracking. This conclusion was supported by regressions of relative pattern tracking (monotonic vs. nonmonotonic) with respect to sequence length for both parameter conditions, shown in Figure 8. Both regressions were significant ($p < .05$) and accounted for 64% and
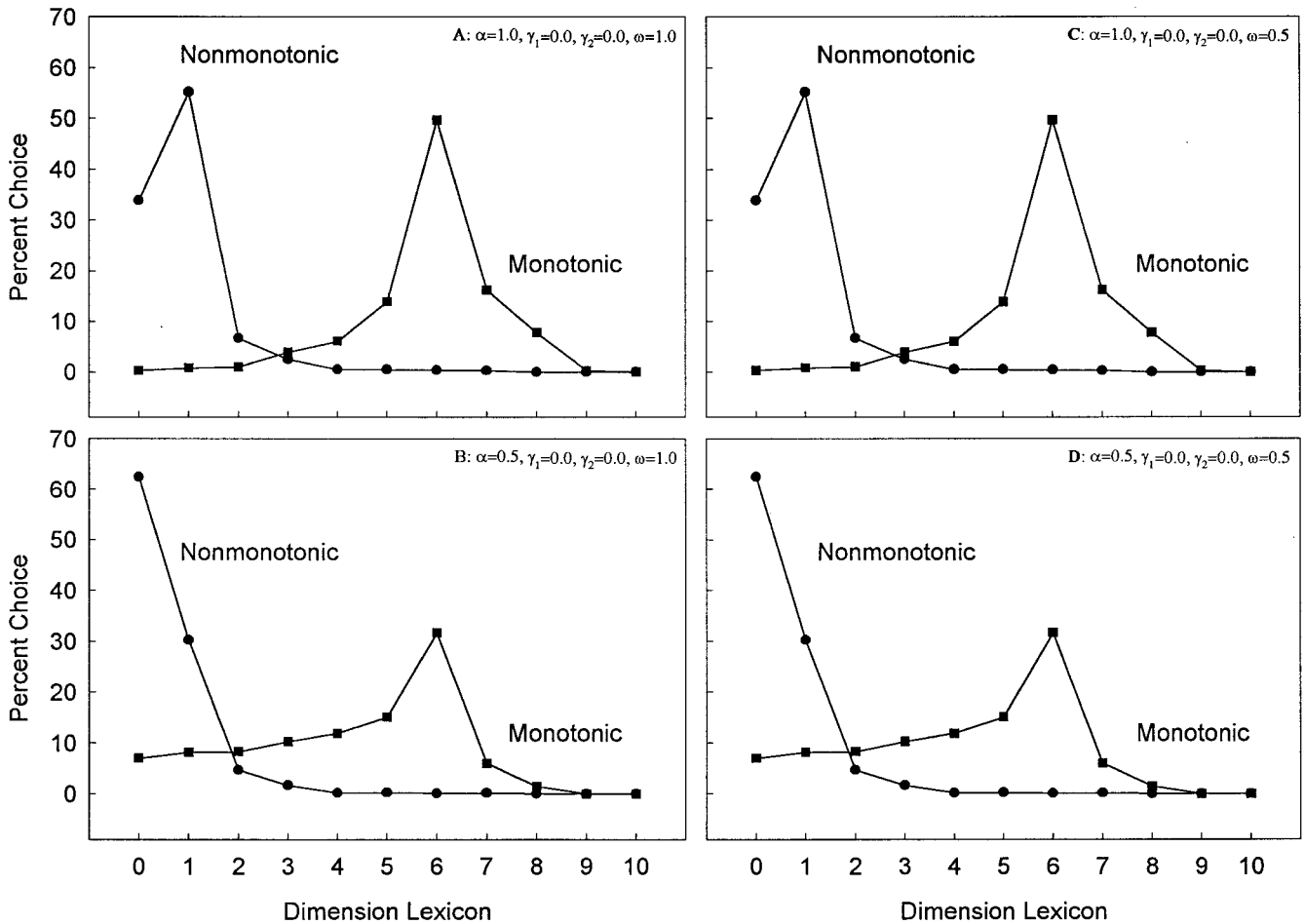
Figure 6. The distribution of responding to items in the lexicon for the final element of the monotonic 20–10–0 and nonmonotonic 1–29–0 sequences of Experiment 1 for each of the four parameter conditions shown in Figure 4.

88% of the variance for Parameter Conditions A and B, respectively. However, on the basis of these data alone it is unclear how sequence length is related to the observed effect that a nonmonotonic arrangement of items is better for short patterns and a monotonic arrangement is better for long patterns. It should be noted that, according to IA theory and SPAM, the monotonic versus nonmonotonic distinction is essentially artificial in this case. However, the results suggest a novel prediction about what appears to be a potentially important issue, namely, sequence length effects in serial-pattern learning. The results from these simulations also indicate that across a variety of stimulus codings, the effects observed in Experiment 1 are also found, further supporting the generality of the findings of Experiment 1. However, only a small set of all possible sequences have been simulated here. The lack of relevant behavioral data corresponding to these novel predictions indicates that more work, both behavioral and computational, is warranted to explore the possible significance of sequence length in serial learning from a discrimination-learning perspective.

Finally, other simulations showed that when the stimulus dimension was removed by using independent rather than related items, the sequence-length effect was not observed, that is, pattern

length and item order no longer influenced sequence tracking.[2] When independent items were chosen, the model no longer made predictions reminiscent of rats' behavior when learning serial patterns of food quantities.

---

[2] Sequences comprised of independent items were examined for sequence order and length effects, namely, that one ordering of stimuli is better for short sequences than for long sequences. The independent items were 63-feature vectors, where each feature was drawn from a random number generator, but the corresponding dot product between any two vectors was required to be approximately 0. As in earlier simulations, the next to last item of the sequence was considered the cue, and the final item of any sequence was the target to be anticipated for each simulation. Sequences of 3, 4, 5, 6, and 7 items with either "X" or reordered "Y" arrangements were simulated. As in Experiment 2, the only differences between simulations were the sequence length and item arrangement. Parameters for the simulations were those used in Experiment 2 (i.e., Parameter Condition A: $\alpha = 1.0$, $\gamma_1 = \gamma_2 = 0.0$, and $\omega = 1.0$; Parameter Condition B: $\alpha = .5$, $\gamma_1 = \gamma_2 = 0.0$, and $\omega = 1.0$). No differences were observed between parameter conditions. When the stimulus dimension was removed, the sequence-length effect was not observed, that is, pattern length and item order no longer influenced sequence tracking. When
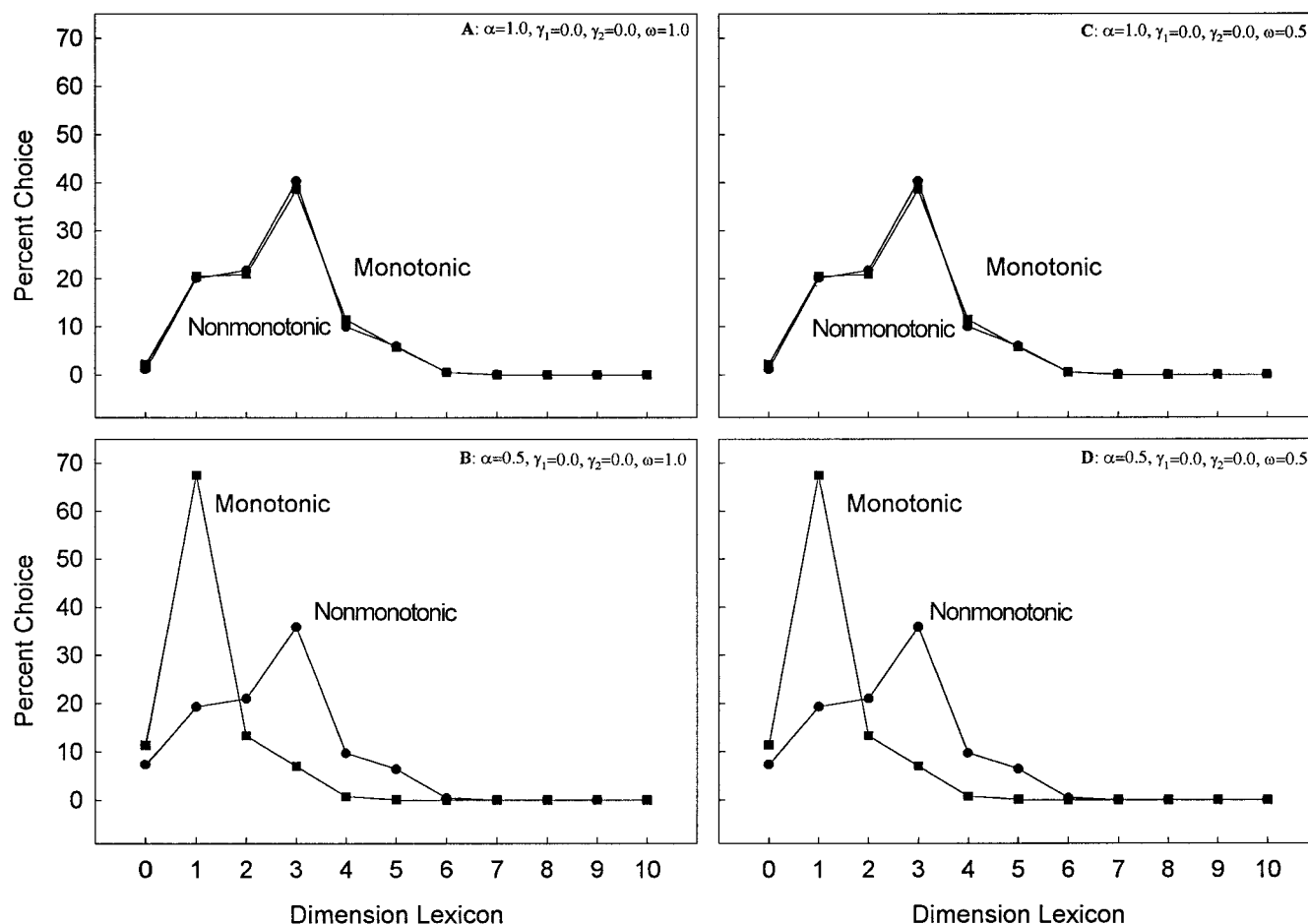
Figure 7. The distribution of responding to items in the lexicon for the final element of the monotonic 14–7–3–1–0 and nonmonotonic 14–1–3–7–0 sequences of Experiment 1 for each of the four parameter conditions shown in Figure 4.

Experiment 3: SPAM Model Simulations of Rule Generalization From Hulse and Dorsky (1979) and Associative Transfer From Haggbloom (1985)

Hulse and Dorsky (1979) demonstrated what they believed to be rule generalization in rats by training rats with one set of

independent items were chosen, the model no longer made predictions reminiscent of rats' behavior when learning serial patterns of food quantities.

RL theory's concept of formal structure and IA theory's concept of generalization are both based on the assumption of a stimulus dimension. The RL theory requires a stimulus dimension so that a stimulus has properties that allow ordinal relationships to exist between it and other stimuli, for example, less than or greater than (Hulse, 1978; Jones, 1974). These relations are the basis for the relational rules that create pattern structure. The IA theory demands a stimulus dimension so that generalization can occur between stimuli predicting different outcomes. The results of this series of simulations do not differentially support IA theory over RL theory or vice versa, but they do indicate the importance of using elements from a stimulus dimension to simulate serial-pattern learning phenomena with SPAM. The results show that SPAM's predictions in Experiments 1 and 2 depended critically on drawing stimuli from a stimulus dimension. More details of these simulations can be obtained from the authors.

patterns then transferring the rats to a new pattern. When the pattern presented in transfer was structurally consistent with training patterns, pattern learning in the transfer phase was fast. In contrast, when training and transfer patterns were structurally different, pattern learning in the transfer phase was retarded. Considering that item and association information could not be reliable predictors of sequence items on transfer, Hulse and Dorsky (1979) concluded that in the transfer phase rats had generalized the rules abstracted during training. Their results are considered strong evidence for the RL position.

Haggbloom (1985) provided evidence that stimulus associations learned during training mediate pattern tracking during transfer, not rule or serial-position information. In a procedure somewhat different from that of Hulse and Dorsky (1979), Haggbloom required all rats to learn a single pattern in training before transfer to one of four different patterns. The four transfer patterns manipulated the food quantities used in the pattern, the order of food quantities, and pattern length to evaluate the role of item associations, pattern structure, and serial position in determining pattern tracking. His use of transfers revealed that when rule or serial-position information was manipulated, serial-pattern tracking was unaffected. Manipulating pairwise associations in transfer, however, disrupted serial-pattern tracking. The simulations in Experi-

Table 3
Results of SPAM Model Simulations of Pattern Tracking for 3-, 4-, 5-, 6-, and 7-Element Monotonic (M) and Nonmonotonic (N) Sequences in Experiment 2

Sequence length	Sequence	Anticipated final food quantity				Sequence length	Sequence	Anticipated final food quantity			
		$\alpha = 1.0$	Better	$\alpha = .5$	Better			$\alpha = 1.0$	Better	$\alpha = .5$	Better
3	10–1–0	1.466		0.839		5	7–5–3–1–0	2.561	M	1.184	M
3	1–10–0	0.723	N	0.379	N	5	7–1–3–5–0	2.583		2.422	
3	8–6–0	5.851		4.230		5	8–6–5–1–0	3.661		1.187	M
3	6–8–0	5.759	N	3.525	N	5	8–1–5–6–0	3.618	N	3.423	
3	8–1–0	1.360		0.875		5	8–5–3–1–0	2.478	M	1.174	M
3	1–8–0	0.990	N	0.601	N	5	8–1–3–5–0	2.624		2.440	
3	10–5–0	4.958		3.412		5	10–8–6–5–0	6.055		4.806	M
3	5–10–0	3.915	N	1.615	N	5	10–5–6–8–0	5.787	N	4.886	
3	10–3–0	3.091		2.174		5	10–6–3–1–0	2.298	M	1.129	M
3	3–10–0	2.084	N	0.943	N	5	10–1–3–6–0	2.570		2.244	
3	8–3–0	3.075		2.277		5	10–5–3–1–0	2.371	M	1.154	M
3	3–8–0	2.723	N	1.583	N	5	10–1–3–5–0	2.746		2.470	
4	8–6–3–0	4.415		2.774	M	5	10–6–5–3–0	4.527	M	2.970	M
4	8–3–6–0	3.606	N	2.838		5	10–3–5–6–0	4.529		4.035	
4	8–6–1–0	2.559		1.017	M	6	10–8–6–3–1–0	2.863	M	1.161	M
4	8–1–6–0	1.687	N	1.218		6	10–1–3–6–8–0	3.702		3.486	
4	6–4–3–0	3.500		2.784	M	6	10–8–6–5–1–0	4.138	M	1.219	M
4	6–3–4–0	3.337	N	2.934		6	10–1–5–6–8–0	4.628		4.159	
4	6–3–1–0	1.882		1.079	M	6	8–6–5–3–1–0	3.163	M	1.239	M
4	6–1–3–0	1.834	N	1.579		6	8–1–3–5–6–0	3.614		3.713	
4	10–8–1–0	2.434		0.952		6	10–6–5–3–1–0	3.045	M	1.229	M
4	10–1–8–0	1.278	N	0.801	N	6	10–1–3–5–6–0	3.743		3.723	
4	10–8–3–0	4.810		2.610		6	10–8–6–5–3–0	5.034	M	3.052	M
4	10–3–8–0	3.343	N	2.237	N	6	10–3–5–6–8–0	5.082		4.496	
4	10–6–3–0	4.228		2.713	M	7	10–8–6–5–3–1–0	3.492	M	1.251	M
4	10–3–6–0	3.690	N	2.820		7	10–1–3–5–6–8–0	4.333		4.312	
4	10–6–5–0	5.497		4.577		7	10–8–6–4–2–1–0	2.653	M	1.148	M
4	10–5–6–0	5.348	N	4.561	N	7	10–1–2–4–6–8–0	3.750		3.887	
4	10–6–1–0	2.168		0.980	M	7	8–7–6–5–4–2–0	4.498	M	2.163	M
4	10–1–6–0	1.784	N	1.236		7	8–2–4–5–6–7–0	4.741		4.838	
4	10–5–1–0	2.022	M	1.015	M	7	8–6–5–4–3–1–0	3.451	M	1.303	M
4	10–1–5–0	2.023		1.498		7	8–1–3–4–5–6–0	3.835		4.030	
4	10–5–3–0	3.818		2.750	M	7	10–9–6–5–3–1–0	3.350	M	1.249	M
4	10–3–5–0	3.683	N	3.049		7	10–1–3–5–6–9–0	4.080		3.739	
4	10–8–5–0	6.486		4.531		7	7–6–5–3–2–1–0	2.673	M	1.177	M
4	10–5–8–0	5.472	N	3.950	N	7	7–1–2–3–5–6–0	3.086		3.636	
4	10–8–6–0	7.108		5.438							
4	10–6–8–0	6.438	N	5.078	N						

Note. Parameter conditions: $\alpha = 1.0$ or $.5$; $\gamma_1 = \gamma_2 = 0.0$; $\omega = 1.0$. SPAM = sequential pairwise associative memory.

ment 3 were directed toward accounting for both Hulse and Dorsky's (1979) rule generalization phenomena and Haggbloom's work that challenged the RL interpretation of serial-pattern learning phenomena.

Method

Hulse and Dorsky (1979) tested rule generalization by training rats on monotonic or nonmonotonic combinations of the food quantities, 10, 5, 3, 1, or 0 pellets. Rats were then transferred to either a monotonic 16–9–3–1–0 pattern or a nonmonotonic 16–1–3–9–0 pattern. Hulse and Dorsky's (1979) rule-generalization study was modeled using the list of convolutions found in Table 4. Training consisted of pairwise convolutions of items from a stimulus dimension. Though analogous in other regards, Hulse and Dorsky's (1979) study and the simulations differed in that for the simulations, the item associations used in training were not used at all in the patterns of the transfer phase. For example, the monotonic training sequence was **10*8**, **8*6**, **6*4**, **4*2**, and
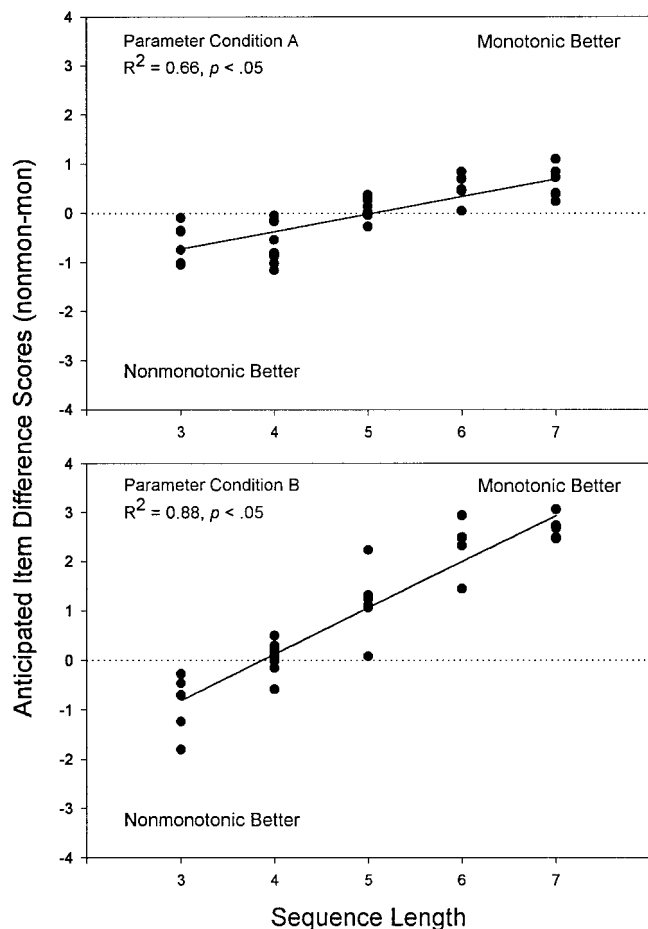
Figure 8. A scatter plot and regression line ($R^2 = .64$, $p < .05$) relating sequence length to anticipated item difference scores for the 3-, 4-, 5-, 6-, and 7-item sequences of Experiment 2. Difference scores less than 0 indicate that tracking for the nonmonotonic pattern was better than tracking for the monotonic pattern, whereas difference scores greater than 0 indicate that tracking for the monotonic pattern was better. nonmon–mon = nonmonotonic–monotonic.

2*0, whereas the monotonic pattern of the transfer phase was **9*5**, **5*3**, **3*1**, and **1*0**. Thus, these patterns logically result in a stronger test of generalization than patterns containing some of the same items, as in Hulse and Dorsky's (1979) experiment. All of the foregoing were convolved and entered into the composite memory trace. Pattern tracking for the final element of the transfer phase pattern was tested by using 1 as a cue to probe the composite trace.

For other training and transfer sequences, pairwise convolutions of items were entered in the same manner, and pattern tracking of the final element of the transfer phase patterns was tested by using 1 or 7 as indicated to probe the composite trace. The same techniques were applied for modeling the associative transfer results from Haggbloom (1985), as shown in Table 4. In the simulations of Haggbloom's associative transfer experiment, SPAM was trained with a monotonic **7–5–3–1–0** pattern before transfer to one of four patterns. Haggbloom used a 5-element **14–7–3–1–0** pattern as the training pattern. For the simulations, this pattern was coded as **7–5–3–1–0** as it was in Experiment 1 above. Haggbloom's transfer patterns were two 5-element patterns and two 6-element patterns, and they were coded for simulations as shown in Table 4. It should be noted that Haggbloom included patterns longer than the training pattern to test for

serial position as a factor controlling rats' response to the pattern, so the simulations likewise included 6-element patterns in transfer for the same purpose. As in the simulations of Hulse and Dorsky's (1979) experiment, for training and transfer sequences, pairwise convolutions of items were entered in the same manner as before, and pattern tracking of the final element of the later transfer phase patterns was tested by using 1, 3, or 7 as indicated as a cue to probe the composite trace. Parameters for the simulations were those used in Experiment 2 (i.e., Parameter Condition A: $\alpha = 1.0$, $\gamma_1 = \gamma_2 = 0.0$, and $\omega = 1.0$; Parameter Condition B: $\alpha = .5$, $\gamma_1 = \gamma_2 = 0.0$, and $\omega = 1.0$).

Results and Discussion

The results, shown in Table 5, for Parameter Condition A parallel behavioral data reported by both Hulse and Dorsky (1979) and Haggbloom (1985). In the rule generalization simulations, both groups receiving a monotonic pattern in the transfer phase anticipated smaller quantities on the final 0 trial than the groups receiving the nonmonotonic pattern in the transfer phase. This result might be expected simply on the basis of the results of Experiment 1, where 5-element monotonic patterns generally produced better tracking of the final 0 element than corresponding nonmonotonic patterns. However, training with a set of monoton-

Table 4
SPAM Model Simulations of Behavioral Rule Generalization and Associative Transfer in Experiment 3: Convolutions Contained in the Trace Followed by the Cue Items for Each Simulation

Simulation of Hulse & Dorsky (1979) Group name reflects training and transfer patterns	
M–M group Training: M **10*8, 8*6, 6*4, 4*2, 2*0** Transfer: M **9*7, 7*3, 3*1, 1*0** Cue item: **1**	M–N group Training: M **10*8, 8*6, 6*4, 4*2, 2*0** Transfer: N **9*1, 1*3, 3*7, 7*0** Cue item: **7**
R–M group Training: R **10*4, 4*6, 6*8, 8*2, 2*0** Transfer: M **9*7, 7*3, 3*1, 1*0** Cue item: **1**	R–N group Training: R **10*4, 4*6, 6*8, 8*2, 2*0** Transfer: N **9*1, 1*3, 3*7, 7*0** Cue item: **7**

Simulation of Haggbloom (1985) Group name reflects information conserved on transfer	
Serial position + rule group Training: M **Start*7, 7*5, 5*3, 3*1, 1*0** Transfer: Serial position + rule **Start*9, 9*7, 7*5, 5*3, 3*0** Cue item: **3**	Serial position group Training: M **Start*7, 7*5, 5*3, 3*1, 1*0** Transfer: Serial position **Start1*9, 9*5, 5*3, 3*7, 7*0** Cue item: **7**
Association + rule group Training: M **Start*7, 7*5, 5*3, 3*1, 1*0** Transfer: Association + rule **Start*9, 9*7, 7*5, 5*3, 3*1, 1*0** Cue item: **1**	Association group Training: M **Start*7, 7*5, 5*3, 3*1, 1*0** Transfer: Association **Start*9, 9*5, 5*7, 7*3, 3*1, 1*0** Cue item: **1**

Note. Parameter conditions: $\alpha = 1.0$ or $.5$; $\gamma_1 = \gamma_2 = 0.0$; $\omega = 1.0$. M = monotonic; N = nonmonotonic; R = random; SPAM = sequential pairwise associative memory.

Table 5
Results of SPAM Model Simulations of Rule Generalization and Associative Transfer in Experiment 3

Simulation of Hulse and Dorsky's (1979) Rule Generalization Study			
		Anticipation on 0 trial	
Training	Transfer	$\alpha = 1.0$	$\alpha = .5$
Monotonic	Monotonic	3.124	1.146
Monotonic	Nonmonotonic	4.277	2.013
Random	Monotonic	3.620	1.143
Random	Nonmonotonic	4.021	2.013

Simulation of Haggbloom's (1985) Associative Transfer Study		
	Anticipation on 0 trial	
Information preserved in transfer	$\alpha = 1.0$	$\alpha = .5$
Association + rule	2.800	1.194
Association	2.845	1.219
Serial position + rule	3.434	2.875
Serial position	5.040	2.914

Note. SPAM = sequential pairwise associative memory.

ically decreasing item associations facilitated tracking a monotonic pattern in the transfer phase compared with the random training condition. Furthermore, monotonic training reduced tracking of a nonmonotonic pattern in the transfer phase compared with the random training condition. Both of these effects are consistent with the predictions of RL theory, because positive transfer was observed for structurally similar training and transfer patterns, and negative transfer was observed for structurally dissimilar patterns. These simulation results thus parallel the Hulse and Dorsky (1979) behavioral results that have been taken as strong evidence of rule generalization in rats, yet SPAM has no rule-induction mechanism. The same effects Hulse and Dorsky (1979) observed in rats were produced in these simulations entirely by interitem associations and generalization. However, Parameter Condition B reduced the retention parameter from $\alpha = 1.0$ to $\alpha = .5$, thereby decreasing the influence of training on item anticipation for the transfer sequences. One general result of reducing the strength of prior interfering associations was that all groups performed better (anticipated smaller quantities) when α equalled .5 than when α equalled 1.0. Reducing retention of training associations and improving overall performance should naturally reduce any transfer of training effects, which was the result observed when α was reduced from 1.0 to .5 in this experiment. Under Parameter Condition B, where retention was reduced, item anticipation was determined entirely by the pattern encoded in the transfer phase.

In the simulations of Haggbloom's (1985) associative transfer experiment, SPAM was trained with a monotonic **7–5–3–1–0** pattern before transfer to one of four patterns. Under both parameter conditions, the best anticipation of the final 0 element (the smallest anticipated item on the final 0 trial) in transfer was for the nonmonotonic **9–5–7–3–1–0** pattern of the association condition, followed closely by the monotonic **9–7–5–3–1–0** pattern of the association + rule condition (as shown in Table 5). In the association condition, the terminal **3–1–0** associations were preserved

between the monotonic pattern of training and the nonmonotonic transfer pattern, but the nonmonotonic transfer pattern was structurally different from the monotonic pattern of training, and the 0 element appeared in a different serial position in transfer compared with training. Conditions without associative information, namely the serial position condition and the serial position + rule condition, produced poorer anticipation of the final 0 element (i.e., their anticipated item scores were higher). Thus, in these simulations, when various potential information relevant to guiding pattern tracking was eliminated in transfer, patterns that preserved associative information produced better pattern tracking than those that preserved rule or serial-position information. The simulation results paralleled the behavioral results reported by Haggbloom. In Experiment 3, SPAM simulated both the rule generalization phenomenon, previously embraced by Hulse and Dorsky (1979) as evidence favoring RL theory, and associative transfer phenomena, which had been presented by Haggbloom as a challenge to RL theory.

Experiment 4: SPAM Model Simulations of Pattern Extrapolation Results of Fountain and Hulse (1981) and Haggbloom and Brooks (1985)

The RL and IA theories prompted studies examining rats' ability to extrapolate patterns of food quantities. Fountain and Hulse (1981) studied rats' ability to extrapolate patterns of varied structural complexity. Rats initially learned 4-element patterns of food quantities. The patterns were 14–7–3–1, 14–5–5–1, and 14–3–7–1, where the patterns are listed in order of increasing formal complexity. After training on 4-element patterns, rats received a test of their ability to extrapolate the sequence to anticipate a 0-pellet element added to the sequence. When a nonrewarded fifth trial was added to the sequence, rats' extrapolation performance was best for the monotonic 14–7–3–1 pattern and progressively worse as structural complexity increased (Fountain & Hulse, 1981). That is, on the added fifth trial, rats in the monotonic pattern condition ran slowest in anticipation of the added 0-pellet trial compared with other conditions. These results were taken as support for the RL view of pattern learning. In contrast, when Haggbloom and Brooks (1985) manipulated element discriminability, they found an instance in which a structurally more complex pattern, 14–9–1–1, was extrapolated better than a structurally simpler pattern, 14–7–3–1. This latter result challenged the RL explanation of extrapolation behavior.

Pattern extrapolation is perhaps the strongest evidence in support of RL theory because it requires rats to anticipate a sequence item that has never been experienced before. The fact that rats respond differentially on the extrapolation trial is taken as prima facie evidence for rule learning that would seem to defy explanation by traditional associative views (Fountain & Hulse, 1981). How can a rat anticipate an item that it has never experienced except by rule extrapolation? Although it was not too difficult for us to conceive how (or concede that) other serial-pattern learning phenomena might be explained by SPAM's generalization phenomena, we approached simulating pattern extrapolation with SPAM with considerable skepticism. However, no model of serial-pattern learning would be complete without accounting for serial-pattern extrapolation studies. With the foregoing in mind, Experiment 4 was designed to determine whether SPAM could simulate

serial-pattern extrapolation of sequences of varying structural complexity and element discriminability. Given SPAM's success in our earlier studies, one might imagine that SPAM would be a particularly good candidate model for accounting for extrapolation behavior if a sufficient associative mechanism actually exists. With that in mind, we also reasoned that failure to simulate pattern extrapolation would provide support for tentatively rejecting the simpler associative–generalization explanation common to IA theory and SPAM in favor of a symbolic construct, namely rule learning, in the domain of pattern extrapolation behavior.

Method

Sequence extrapolation was modeled by associating items as shown in Table 6. For example, the composite memory trace for the sequence **7–5–3–1** would contain the convolutions **Start*7**, **7*5**, **5*3**, and **3*1**. To generate the anticipated item during extrapolation, 1 was used as a cue. Because 0 had not been used in the convolutions, anticipation after 1 reflected a capacity to predict a food quantity on the basis of generalization from experiences with other food quantity stimuli. Parameters for the simulations were those used in Experiments 2 and 3 (i.e., Parameter Condition A: $\alpha = 1.0$, $\gamma_1 = \gamma_2 = 0.0$, and $\omega = 1.0$; Parameter Condition B: $\alpha = .5$, $\gamma_1 = \gamma_2 = 0.0$, and $\omega = 1.0$).

Results and Discussion

The results from both parameter conditions, shown in Table 7, were consistent with the behavioral literature. The weak monotonic, highly discriminable sequence, **7–6–1–1** (coded to reflect the food quantity pattern 14–9–1–1), produced the best extrapolation relative to the other three sequences. This is consistent with the results of Haggbloom and Brooks (1985), in which increased item discriminability produced a structurally complex pattern that fostered better extrapolation than structurally simpler patterns. Further, the other sequences, **7–5–3–1**, **7–4–4–1**, and **7–3–5–1**

Table 6

SPAM Model Simulations of Strong Monotonic, Weak Monotonic, and Nonmonotonic Sequence Extrapolation in Experiment 4: Information Contained in the Trace and Cue Items for Each Simulated Sequence

Convolutions contained in the strong monotonic trace: 7–5–3–1	Convolutions contained in the nonmonotonic trace: 7–3–5–1
Start*7	**Start*7**
7*5	**7*3**
5*3	**3*5**
3*1	**5*1**
+4 irrelevant convolutions	+4 irrelevant convolutions
Cue item: **1**	Cue item: **1**
Convolutions contained in the weak monotonic trace: 7–4–4–1	Convolutions contained in the weak monotonic trace: 7–6–1–1
Start*7	**Start*7**
7*4	**7*6**
4*4	**6*1**
4*1	**1*1**
+4 irrelevant convolutions	+4 irrelevant convolutions
Cue item: **1**	Cue item: **1**

Note. Parameter conditions: $\alpha = 1.0$ or .5; $\gamma_1 = \gamma_2 = 0.0$; $\omega = 1.0$. SPAM = sequential pairwise associative memory.

Table 7

Results of SPAM Model Simulations of Pattern Extrapolation in Experiment 4

Sequence	Anticipated food quantity on the extrapolation trial	
	$\alpha = 1.0$	$\alpha = .5$
Weakly monotonic: 7–6–1–1	2.608	1.214
Strong monotonic: 7–5–3–1	3.703	2.905
Weakly monotonic: 7–4–4–1	4.033	3.777
Nonmonotonic: 7–3–5–1	4.588	4.404

Note. SPAM = sequential pairwise associative memory.

(coded for the patterns 14–7–3–1, 14–5–5–1, and 14–3–7–1, respectively), follow the pattern of results reported by Fountain and Hulse (1981). That is, pattern extrapolation was progressively worse for patterns described as having increasingly more structural complexity. SPAM, as do discrimination-learning theories in general, relies on the association of items in sequence and generalization between items to account for sequential learning and extrapolation. According to Haggbloom and Brooks (1985), the memory for the 1-pellet quantity in the **7–6–1–1** pattern was a more discriminable signal of small reward or nonreward than 1 in other sequences simulated here. This results in reduced generalization from other pattern elements that would tend to signal larger rewards, and the result is that 1 in **7–6–1–1** produces anticipation of a smaller quantity—better extrapolation—than in the patterns **7–5–3–1**, **7–4–4–1**, and **7–3–5–1**. In these latter patterns, there was more generalization between 1 and other elements in the pattern that signaled larger rewards. Despite our initial doubts and expectations, SPAM was able to simulate well-established behavioral extrapolation phenomena. From our perspective, this outcome was almost completely unanticipated.

General Discussion

One of the problems in the serial-pattern learning literature that complicated the RL versus IA debate was that IA theory with generalization (Capaldi & Molina, 1979; Capaldi, Verry, & Davidson, 1980) could not make strong predictions concerning patterns that contained more than a few items. For example, Capaldi, Verry, and Davidson (1980) stated that the generalized reward–signal capacity received by 1 in sequences such as 14–7–3–1–0 or 14–5–5–1–0 was "completely indeterminate" (p. 583). In other words, even though 1 signaled nonreward (0) by way of direct association, IA theory predicts that cue generalization should cause 1 to acquire the ability to signal reward through generalization from other similar items in series. It was thought that the amount of generalization was crucial in that it determined how much 1 would signal reward versus nonreward. However, the amount of reward–signal capacity 1 received through generalization was indeterminate presumably because, as suggested by Roitblat (1982), Capaldi et al. (1980) could not explicitly describe the requisite generalization functions for food reward quantities. For that reason, Capaldi et al. were unable to make strong predictions regarding learning for 5-item series that would have allowed them to claim that IA theory could describe the available behavioral

data. Instead, at the time they were forced to adopt the weaker position that the cue generalization interpretation "cannot be excluded" (Capaldi et al., 1980, p. 584).

The simulations of Experiments 1–4 demonstrated the power of simple principles of pairwise association and stimulus generalization to explain a number of classic RL behavioral phenomena. SPAM simulations showed that a mathematical model with the general features of simple associative learning, like those observed in discrimination learning, can account for the critical phenomena of serial-pattern tracking in patterns of different length and item arrangements (Experiment 1 and 2), rule generalization (Experiment 3), and pattern extrapolation (Experiment 4) from the rat sequential-learning literature. SPAM provides an existence proof for a purely associative, subsymbolic mechanism that can account for several critical behavioral phenomena previously taken as strong evidence for rule learning in rats. The results from Experiment 2 also lead to the conclusion that sequence length may be an important factor influencing serial-pattern tracking as a result of purely associative processes, and this idea needs to be confirmed by additional behavioral tests. Thus, although rats may in fact use rule-induction processes to learn serial patterns comprised of food quantities, the results of the SPAM simulations reported above indicate that this is not a necessary conclusion from the behavioral experiments that motivated this work.

SPAM is nearly the simplest possible formulation of this associative-memory model because it does not store information about remote associations between sequential events, the effects of extraneous cues like phrasing, or the passage of time or serial position. This is not to say that these factors play no role in serial-pattern learning; the reported simulations simply tested whether a model without these features could be sufficient to describe what may be considered the critical phenomena for the RL versus IA theory debate in the reward magnitude serial-pattern learning literature. More recent work, particularly Murdock's work on his model known as TODAM (Murdock, 1982, 1983, 1985), has significantly elaborated the basic model to account for a broader range of human memory phenomena, including sequential memory effects (Murdock, 1992, 1993, 1995a). From this perspective, it is perhaps remarkable that a simple associative model like SPAM based on pairwise associations and generalization between sequential items can account for so many of the critical phenomena of reward magnitude pattern learning in rats.

It should be noted that SPAM is an example of a class of "holographic" associative memory models that has accounted for a wide variety of psychological phenomena in the human literature: speed–accuracy trade-off in item recognition, list–strength effect, list–length effect, prototype abstraction, A–B A–D paradigm, Osgood transfer surface, novelty monitoring in control and Korsakoff subjects, serial-order information, recognition–failure function, and blended memories in eye-witness testimony (Eich, 1982; Hockley & Murdock, 1987; Metcalfe, 1990, 1991, 1993; Murdock, 1983, 1995a, 1995b). To this impressive list can now be added two important features of animal learning, namely discriminative control through pairwise associations and cue generalization, which have been used to account for animal behavior in sequential learning and schedules of reinforcement paradigms.

SPAM as an Associative Model of Reward Magnitude Serial-Pattern Learning

Associative theories posit that stimuli can come to control behavior. According to associative theories, when multiple stimuli signal different events, say reward and nonreward events, the similarity of the stimuli that serve as cues determines the difficulty of acquiring discriminative responding. The more similar the stimuli are, the more difficult the discrimination will be, because the signal capacity of one stimulus can generalize to other related stimuli. For example, when one stimulus, S+, signals reinforcement and another stimulus, S−, signals nonreinforcement, S− can receive capacity to signal reinforcement from S+. How much S− signals reinforcement via generalization from S+ increases as similarity between S+ and S− increases. As Roitblat (1982) aptly observed, IA theory "gains its power not from the association mechanism but from generalization" (p. 366), and its emphasis on generalization distinguishes it from other theoretical perspectives on sequential learning and memory such as RL theory (Fountain et al., 1983; Fountain & Hulse, 1981; Hulse, 1978; Hulse & Dorsky, 1977, 1979) and working or episodic memory models (e.g., Kesner, Measom, Forsman, & Holbrook, 1984; Olton, Shapiro, & Hulse, 1984; Sands & Wright, 1980). SPAM is an associative model with cue generalization in that related items can come to signal different events and generalization can occur between related cue items. As has been shown, SPAM is also able to simulate critical phenomena described in the reward magnitude pattern learning literature. We now must examine SPAM's assumptions and characteristics to determine which are necessary or sufficient for producing the pattern of results obtained in the foregoing studies.

As the name *sequential pairwise associative memory* suggests, the SPAM model encodes sequential information as pairwise associations between successive events. It should be recognized, however, that SPAM, like the cue generalization idea, gains its power from generalization, not from the associative mechanism per se. Generalization is independent of the association (convolution) and retrieval (correlation) processes in SPAM. Generalization depends on similarity between vectors representing the items to be remembered, and therefore generalization can be reduced or eliminated by reducing or eliminating similarity between vectors. When generalization is removed by using independent item vectors, SPAM no longer simulates behavioral data from the rat reward magnitude serial-learning literature (see Footnote 2). This would seem to support the view that generalization is in fact the sine qua non of reward magnitude serial learning, as claimed by the associative theories generally and by the cue generalization idea in particular.

It should be mentioned that this latter conclusion, that generalization among cues is the critical factor in determining serial-pattern performance, has been evaluated and accepted in the face of our initial predisposition to the contrary. First, it should be clear that our bias going into this project was that animals do, in fact, use symbolic rule-induction processes in reward magnitude serial learning (Fountain, 1986; Fountain et al., 1983, 1984; Fountain & Hulse, 1981). Second, SPAM also has characteristics other than generalization that could figure prominently in its ability to simulate the behavioral data. In particular, SPAM is also characterized by a distributed (vector) representation of items, a composite

memory trace, pairwise associations, symmetrical associations, and parameters for determining encoding and forgetting. Thus, it is possible that SPAM's successful simulation of the behavioral data might depend on the interaction of generalization based on vector similarity with some other feature such as distributed representation or symmetrical associations. Let us consider these other features of SPAM successively.

First, does SPAM's ability to simulate the behavioral data depend on representing items in distributed form? We did not test this issue directly. A stimulus dimension can be represented by a series of item vectors that differ in the proportion of features they share with independent items representing the extremes of a dimension. For example, Metcalfe (1990) created an 11-item dimension to represent a color dimension with successive vectors differing by approximately 10%. Metcalfe (1990) used this color dimension to simulate "memory blending" in eyewitness testimony. In Experiments 1–4, SPAM used an analogous procedure to create a stimulus dimension of food quantities with ordinal relations. However, it is clear that item generalization can be modeled using a scalar, rather than a distributed, representation of items and their similarity, as shown in Blough's (1975) model of discrimination learning and generalization, which was derived from the Rescorla–Wagner model (Rescorla & Wagner, 1972). Roitblat (1982) showed how one interpretation of Blough's model failed to account for the behavioral outcomes of Hulse and Dorsky (1977), but further scrutiny of Blough's model may be in order in light of the fact that SPAM and Blough's model have similar generalization processes and that SPAM was indeed able to simulate the general features of the reward magnitude serial-pattern learning behavioral data.[3] In principle, although distributed representation adds potentially desirable features such as "graceful degradation," it does not seem likely as a factor necessary for simulating the behavioral data.

Second, does SPAM's ability to simulate the behavioral data depend on representing items in a composite trace? This question was not directly tested, but a few comments are in order. In SPAM, associations in memory are stored as a weighted sum of the vectors representing past associations. As indicated before, SPAM is essentially the same model as the version of Metcalfe's (1990) CHARM used to simulate "misleading information" and "memory blending" effects in human eyewitness testimony, but SPAM is simpler than later "novelty monitoring" versions of CHARM (Metcalfe, 1993). It is clear from Metcalfe's (1990) work on misleading information and memory blends that these effects depend on simultaneous recall of information retrieved by similar cues, not on composite storage per se. Misleading information effects, memory blending, and generalization among food quantities all result from competing information at retrieval. The same effects could be achieved from simultaneous recall into working memory of independently stored traces because these effects do not depend on alteration of the original memories by later experiences (resulting from composite storage, presumably) but by recovery of multiple unmodified memories that compete in working memory at the time of test. The conclusion, then, is that composite storage is not critical for simulating the behavioral data.

Third, does SPAM's ability to simulate the behavioral data depend on representing sequential information as pairwise associations? Murdock's (1993, 1995a) TODAM2 has recently been elaborated to include "multiple convolutions" to produce effects akin to remote associations. Capaldi (1994) has also argued for the importance of remote associations in sequential memory. Surprisingly, SPAM has proven remarkably effective at simulating rat behavioral data without resorting to other elaborations that have been required to simulate human cognitive phenomena, including list learning. Only additional work will determine whether elaborations of the model will be required to simulate other features of rat sequential learning such as Capaldi's "remote anticipations" (cf. Capaldi et al., 1983; Capaldi & Miller, 1988) and whether such additions to the model will preclude modelling the phenomena discussed in this article.

Fourth, does SPAM's ability to simulate the behavioral data depend on symmetrical item associations? This assumption depends on the observation of associative symmetry in behavioral data (cf. Asch & Ebenholtz, 1962). The data are mixed with respect to the presence of associative symmetry as a result of conditional discrimination learning. Pigeons (Hogan & Zentall, 1977), rhesus monkeys, and baboons (D'Amato, Salmon, Loukas, & Tomie, 1985; Sidman et al., 1982) have either lacked associative symmetry or have only shown weak symmetry. This is in contrast with other more recent studies that demonstrate strong associative symmetry in pigeons (Zentall & Urcuioli, 1993), sea lions (Schusterman & Kastak, 1993), chimpanzees (Tomonaga, Matsuzawa, Fujita, & Yamamoto, 1991), and rats (Bunsey & Eichenbaum, 1996). In addition, associative symmetry has also been demonstrated in normal adults (Lazar, 1977), normal children (Sidman & Tailby, 1982), and mentally retarded teenagers (Sidman, 1971). Although there have been studies demonstrating associative symmetry, the factors that influence it require more study. Therefore, one must question the utility of assuming symmetrical associations. Although removing associative symmetry from SPAM is currently not an option, comparison of SPAM to other models not endowed with associative symmetry may provide insight into the validity of assuming symmetrical associations.

Fifth, does SPAM's ability to simulate the behavioral data depend on specific parameters for determining encoding and forgetting? In Experiments 1–4, manipulations of the retention and encoding parameters (i.e., α and ω, respectively) did not appear to change the overall outcome (i.e., whether the monotonic or nonmonotonic pattern produced the better pattern tracking). As an explanation, it should be noted that SPAM's association process, convolution, is essentially Hebbian rather than error correcting in nature. The fundamental process involves multiplication of corresponding feature values for the associated events, then weighting

---

[3] Two points deserve attention concerning the predictions made by Roitblat (1982) in applying the Blough (1975) model to the serial-pattern learning phenomena reported by Hulse and Dorsky (1979). First, one result neglected by Roitblat was that between-groups predictions from the Blough model matched behavioral results obtained by Hulse and Dorsky (1979)—namely, a 5-element monotonic pattern produced better anticipation of the terminal element than did a nonmonotonic pattern. Secondly, Roitblat focused on the difference between the final 2 items of sequences as an index of pattern learning. Hulse and Dorsky (1979) never evaluated statistically the differences between the final two items of sequences in their article, but they did evaluate the number of pattern repetitions before differential responding was observed between the initial and terminal quantities in each sequence. The results were consistent with predictions of the Blough model for initial and terminal item scores.

the resulting association vector by a learning parameter, ω, before adding it to the memory trace vector. The result is that SPAM is an end-state model that learns a scaled version of the final pattern of weights in a single pass. Different values of ω will not change the pattern of output, nor will repetition of the same set of convolutions change the output unless the retention parameter, α, is less than 1.0. This was the pattern of results observed in the reported experiments, where changing the learning parameter had no effect but changing the retention parameter modulated the size, but not the direction of observed differences between patterns. Blough's (1975) model relies on a different association process derived from the Rescorla–Wagner model (1972). Thus, Blough's model relies on an error-correcting (delta) rule equivalent to the Widrow-Hoff least-means-squares rule (cf. Gluck, Reifsnider, & Thompson, 1990). Given that Blough's model simulates generalization for similar stimuli in discrimination-learning tasks and may yet be able to simulate the behavioral data, the distinction between learning rules used in SPAM and Blough's model is likely to be inconsequential. Thus, SPAM can be restated in terms of an error-correcting rather than a Hebbian learning rule so that characteristics of acquisition and extrapolation of monotonic versus nonmonotonic patterns can be examined, as recent work in our lab shows (Wallace, Lewis, Fountain, & Block, 1999). This suggests that SPAM is a member of a class of formally equivalent or similar associative models that share common features—association and generalization—that produce common behavioral outcomes in sequential learning paradigms. The significant conclusion is that there exists at least one subsymbolic mechanism that can explain sequential behavior of a type previously thought to require symbolic processes (Gallistel, 1995; Hulse, 1978; Lashley, 1951; Roitblat & von Fersen, 1992).

Conclusion

Roitblat (1982), in a somewhat ironic twist at the time, appealed to parsimony to support the RL theory over associative memory and cue generalization as an explanation of serial-pattern learning in rats. He suggested that subsymbolic associative theories actually require rats to maintain more information about stimuli than is required by a symbolic rule-based model of sequential learning. This idea is based on the fact that associative theories require generalization based on interval-level properties of item similarity, whereas RL theory requires only ordinal-level properties—that is, RL theory requires only that rats recognize that the item set is ordered. Although in theory this may be true, in actuality both psychophysical and stimulus-generalization studies with a variety of species and stimulus dimensions support the view that animals are indeed sensitive to the interval-level properties of stimuli drawn from a dimension. Classical views of generalization and stimulus discrimination depend on this latter notion (cf. Spence, 1937).

As Capaldi repeatedly claimed, his IA theory viewed sequential learning as an albeit complex example of traditional instrumental learning, whereas RL theory requires postulating a "unique form of learning that requires human serial-pattern learning models for its explanation" (Capaldi et al., 1980, p. 583). In fact, the theoretical significance of the RL hypothesis in animal-learning theory centers on the proposed nature of pattern representation as depending on interitem relations or hierarchical relations—that is, rules and

the requisite symbolic rule-induction processes—rather than subsymbolic item associations having properties reflected in traditional generalization studies. This debate over representation in animal memory—namely, whether what is encoded is symbolic relational rules versus subsymbolic stimulus properties—continues vigorously in several areas of animal learning and cognition research, most recently in the domain of "transitive inference" research with species as diverse as pigeons, rats, rhesus monkeys, and chimpanzees (Boysen, Berntson, Shreyer, & Quigley, 1993; Couvillon & Bitterman, 1992; Davis, 1992; Gillan, 1981; Markovits & Dumas, 1992; McGonigle & Chalmers, 1992; Roberts & Phelps, 1994; Steirn, Weaver, & Zentall, 1995; Treichler & Van Tilburg, 1996; Weaver, Steirn, & Zentall, 1997; Wynne, 1997; Zentall & Sherburne, 1994). Similar debates can also be found in the human literature, in which the argument over the necessity for postulating specialized rule-induction processes to explain complex behavior has been recently reinvigorated by research on connectionist models of learning, memory, and language acquisition and production (e.g., Rumelhart & McClelland, 1987). In this article, it has been demonstrated that SPAM, a subsymbolic associative memory model, can simulate what have been taken to be examples of cognitive phenomena in rats—rule induction, rule generalization, and rule extrapolation. The results add to a growing body of evidence that animal sequential behavior, even putatively "complex" processes such as phrasing effects on chunking (Capaldi, 2002; Capaldi et al., 1984; Stempowski et al., 1999; Terrace, 1987), can be understood in terms of subsymbolic mediating processes. Whereas one must be cautious not to throw out the baby with the bath water, our results suggest that a healthy skepticism should be maintained toward nonassociative and "emergent" processes in cognitive theorizing.

References

Asch, S. E., & Ebenholtz, S. M. (1962). The principle of associative symmetry. *Proceedings of the American Philosophical Society, 106,* 135–163.

Blough, D. S. (1975). Steady state data and a quantitative model of operant generalization and discrimination. *Journal of Experimental Psychology: Animal Behavior Processes, 104,* 3–21.

Boysen, S. T., Berntson, G. G., Shreyer, T. A., & Quigley, K. S. (1993). Processing of ordinality and transitivity by chimpanzees (*Pan troglodytes*). *Journal of Comparative Psychology, 107,* 208–215.

Bunsey, M., & Eichenbaum, H. (1996). Conservation of hippocampal memory function in rats and humans. *Nature, 379,* 225–257.

Burns, R. A., Dunkman, J. A., Jr., & Detloff, S. L. (1999). Ordinal position in the serial learning of rats. *Animal Learning & Behavior, 27,* 272–279.

Burns, R. A., & Gordon, W. U. (1988). Some further observations on serial enumeration and categorical flexibility. *Animal Learning & Behavior, 16,* 425–428.

Burns, R. A., Hulbert, L. G., & Cribb, D. (1990). A test for order relevance in a three-element serial learning task. *Journal of General Psychology, 117,* 91–98.

Capaldi, E. J. (1985). Anticipation and remote associations: A configural approach. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11,* 444–449.

Capaldi, E. J. (1994). The sequential view: From rapidly fading stimulus traces to the organization of memory and the abstract concept of number. *Psychonomic Bulletin & Review, 2,* 156–181.

Capaldi, E. J. (2002). The discriminative stimulus and response enhancing properties of reward produced memories. In S. B. Fountain, M. Bunsey,

J. H. Danks, & M. K. McBeath (Eds.), *Animal cognition and sequential behavior: Behavioral, biological, and computational perspectives* (pp. 91–113). Norwell, MA: Kluwer Academic.

Capaldi, E. J., & Miller, D. J. (1988). The rat's simultaneous anticipation of remote events and current events can be sustained by event memories alone. *Animal Learning & Behavior, 16,* 1–7.

Capaldi, E. J., & Molina, P. (1979). Element discriminability as a determinant of serial-pattern learning. *Animal Learning & Behavior, 7,* 318–322.

Capaldi, E. J., Nawrocki, T. M., & Verry, D. R. (1983). The nature of anticipation: An inter- and intraevent process. *Animal Learning & Behavior, 11,* 193–198.

Capaldi, E. J., Verry, D. R., & Davidson, T. (1980). Memory, serial anticipation pattern learning, and transfer in rats. *Animal Learning & Behavior, 8,* 575–585.

Capaldi, E. J., Verry, D. R., Nawrocki, T. M., & Miller, D. J. (1984). Serial learning, interitem association, phrasing cues, interference, overshadowing, chunking, memory, and extinction. *Animal Learning & Behavior, 12,* 7–20.

Chen, S., Swartz, K. B., & Terrace, H. S. (1997). Knowledge of the ordinal position of list items in rhesus monkeys. *Psychological Science, 8,* 80–86.

Church, R. M. (1997). Quantitative models of animal learning and cognition. *Journal of Experimental Psychology: Animal Behavior Processes, 23,* 379–389.

Couvillon, P. A., & Bitterman, M. E. (1992). A conventional conditioning analysis of "transitive inference" in pigeons. *Journal of Experimental Psychology: Animal Behavior Processes, 18,* 308–310.

Dallal, N. L., & Meck, W. H. (1990). Hierarchical structures: Chunking by food type facilitates spatial memory. *Journal of Experimental Psychology: Animal Behavior Processes, 16,* 69–84.

D'Amato, M., Salmon, D. P., Loukas, E., & Tomie, A. (1985). Symmetry and transitivity of conditional relations in monkeys (*Cebus apella*) and pigeons (*Columba livia*). *Journal of the Experimental Analysis of Behavior, 44,* 35–47.

Davis, H. (1992). Transitive inference in rats (*Rattus norvegicus*). *Journal of Comparative Psychology, 106,* 342–349.

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

Fountain, S. B. (1986). Serial-pattern learning: A unitary process? *Physiological Psychology, 14,* 67–70.

Fountain, S. B. (1990). Rule abstraction, item memory, and chunking in rat serial-pattern tracking. *Journal of Experimental Psychology: Animal Behavior Processes, 16,* 96–105.

Fountain, S. B., & Annau, Z. (1984). Chunking, sorting, and rule-learning from serial patterns of brain-stimulation reward by rats. *Animal Learning & Behavior, 12,* 265–274.

Fountain, S. B., Benson, A. M., & Wallace, D. G. (2000). Number, but not rhythmicity, of temporal cues determines phrasing effects in rat serial-pattern learning. *Learning and Motivation, 31,* 301–322.

Fountain, S. B., Evensen, J. C., & Hulse, S. H. (1983). Formal structure and pattern length in serial pattern learning by rats. *Animal Learning & Behavior, 11,* 186–192.

Fountain, S. B., Henne, D. R., & Hulse, S. H. (1984). Phrasing cues and hierarchical organization in serial pattern learning by rats. *Journal of Experimental Psychology: Animal Behavior Processes, 10,* 30–45.

Fountain, S. B., & Hulse, S. H. (1981). Extrapolation of serial stimulus patterns by rats. *Animal Learning & Behavior, 9,* 381–384.

Fountain, S. B., Raffaele, K. C., & Annau, Z. (1986). Behavioral consequences of intraperitoneal carbon monoxide administration in rats. *Toxicology and Applied Pharmacology, 83,* 546–555.

Fountain, S. B., & Rowan, J. D. (1995a). Coding of hierarchical versus linear pattern structure in rats and humans. *Journal of Experimental Psychology: Animal Behavior Processes, 21,* 187–202.

Fountain, S. B., & Rowan, J. D. (1995b). Sensitivity to violations of "run" and "trill" structures in rat serial-pattern learning. *Journal of Experimental Psychology: Animal Behavior Processes, 21,* 78–81.

Fountain, S. B., Rowan, J. D., & Benson, D. M., Jr. (1999). Rule learning in rats: Serial tracking in interleaved patterns. *Animal Cognition, 2,* 41–54.

Fountain, S. B., Schenk, D. E., & Annau. Z. (1985). Serial-pattern-learning processes dissociated by trimethyltin exposure in rats. *Physiological Psychology, 13,* 53–62.

Fountain, S. B., Wallace, D. G., & Rowan, J. D. (2002). The organization of sequential behavior. In S. B. Fountain, M. Bunsey, J. H. Danks, & M. K. McBeath (Eds.), *Animal cognition and sequential behavior: Behavioral, biological and computational perspectives* (pp. 115–150). Norwell, MA: Kluwer Academic.

Gallistel, C. R. (1995). The replacement of general-purpose theories with adaptive specializations. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 1255–1267). Cambridge, MA: MIT Press.

Gillan, D. J. (1981). Reasoning in the chimpanzee: II. Transitive inference. *Journal of Experimental Psychology: Animal Behavior Processes, 7,* 150–164.

Gluck, M. A., Reifsnider, E. S., & Thompson, R. F. (1990). Adaptive signal processing and the cerebellum: Models of classical conditioning and VOR adaptation. In M. A. Gluck & D. E. Rumelhart (Eds.), *Neuroscience and connectionist theory* (pp. 131–185). Hillsdale, NJ: Erlbaum.

Haggbloom, S. J. (1985). Serial learning and transfer in rats: Effects of changes in stimulus–stimulus associations, pattern structure, and serial position information. *Animal Learning & Behavior, 13,* 370–374.

Haggbloom, S. J., & Brooks, D. M. (1985). Serial anticipation and pattern extrapolation in rats as a function of element discriminability. *Animal Learning & Behavior, 13,* 303–308.

Hockley, W. E., & Murdock, B. B. (1987). A decision model for accuracy and response latency in recognition memory. *Psychological Review, 94,* 341–358.

Hogan, D. E., & Zentall, T. R. (1977). Backward associations in the pigeon. *American Journal of Psychology, 90,* 3–15.

Hull, C. L. (1931). Goal attraction and directing ideas conceived as habit phenomena. *Psychological Review, 38,* 487–506.

Hulse, S. H. (1978). Cognitive structure and serial pattern learning by animals. In S. H. Hulse, H. Fowler, & W. K. Honig (Eds.), *Cognitive processes in animal behavior* (1st ed., pp. 311–340). Hillsdale, NJ: Erlbaum.

Hulse, S. H. (1980). The case of the missing rule: Memory for reward vs. formal structure in serial-pattern learning by rats. *Animal Learning & Behavior, 8,* 689–690.

Hulse, S. H., & Dorsky, N. P. (1977). Structural complexity as a determinant of serial pattern learning. *Learning and Motivation, 8,* 488–506.

Hulse, S. H., & Dorsky, N. P. (1979). Serial pattern learning by rats: Transfer of a formally defined stimulus relationship and the significance of nonreinforcement. *Animal Learning & Behavior, 7,* 211–220.

Hunter, W. S. (1920). The temporal maze and kinaesthetic sensory processes in the white rat. *Psychobiology, 2,* 1–17.

Jensen, A. R., & Rohwer, W. D., Jr. (1965). What is learned in serial learning? *Journal of Verbal Learning and Verbal Behavior, 4,* 62–72.

Jones, M. R. (1974). Cognitive representations of serial patterns. In B. Kantowitz (Ed.), *Human information processing: Tutorials in performance and cognition* (1st ed., pp. 187–228) Hillsdale, NJ: Erlbaum.

Kesner, R. P., Measom, M. O., Forsman, S. L., & Holbrook, T. H. (1984). Serial-position curves in rats: Order memory for episodic spatial events. *Animal Learning & Behavior, 12,* 378–382.

Lashley, K. S. (1951). The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior* (1st ed., pp. 112–146). New York: Wiley.

Lazar, R. (1977). Extending sequence-class membership with matching to sample. *Journal of the Experimental Analysis of Behavior, 27,* 381–392.

Macuda, T., & Roberts, W. A. (1995). Further evidence for hierarchical chunking in rat spatial memory. *Journal of Experimental Psychology: Animal Behavior Processes, 21,* 20–32.

Markovits, H., & Dumas, C. (1992). Can pigeons really make transitive inferences? *Journal of Experimental Psychology: Animal Behavior Processes, 18,* 311–312.

McGonigle, B., & Chalmers, M. (1992). Monkeys are rational! *Quarterly Journal of Experimental Psychology, 45B,* 189–228.

Metcalfe, J. (1990). Composite holographic associative recall model (CHARM) and blended memories in eyewitness testimony. *Journal of Experimental Psychology: General, 119,* 145–160.

Metcalfe, J. (1991). Recognition failure and the composite memory trace in CHARM. *Psychological Review, 98,* 529–553.

Metcalfe, J. (1993). Novelty monitoring, metacognition, and control in a composite holographic associative recall model: Implications for Korsakoff amnesia. *Psychological Review, 100,* 3–22.

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

Murdock, B. B. (1983). A distributed memory model for serial-order information. *Psychological Review, 90,* 316–336.

Murdock, B. B. (1985). Convolution and matrix systems: A reply to Pike. *Psychological Review, 92,* 130–132.

Murdock, B. B. (1992). Serial organization in a distributed memory model. In A. F. Healy, S. M. Kosslyn, & R. M. Shiffrin (Eds.), *From learning theory to connectionist theory: Essays in honor of William K. Estes* (pp. 201–225). Hillsdale, NJ: Erlbaum.

Murdock, B. B. (1993). TODAM2: A model for the storage and retrieval of item, associative, and serial-order information. *Psychological Review, 100,* 183–203.

Murdock, B. B. (1995a). Developing TODAM: Three models for serial-order information. *Memory and Cognition, 23,* 631–645.

Murdock, B. B. (1995b). Similarity in a distributed memory model. *Journal of Mathematical Psychology, 39,* 251–264.

Neath, I., & Capaldi, E. J. (1996). A "random-walk" simulation model of multiple-pattern learning in a radial-arm maze. *Animal Learning & Behavior, 24,* 206–210.

Newell, A., & Simon, H. A. (1961). Computer simulation of human thinking. *Science, 134,* 2011–2017.

Olton, D. S., Shapiro, M. L., & Hulse, S. H. (1984). Working memory and serial patterns. In H. L. Roitblat, T. G. Bever, & H. S. Terrace (Eds.), *Animal cognition* (pp. 171–182). Hillsdale, NJ: Erlbaum.

Press, W. H., Flannery, B. P., Teukolsky, S. A., & Vetterling, W. T. (1988). *Numerical recipes in C: The art of scientific computing.* New York: Cambridge University Press.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning: II. Current research and theory* (pp. 64–99). New York: Appleton-Century-Crofts.

Restle, F., & Brown, E. R. (1970a). Organization of serial pattern learning. In G. H. Bower (Ed.), *Psychology of learning and motivation* (1st ed., pp. 249–331). New York: Academic Press.

Restle, F., & Brown, E. R. (1970b). Serial pattern learning. *Journal of Experimental Psychology, 83,* 120–125.

Restle, F., & Brown, E. R. (1970c). Serial pattern learning: Pretraining of runs and trills. *Psychonomic Science, 19,* 321–322.

Roberts, W. A. (1979). Spatial memory in the rat on a hierarchical maze. *Learning and Motivation, 10,* 117–140.

Roberts, W. A., & Phelps, M. T. (1994). Transitive inference in rats: A test of the spatial coding hypothesis. *Psychological Science, 5,* 368–374.

Roitblat, H. L. (1982). The meaning of representation in animal memory. *Behavioral and Brain Sciences, 5,* 353–406.

Roitblat, H. L. (1984). Representations in pigeon working memory. In H. L. Roitblat, T. G. Bever, & H. S. Terrace (Eds.), *Animal cognition* (pp. 79–97). Hillsdale, NJ: Erlbaum.

Roitblat, H. L., Pologe, B., & Scopatz, R. A. (1983). The representation of items in serial position. *Animal Learning & Behavior, 11,* 489–498.

Roitblat, H. L., & von Fersen, L. (1992). Comparative cognition: Representations and processes in learning and memory. *Annual Review of Psychology, 43,* 671–710.

Rumelhart, D. E., & McClelland, J. L. (1987). Learning the past tense of English verbs: Implicit rules or parallel distributed processing? In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 195–248). Hillsdale, NJ: Erlbaum.

Sands, S. F., & Wright, A. A. (1980). Primate memory: Retention of serial list items by a rhesus monkey. *Science, 209,* 938–940.

Schusterman, R. J., & Kastak, D. (1993). A California sea lion (*Zalophus californianus*) is capable of forming equivalence relations. *Psychological Record, 43,* 823–839.

Sidman, M. (1971). Reading and auditory-visual equivalences. *Journal of Speech & Hearing Research, 14,* 5–13.

Sidman, M., Rauzin, R., Lazar, R., Cunningham, S., Tailby, W., & Carrigan, P. (1982). A search for symmetry in the conditional discriminations of rhesus monkeys, baboons, and children. *Journal of the Experimental Analysis of Behavior, 37,* 23–44.

Sidman, M., & Tailby, W. (1982). Conditional discrimination vs. matching to sample: An expansion of the testing paradigm. *Journal of the Experimental Analysis of Behavior, 37,* 5–22.

Simon, H. A., & Kotovsky, K. (1963). Human acquisition of concepts for sequential patterns. *Psychological Review, 70,* 534–546.

Skinner, B. F. (1934). The extinction of chained reflexes. *Proceedings of the National Academy of Sciences, 20,* 234–237.

Spence, K. W. (1936). The nature of discrimination learning in animals. *Psychological Review, 43,* 427–449.

Spence, K. W. (1937). The differential response in animals to stimuli varying within a single dimension. *Psychological Review, 44,* 430–444.

Steirn, J. N., Weaver, J. E., & Zentall, T. R. (1995). Transitive inference in pigeons: Simplified procedures and a test of value transfer theory. *Animal Learning & Behavior, 23,* 76–82.

Stempowski, N. K., Carman, H. M., & Fountain, S. B. (1999). Temporal phrasing and overshadowing in rat serial-pattern learning. *Learning and Motivation, 30,* 74–100.

Swartz, K. B., Chen, S., & Terrace, H. S. (1991). Serial learning by rhesus monkeys: I. Acquisition and retention of multiple four-item lists. *Journal of Experimental Psychology: Animal Behavior Processes, 17,* 396–410.

Terrace, H. S. (1987). Chunking by a pigeon in a serial learning task. *Nature, 325,* 149–151.

Terrace, H. S. (1991). Chunking during serial learning by a pigeon: I. Basic evidence. *Journal of Experimental Psychology: Animal Behavior Processes, 17,* 81–93.

Terrace, H. S. (2002). The comparative psychology of chunking. In S. B. Fountain, M. Bunsey, J. H. Danks, & M. K. McBeath (Eds.), *Animal cognition and sequential behavior: Behavioral, biological, and computational perspectives* (pp. 23–55). Norwell, MA: Kluwer Academic.

Terrace, H. S., & Chen, S. (1991a). Chunking during serial learning by a pigeon: II. Integrity of a chunk on a new list. *Journal of Experimental Psychology: Animal Behavior Processes, 17,* 94–106.

Terrace, H. S., & Chen, S. (1991b). Chunking during serial learning by a pigeon: III. What are the necessary conditions for establishing a chunk? *Journal of Experimental Psychology: Animal Behavior Processes, 17,* 107–118.

Tomonaga, M., Matsuzawa, T., Fujita, K., & Yamamoto, J. (1991). Emergence of symmetry in a visual conditional discrimination by chimpanzees (*Pan troglodytes*). *Psychological Reports, 68,* 51–60.

Treichler, F. R., & Van Tilburg, D. (1996). Concurrent conditional discrimination tests of transitive inference by macaque monkeys: List linking. *Journal of Experimental Psychology: Animal Behavior Processes, 22,* 105–117.

Wallace, D. G., Lewis, C. F., Fountain, S. B., & Block, C. H. (1999). Rat "rule induction" can be modeled by a neural network with supervised learning. *Society for Neuroscience Abstracts, 25,* 85.

Weaver, J. E., Steirn, J. N., & Zentall, T. R. (1997). Transitive inference in pigeons: Control for differential value transfer. *Psychonomic Bulletin & Review, 4,* 113–117.

Wynne, C. D. L. (1997). Pigeon transitive inference: Tests of simple accounts of a complex performance. *Behavioural Processes, 39,* 95–112.

Zentall, T. R., & Sherburne, L. M. (1994). Transfer of value from S+ to S− in a simultaneous discrimination. *Journal of Experimental Psychology: Animal Behavior Processes, 20,* 176–183.

Zentall, T. R., & Urcuioli, P. J. (1993). Emergent relations in the formation of stimulus classes by pigeons. *Psychological Record, 43,* 795–810.