ASSOCIATIVE LEARNING VERSUS RULE-LEARNING: A COMPUTER MODEL OF PATTERN PHRASING EFFECTS

A thesis submitted to the Kent State University Honors College in partial fulfillment of the requirements for University Honors

by

Alison Dickerhoof

May, 2012
TABLE OF CONTENTS

LIST OF FIGURES ................................................................................. iv
LIST OF TABLES ....................................................................................... v
ACKNOWLEDGMENTS ............................................................................. vi

CHAPTER

I. INTRODUCTION .................................................................................. 1
   Rule Learning versus Associative Learning in Rat Serial Pattern Learning ......................................................... 4
   The Sequential Pairwise Associative Memory (SPAM) Computational Model ...................................................... 8

II. METHOD ................................................................................................. 19

III. RESULTS ............................................................................................. 21

IV. DISCUSSION ......................................................................................... 26

V. REFERENCES ......................................................................................... 30
LIST OF FIGURES

Figure 1  Flow diagram for one iteration of the 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. Note. From “What is learned in sequential learning? An associative model of reward magnitude serial-pattern learning,” by D. G. Wallace and S. B. Fountain, 2002, Journal of Experimental Psychology: Animal Behavior Processes, 28, 43-63. Copyright 2002 by the American Psychological Association. Reprinted with permission.

Figure 2  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. Note. From “What is learned in sequential learning? An associative model of reward magnitude serial-pattern learning,” by D. G. Wallace and S. B. Fountain, 2002, Journal of Experimental Psychology: Animal Behavior Processes, 28, 43-63. Copyright 2002 by the American Psychological Association. Reprinted with permission.

Figure 3  The WM here represents the average choice after ten trials that was predicted with the cue of 1 pellet. (Further description of how the WM is calculated is provided in Fig. 2.) Although each of the groups, when using a cue of 1, predicted an average number close to 3, the groups which included cues produced significantly lower WMs than did the no cue group. * indicates $p < .05$.

Figure 4  SPAM produced results consistent with the idea that the pattern which did not include cues would prove more difficult to learn than any of the other three conditions which did use cues. The no cues condition produced fewer of the lower-item choices and more of the higher-item choices than did the three cued conditions, which supports the idea that the no cue condition, with a cue item of 1, did more poorly in predicting 0 pellets than did any of the other three cued conditions.
LIST OF TABLES

Table 1  Programming of the Patterns with and without Phrasing Cues used in Fountain, Henne, and Hulse (1984) into SPAM .......................... 20

Table 2  T-test of Significance Between Groups ........................................... 24
ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor Dr. Stephen B. Fountain. His high level of expertise in the subject area and experience with this type of work were of immeasurable benefit to my project. I would also like to thank him for his patience and encouragement, which I could not have done without. Additionally, I would like to express my gratitude towards my thesis committee members, Dr. Sara Newman, Dr. Marilyn Norconk and Dr. David Riccio for their time and insight.
Introduction

One classic question amongst researchers using rats for pattern learning experiments is whether or not the rats can learn abstract rules. Rats are able to learn patterns, but what mechanisms are they using to learn them? There is a fundamental debate about how rats learn and just how similar their brains, or, in particular, their memory encoding processes, might be to those of humans. For animals in general, it may not be a far stretch of the imagination to think that a primate has a similar memory encoding process to a human, but what about the seemingly simple-minded rat? There are two main views into the processes of a rat's memory; there are those who believe rat brains function quite like human brains and that rats are able to learn rules in order to retain memory about a pattern, and there are those who believe rat brains work like simple computers, memorizing the items in a pattern simply by associating them to one another. For example, someone who believes rats have the capability of creating rules and using them to memorize patterns might think the rat had been able to remember a pattern of 1-3-5-7-9 of food pellets because the rat had created a “+2” rule. So, when the rat experiences a number of pellets in the pattern, such as “3”, it would apply the “+2” rule and correctly anticipate “5” as the next item in the pattern. On the other hand, someone who believed in the simple association theory of rat memory might believe that the same 1-3-5-7-9 pattern would be memorized easily because the rat would learn to associate 1 pellet with 3, 3 with 5, 5 with 7, and 7 with 9. Therefore, when the rat saw any number of pellets in this pattern, it would be able to predict the next number by using the
associations it had stored in memory during training.

   Notably, experiments involving reward magnitude serial pattern learning have been used to support these two theories of how rats learn. Serial pattern learning involves training an experimental participant to learn a certain pattern. The reward magnitudes used in these studies were arranged to form patterns. Specifically, a particular reward size (the reward being food pellets) was chosen to represent each item within a determined serial pattern. The correct number of food pellets was placed at the end of a straight runway before each trial began. The animal was placed in a start box, and was then released to travel down the straight runway in search of the food pellets at the other end. On each trial, the number of food pellets would change based on the reward magnitude dictated by the serial pattern. The animal then learned these patterns of reward sizes, and demonstrated this learning in terms of how quickly or slowly it ran towards the end of the runway. (A rat who predicted a large reward ran faster than if it predicted a small reward).

   When trying to determine if a participant (whether it be a human or an animal) is using rules, typically the participants are split into groups, one trained with structured patterns and another trained with unstructured patterns. The structured patterns are formally simple; that is, they can be described by simple rules. For example, the serial pattern 10-9-8-7-6-5-4-3-2-1 is very simple and can be remembered by a “less than” rule that describes the relationships between the successive items of the list. A simple decreasing (or increasing) pattern can be described as monotonic, since the values within the pattern decrease monotonically. An unstructured, formally complex pattern would not be able to be described by any simple rule, and would generally be represented by a
random order of numbers within a serial pattern. Taking the same numbers used in the structured pattern example, an example of an unstructured pattern would be 10-3-7-5-9-2-4-6-8-1. A complex pattern can also be described as nonmonotonic. The research of Hulse and Dorsky (1977), Fountain, Evensen and Hulse (1983) and Capaldi and Molina (1979) all used rats learning serial patterns to support or refute these two opposing views of behavioral learning in animals.
Rule Learning versus Associative Learning in Rat Serial Pattern Learning

The first research into whether or not rats can learn rules was conducted in the Hulse and Dorsky (1977) study investigating whether structural complexity was a determinant of how well a pattern could be learned. Previous research done with human subjects supported the idea that humans are better at remembering sequences organized in such a way that a simple rule can describe their patterns (e.g., Jones, 1974; Kotovsky, 1963; Restle & Brown, 1970). Hulse and Dorsky were interested in testing rats to see if they found patterns with simple rules easier to learn than patterns without those simple rules. To demonstrate this difference in complexity, Hulse and Dorsky chose to use monotonic and nonmonotonic serial patterns. Hulse and Dorsky postulated that the same learning differences would be observed in rats, supporting the idea that the rats were using rules to learn the patterns.

The experimental design included two groups of rats, one trained on a monotonically decreasing pattern of food pellets (14-7-3-1-0), and the other group trained with a similar pattern of nonmonotonically decreasing food pellets (14-1-3-7-0). Each day, each rat was subject to four trials, each consisting of five runs (one for each quantity of food pellets). In order to obtain the food, the rat was required to run down to a goal box which was at the end of a straight runway. The running time for each rat on each run was recorded, and was later used as a measure of anticipation. After seventeen days, the experiment stopped, and the results were recorded. The results showed that the rats in the group trained on the monotonic pattern were better able to anticipate zero food pellets.
than the rats which were trained using the nonmonotonic pattern. That is, the rats with the monotonic pattern ran more slowly towards the goal box on the final (zero pellets) trial than did their counterparts in the group trained on the nonmonotonic patterns. Hulse and Dorsky (1977) then concluded that this was strong evidence in favor of the rule learning theory in rats.

When Capaldi and Molina (1979) countered the Hulse and Dorsky (1977) findings with their own research on learning in rats, they approached the question with a different view. Capaldi was of the opinion that rats do not use rules, but rather they use simple associative mechanisms to learn patterns. In order to provide evidence against the idea that monotonically arranged patterns are easier to learn, he set up an experiment using monotonic and nonmonotonic patterns. However, the patterns were designed so that the learning was based more upon how easily one item could be discriminated from the next than on whether the elements in the pattern were arranged in a monotonic fashion. The experiment was designed using the assumption that items easily distinguished from one another would be easier to learn, or that the element discriminability was most important in pattern learning. The research involved two separate experiments, the first of which utilized three different patterns with varying discriminability, with one having the least discriminability and being arranged in a monotonic pattern (15-10-5-0), with the other two containing more easily discriminable elements arranged in weakly monotonic patterns (15-15-0-0 and 14-14-2-0). The second experiment utilized two different, shorter patterns, one with difficult to discriminate elements arranged in a monotonic pattern (20-10-0) and the other with easily discriminable elements arranged in a nonmonotonic
pattern (1-29-0).

The results of these experiments showed that the rats were able to predict zero pellets better when the item preceding zero was very distinctly different from zero (Capaldi & Molina, 1979). For example, in Experiment 2, the nonmonotonic sequence (1-29-0) proved to lead to better and faster learning than did the monotonic (20-10-0) sequence. The rats ran more slowly towards zero pellets when trained with the nonmonotonic 1-29-0 pattern than they did when trained with the monotonic 20-10-0 pattern. It was not a large difference, but the fact that learning of the nonmonotonic sequence did not produce poorer results (and in fact, produced better results) than learning of the monotonic sequence was enough for Capaldi to conclude that monotonic sequences (which facilitate rule learning) are not necessarily easier for rats to learn when the elements are not easy to discriminate from one another. Thus, his research supported his theory that rats use simple associative mechanisms to learn patterns.

Later, Fountain, Evensen, and Hulse (1983) conducted another study that also supported the idea that rats learned rules because they were able to learn long monotonic patterns better than long nonmonotonic patterns of food quantities. The experiment was set up so that there were four groups of rats, each under different serial pattern conditions. One group was trained to learn the long monotonic (simple) pattern of 18-10-6-3-0 pellets. Another group was trained to learn the short monotonic (simple) pattern of 18-1-0 pellets. The third group was trained to learn the long nonmonotonic (complex) pattern of 10-1-3-6-18-0 pellets. The final group was trained to learn the short nonmonotonic (complex) pattern of 1-18-0 pellets. The results showed that the group trained with the
long monotonic pattern reached asymptote earlier and was able to predict zero pellets better than any other group through the end of the study. The second best group in terms of speed of learning and ability to predict zero pellets was the short monotonic group. This is expected because the lack of repetition of the “less than” rule demonstrated in the short monotonic patterns as compared to the long monotonic patterns, so it is slightly more difficult to learn. The data that Capaldi and Molina (1979) found showing that rats learn short monotonic and nonmonotonic reward magnitude patterns using only simple associations conflicts with both of the Hulse studies mentioned above (Hulse & Dorsky, 1977; Fountain, Evensen, & Hulse, 1983). The fact that the rats in the Capaldi and Molina (1979) study were able to learn both short monotonic and short nonmonotonic patterns equally well suggests that the rats were using only associations to learn the patterns and not rules.

To help clear up the disparity, the Sequential Pairwise Associative Memory (SPAM) computational model was developed by Wallace and Fountain (2002). It was developed in an attempt to replicate the data from the reward magnitude experiments. Since it was programmed to use only simple association mechanisms to produce output, whatever results were obtained from running SPAM could only be the result of associations.
The Sequential Pairwise Associative Memory (SPAM) Computational Model

SPAM works by using associations to predict an outcome with a given cue. When programming SPAM to replicate lab data, first, the original reward quantities from the experiments must be coded as vectors of random numbers. SPAM can create sets of related vectors to represent pattern items, and these vectors all vary systematically so the numbers within them are related in the same way that food quantities in a pattern are related to each other. For example, vectors numbered 120-130 are a set where successive vectors differ by ten percent. It is worth noting that SPAM also has the capability to pair items that are represented by vectors that are truly random and not related to other vectors. For this particular research, it was necessary to use related vector items. In the Fountain, Henne, and Hulse (1984) study later replicated in this research, the reward magnitudes follow a certain pattern. In other words, the number of food pellets that makes up the reward magnitude on any given trial is not random; rather, a specific trial's reward magnitude has been determined by the experiment's other reward magnitudes, creating a stimulus dimension of reward magnitudes of systematically varying food quantities.

When choosing vectors to represent real-world reward magnitude values, one must consider how each number relates to the others in the set and in what way they vary. For example, when SPAM was first used to replicate data with monotonic and nonmonotonic patterns, one pattern used was that of Hulse and Dorsky (1977): 14-7-3-1-0 pellets (Wallace & Fountain, 2002). These quantities of pellets do vary in a systematic
way; therefore, the items chosen from the vector set should also vary in the same systematic way. In the Hulse and Dorsky (1977) experiment, each food quantity was roughly half of the previous food quantity, starting with fourteen and ending with zero. Therefore, the vector items 127, 125, 123, 121 and 120 were used by Wallace and Fountain (2002) since they vary in a similar systematic way as the actual food quantities used by Hulse and Dorsky.

Next, the vector items were associated with one another in SPAM. The vector item 127 was paired with 125, 125 with 123, 123 with 121, and 121 with 120. This represents the idea that each previous food quantity predicts the next food quantity (as shown in part (1) of Figure 1). There is also a random vector (drawn from random vectors numbered 0-100) that is unrelated to food quantity vectors placed before the beginning item to serve as a start cue. This can represent anything that an animal in the lab may use as a cue for the start of the experiment, such as being placed in the chamber. In the lab, these cues often lead to the animal predicting the first value in the pattern, so they are important. In SPAM, associations are formed by a process called convolution that is described in detail by Wallace and Fountain (2002). The results of the aforementioned associations are all added to a common memory trace, which is the composite of all associations (as shown in part (2) of Figure 1). Then, the exponential decay parameters $\alpha, \gamma_1, \gamma_2,$ and $\omega$ can be set to represent forgetting. These parameters were automatically set at $\alpha=1.0, \gamma_1=0.0, \gamma_2=0.0$ and $\omega=1.0$, as they were in Wallace and Fountain (2002). The number of iterations must then be set, determining how many times the program repeats its operations, with one thousand being the default. This is equivalent
of testing one thousand subjects in the experiment. A cue is then selected to probe memory, and a lexicon developed. The items within the lexicon are the possible choices for the cue, usually representing all other items in the original set of related vectors, thus representing all possible food quantities. If multiple cues are used, each cue must have a copy of the lexicon from which to choose. Finally, the number of cue item probes must be selected, and the number of items in the lexicon must be designated. Then the program is run, and output is produced by retrieval through correlation, which is also described in detail in Wallace and Fountain (2002). The output allows the user to see how many times the cue picked each item in a lexicon, and in some cases the Weighted Match (WM) may be used in stimulus dimension tests.
Convolutions Contained in the Sequence: 20-10-0

(1) Pairwise Associations

Start → 8 → 8 → 6 → 6 → 0

Cue Item 6 → T

(2) Composite Memory Trace

Dimension Lexicon

10 → 9 → 8 → 7 → 6 → 5 → 4 → 3 → 2 → 1 → 0

(3) Retrieved Item

(4)

(5)
Figure 1. Flow diagram for one iteration of the 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. Note. From “What is learned in sequential learning? An associative model of reward magnitude serial-pattern learning,” by D. G. Wallace and S. B. Fountain, 2002, *Journal of Experimental Psychology: Animal Behavior Processes, 28*, 43-63. Copyright 2002 by the American Psychological Association. Reprinted with permission.
Figure 2 shows that the WM is calculated by taking the percentage of times each item was chosen by the cue and turning it into a proportion. Then, the proportion is multiplied by the value of the item with which it is associated. For example, if 120 was chosen three percent of the time, then .03 would be multiplied by 0 (because the item 120 represents the value 0), to create a weighted score of 0. This method is repeated with the probability of each item in the vector to create a weighted score for each. Then, all the weighted scores are totaled to create the WM. Therefore, the WM is useful in determining SPAM's overall prediction regarding a certain cue because it represents the average value that was chosen.
Anticipated Item Score Calculation

<table>
<thead>
<tr>
<th>Lexicon Item</th>
<th>Probability of Response</th>
<th>Weighted Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>9</td>
<td>0.003</td>
<td>0.027</td>
</tr>
<tr>
<td>8</td>
<td>0.078</td>
<td>0.624</td>
</tr>
<tr>
<td>7</td>
<td>0.162</td>
<td>1.134</td>
</tr>
<tr>
<td>6</td>
<td>0.497</td>
<td>2.982</td>
</tr>
<tr>
<td>5</td>
<td>0.139</td>
<td>0.695</td>
</tr>
<tr>
<td>4</td>
<td>0.061</td>
<td>0.244</td>
</tr>
<tr>
<td>3</td>
<td>0.039</td>
<td>0.117</td>
</tr>
<tr>
<td>2</td>
<td>0.010</td>
<td>0.020</td>
</tr>
<tr>
<td>1</td>
<td>0.008</td>
<td>0.008</td>
</tr>
<tr>
<td>0</td>
<td>0.003</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Anticipated Food Quantity for the Final Item of the Sequence, 20-10-0: 5.851
Figure 2. 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 1000 iterations of the sequential pairwise associative memory (SPAM) model. Note. From “What is learned in sequential learning? An associative model of reward magnitude serial-pattern learning,” by D. G. Wallace and S. B. Fountain, 2002, *Journal of Experimental Psychology: Animal Behavior Processes, 28*, 43-63. Copyright 2002 by the American Psychological Association. Reprinted with permission.
When there is a large number of associations in the trace, there is a large risk for generalization. Since each item pair is associated both backward and forward in the pattern, generalization occurs when an item within the same vector is used more than once. For example, if a trace only contains an association of 121 and 120, a cue of 121 will always retrieve 120. However, if a trace contains 123 paired with 121 and 121 paired with 120, a cue of 121 will retrieve 120 and 123 because 0 and 3 are similar, and it is paired with both.

Of interest now is whether or not SPAM can model other kinds of animal lab data. It has already been used to replicate serial pattern learning with monotonic and nonmonotonic patterns. The following research aimed to discover if SPAM could demonstrate the same temporal and place cue overshadowing effects that were seen in the Fountain, Henne, and Hulse (1984) study. The results of their study supported the idea that chunking (by using temporal and place cues) helped rats learn serial patterns better and faster than rats who learned the same patterns without cues. All rats were trained to learn the 25-element pattern 14-7-3-1-0-14-7-3-1-0-14-7-3-1-0-14-7-3-1-0-14-7-3-1-0 (a strongly monotonic pattern repeated five times). Some groups of rats were given phrasing cues to chunk the pattern into its five logical pieces. The rats were split into six groups, one group learning the pattern with no cues, another learning the pattern with temporal cues, another group learning the pattern with place cues and three groups learning the pattern with combined place and temporal cues. The food pellets were placed at the end of a T-shaped maze (a straight runway with two perpendicular arms going off in each direction from the end of the runway). The rats being trained with no cues followed the
pattern exactly with no breaks between the logical chunks and with the pellets being placed in the same arm of the T-maze every time. The rats in the place cue group experienced food pellets in a different arm of the maze for every logical chunk within the 25 element pattern. The rats trained with a temporal cue rested for 10-15 minutes between chunks, but always received food pellets in the same arm of the maze. The groups using temporal and place cues together received a 10-15 minute break between chunks and the food pellets were placed in different arms of the maze after each chunk. At the end of phase 1, after 13 days, all five of the cued groups showed better anticipation (slower running times) for the 0 pellet element of each chunk than did the group using no phrasing cues (Fountain, Henne & Hulse 1984).

The results of this study outlined many cue removals and cue changes. By the nature of its design, it is known that SPAM cannot replicate cue removal data because it produces only what results would be demonstrated if the animals were trained to asymptote. Although data on training to asymptote has yet to be produced in the lab, it was the focus of this research. If SPAM shows differences between cued and no cue groups, then that would suggest that rats that were trained with cued and non-cued patterns to the fullest of their abilities would always show differences between the no cue group and the groups who had cues to phrase their patterns. Specifically, the groups who received phrasing cues would always show better learning of the patterns than the group that did not. This supports Capaldi's associative view of animal learning because the cues would lead to easier and less confusing associations for the rats so that they would always be able to learn better than a rat who never had cues and therefore had more difficult
associations to learn. If SPAM was not able to show differences in final learning between
groups of rats with phrasing cues and a group with no phrasing cues, then this would
support Hulse's rule learning view of animal learning. Because both groups would be
doing equally well at learning the patterns in this case, it might suggest that they are
learning some sort of rule about how the pattern works (that it is consistently decreasing
to 0 and then returning to a very large number) instead of simply learning that 14 pellets
predict 7 pellets and so on.
Method

As shown in Table 1, SPAM was programmed with no cues, a temporal cue, a place cue and combined place and temporal cues. The no cue pattern was represented with a random start cue paired with 127, 127 paired with 125, 125 paired with 123, 123 paired with 121, 121 paired with 120, and 120 paired with 127, coding for 0 followed by 14 at the beginning of the next pattern chunk. The other three situations were coded in exactly the same way using the same 120 vector, except in the temporal cue condition, the random number 50 was assigned to represent the temporal cue and was paired with the final 127. In the place cue condition, the random number of 42 was assigned to represent the place cue and was paired with the last 127. In the place and temporal cues condition, both 42 and 50 were paired with the final 127. Convolutions of 0 with 0 were added to allow each situation to contain the same number of item pairs. The program was then run with $\alpha = 1.0$, $\gamma_1 = 0.0$, $\gamma_2 = 0.0$ and $\omega = 1.0$, and iterations set at 1000. Only one cue was tested, and that cue was 121 which represented 1 pellet as a cue for 0 pellets. Each condition was run for 1000 iterations ten separate times, and the weighted match (WM) results from each of the ten runs were obtained.
Table 1.

Programming of the Patterns with and without Phrasing Cues used in Fountain, Henne, and Hulse (1984) into SPAM

<table>
<thead>
<tr>
<th>No Cue</th>
<th>Temporal Cue</th>
<th>Place Cue</th>
<th>Temporal and Place Cues</th>
</tr>
</thead>
<tbody>
<tr>
<td>Start*127</td>
<td>Start*127</td>
<td>Start*127</td>
<td>Start*127</td>
</tr>
<tr>
<td>127*125</td>
<td>127*125</td>
<td>127*125</td>
<td>127*125</td>
</tr>
<tr>
<td>125*123</td>
<td>125*123</td>
<td>125*123</td>
<td>125*123</td>
</tr>
<tr>
<td>123*121</td>
<td>123*121</td>
<td>123*121</td>
<td>123*121</td>
</tr>
<tr>
<td>121*120</td>
<td>121*120</td>
<td>121*120</td>
<td>121*120</td>
</tr>
<tr>
<td>120*127</td>
<td>50*127</td>
<td>42*127</td>
<td>50*127</td>
</tr>
<tr>
<td>0*0</td>
<td>0*0</td>
<td>0*0</td>
<td>42*127</td>
</tr>
<tr>
<td>0*0</td>
<td>0*0</td>
<td>0*0</td>
<td>0*0</td>
</tr>
</tbody>
</table>
Results

When SPAM was programmed with temporal and place cues versus no cues, the runs that included cues predicted zero pellets better than the run which did not include any cues. As shown in Figure 3, the no cue condition yielded an average WM of 2.92; the temporal cue condition yielded an average WM of 2.63; the place cue condition yielded an average WM of 2.64; the place and temporal cue condition yielded an average WM of 2.62. Although the WM for the no cue group could be interpreted as essentially the same as the WMs of the cued groups (since they could all be rounded up to 3), tests for significance did not support this interpretation. Since the no cue group's WMs consistently stayed within the 2.90-2.99 range, and the WMs of all three of the cued groups stayed within the 2.60-2.69 range, the no cue and cued conditions produced statistically different results.
Figure 3. The WM here represents the average choice after ten trials that was predicted with the cue of 1 pellet. (Further description of how the WM is calculated is provided in Fig. 2.) Although each of the groups, when using a cue of 1, predicted an average number close to 3, the groups which included cues produced significantly lower WMs than did the no cue group. * indicates $p < .05$. 
A 2 X 2 analysis of variance (ANOVA) was conducted on the data shown in Figure 3 with the WM of each run of the program considered as resulting from a separate subject and with place and temporal cues as between-subjects factors. The ANOVA revealed significant main effects for place cues, $F(1, 36) = 544.31, p < .001$, and temporal cues, $F(1, 36) = 567.14, p < .001$, and a significant Place Cues X Temporal Cues interaction, $F(3, 36) = 582.63, p < .001$. Table 2 shows the results of $t$-tests conducted as post-hoc comparisons. The $t$-tests between each cued group and the no cue group each yielded a $t$ value that corresponded with $p < .001$ ($t (18) = -10.64, p < .001$, $t (18) = -9.98, p < .001$, $t (18) = -12.57, p < .001$). Table 2 additionally shows that tests for significance between each of the different cued groups yielded $t$ values which corresponded with $p$ values large enough to determine that the cued groups were not significantly different from one another ($t (18) = 0.1, p = .9214$, $t (18) = -0.09, p = .9293$, $t (18) = -0.2, p = .8437$). As shown in Figure 4, when looking at the proportion of times each item in the lexicon was picked when cued with the 121 item (representing one pellet), the data show that the no cue condition chose more large values (item vectors 124, 125, 126) than did the cued conditions, and that the no cue condition chose fewer small values (121, 122) than did the three cued groups.
Table 2.

T-test of Significance Between Groups

<table>
<thead>
<tr>
<th></th>
<th>No Cue and Temporal Cue</th>
<th>No Cue and Place Cue</th>
<th>No Cue and Temporal and Place Cues</th>
<th>Temporal Cue and Place Cue</th>
<th>Temporal Cue and Temporal and Place Cues</th>
<th>Place Cue and Temporal and Place Cues</th>
</tr>
</thead>
<tbody>
<tr>
<td>t value</td>
<td>-10.64</td>
<td>-9.98</td>
<td>-12.57</td>
<td>0.1</td>
<td>-0.09</td>
<td>-0.2</td>
</tr>
<tr>
<td>p value</td>
<td>&lt;.0001</td>
<td>&lt;.0001</td>
<td>&lt;.0001</td>
<td>0.9214</td>
<td>0.9293</td>
<td>0.8437</td>
</tr>
<tr>
<td>Significant?</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
</tbody>
</table>
Figure 4. SPAM produced results consistent with the idea that the pattern which did not include cues would prove more difficult to learn than any of the other three conditions which did use cues. The no cues condition produced fewer of the lower-item choices and more of the higher-item choices than did the three cued conditions, which supports the idea that the no cue condition, with a cue item of 1, did more poorly in predicting 0 pellets than did any of the other three cued conditions.
Discussion

In the Fountain, Henne, and Hulse (1984) study, the pattern used always had one pellet on the trial preceding zero pellets. If the rat was able to learn the pattern well, it would run very slowly after having received one pellet. This is interpreted as the rat being able to anticipate correctly that on the trial following the one in which the rat received one pellet at the end of the runway, the rat would receive zero pellets. In the study, as is commonly found in similar research, rats that learn patterns containing phrasing cues that chunk the pattern into logical pieces are better able to predict zero pellets after a preceding magnitude than are rats that learn the same pattern without any phrasing cues. That is, the rats that were trained with cues run more slowly, if at all, towards the end of the runway on the final trial than rats trained using the same pattern on the same trial but that have not been trained using phrasing cues.

The results produced by SPAM (represented by its WMs), can be interpreted in order to understand whether they match the results of the Fountain, Henne, and Hulse (1984) study. The cue in the current experiment SPAM was one pellet (vector 121), which was associated with zero pellets that followed it (vector 120) and three pellets that preceded it (convolutions 123-121 and 121-120). When using one (vector 121) as a cue, the programmer is asking SPAM what is predicted by one pellet. Therefore, if the results of the computer program match the results of the study, SPAM must provide evidence that one pellet predicts zero pellets better in the trials using phrasing cues than it does in the trials that do not use cues where zero pellets predicts fourteen pellets. This can be
seen by examining the overall WM of each trial. The WM represents the program's weighted average of responses to the cue. The way SPAM is designed, a smaller WM represents a smaller predicted reward magnitude, or number of pellets. The program does not in any way directly represent running times, so these must be inferred by the program's predicted reward magnitude.

In the Fountain, Henne and Hulse (1984) study, it was shown that rats that were anticipating smaller reward magnitudes ran more slowly in the runway on that trial. Therefore, the predicted value, or WM, produced by SPAM may lead indirectly to an interpretation of running time in a rat. When SPAM was independently programmed with a no cue trial, a temporal cue trial, a place cue trial, and a place and temporal cue combination trial, all three of the cued conditions had slightly lower (but still significantly different) WM values than the no cue condition. The small WMs of the temporal, place, and temporal and place cue groups (2.63, 2.64, 2.62, respectively) represent a small cued reward magnitude (of roughly 2.63 pellets), and it can therefore be inferred that the small WMs predict slow running times. The WM associated with the no cue group (2.92), which was larger than any of the WMs of the cued groups, suggests that the pattern including no cues does a poorer job of predicting zero pellets (since it predicts three pellets), and represents a faster run time, than the WMs of the cued groups.

The results of this research with SPAM show that there is a difference between having phrasing cues and no phrasing cues when the learning is final and complete. Therefore, these results support the idea that these cues facilitate pattern learning and therefore lead to a better anticipation of zero, but that is only once learning has reached
its asymptote. The Fountain, Henne, and Hulse (1984) study did not include training rats until they had reached the utmost capacity of their learning, because that was not the goal of that research. Interestingly, no such study has ever involved training rats until their pattern learning has reached asymptote, so it is not clear whether these data would actually match real-world findings. It would be necessary to run a study of this type in order to see how its results would match the results from this study with SPAM.

The findings from this research do, however, support Capaldi's view that rats use only simple associations to learn patterns since SPAM uses only simple associations to make its predictions. These results also provide tentative evidence against Hulse's rule learning theory, because it would go against the law of parsimony to assume that rats were using rules to learn these patterns when SPAM provides evidence that the same kind of learning can be achieved by simple associations alone. A study which trains rats to asymptote with this monotonic pattern with cued and no cue groups needs to be implemented in order to understand truly how these results match real animal research lab data. If future research was able to show that animals in no cue and cued groups were equal in their pattern learning, Hulse's rule learning theory would be supported and it would provide evidence against SPAM and Capaldi's simple associations theory. Also, the cue removal research, which was the focus of the Fountain, Henne, and Hulse (1984) study could not be replicated by SPAM because of its power only to show results at the end of an experiment or when the learning has reached its asymptote. Using SPAM, it would be impossible to program cue removal scenarios, because the effects on learning of cue removals are seen directly after the cue is removed. SPAM would be unable to
demonstrate results of any kind of in-progress manipulations.

The SPAM computational model is capable of demonstrating the difference in learning that occurs when cues are placed within a pattern to form logical chunks. Since SPAM is programmed using only simple association mechanisms, it may be interesting to further investigate other kinds of data that may be represented by this program. For example, if data that was previously believed to be explained by rule learning in animals was able to be replicated by SPAM, it may open up new ideas about the differences and similarities between simple associative learning and rule learning.

It is important to note that associative learning of a pattern does not necessarily indicate simple learning. Even if rats are not using rule learning to memorize patterns, having many associations in memory is not simple either. Rather, the associative memory model is also a fairly complex model. However, it is known that humans can derive rules from patterns in order to learn them, and if it is the case that rats use associative mechanisms to memorize patterns, then it can be concluded that humans and rats use different mechanisms for pattern learning. It must also be considered that the tasks discussed in this research have not tapped into reasoning in rats, and that other serial learning tasks, such as those using other serial pattern learning paradigms (e.g., Fountain, 2008; Fountain & Benson, 2006; Fountain & Rowan, 1995a, b; Fountain, Rowan, & Benson, 1999; Fountain et al., 2006, 2012; Kundey & Fountain, 2011), may better demonstrate whether rats can learn rules. Finally, the results of using SPAM do support the idea that rats use associations to memorize patterns, but this does not necessarily mean that past research has tested all possible situations that may entice rats to use rules.
for learning, and the question of whether or not rats can learn rules remains unanswered.
References


