An Analysis of Error Rate as a Predictor of User Persistence in Machine Systems

Ronald Mellado Miller, Rustin S. Terry, Evan B. Johnson, Don Colton

Brigham Young University - Hawaii
55-220 Kulanui Street #1970, Laie, Hi 96762
millerr@byuh.edu

Abstract

Error rate, the proportion of failures to overall attempts, appears to be an almost ubiquitous measure of user outcomes in assessing the acquisition of hardware and software proficiency. This paper investigates, based on Capaldi's (1994) memory-sequential model, the possibility that an overall description of error rate as a percent is less informative than an analysis of the sequences of successes and failures making up the overall percent rate in terms of predicting user persistence. Three experiments are outlined which investigate the effect of the pattern of user error rates in training (acquisition) on user persistence in the face of constant failure (extinction). The experiments involve a simple computer based task, the use of a common word processing software program, and a novel learning task, respectively. The results are generally supportive of sequential theory predictions. Though further investigations are needed, it appears that sequential theory can aid in designing the training for systems whose goals include controlling for user persistence.

1 Theoretical Background

Sequential theory's memory model of learning has been successfully applied in response contingent instrumental conditioning experiments (Capaldi, 1966, 1994; Capaldi and Miller, 2004) as well as in non-response contingent Pavlovian conditioning (Miller and Capaldi, submitted, a). For the following discussion, reinforcement will be considered comparable to a user's success in a task while nonreinforcement will be utilized as being similar to failure in a task. Some of the variables examined by Sequential theory are N-length (the number of nonreinforced trials preceding a reinforced trial), NR transitions (when a nonreinforced trial is followed by a reinforced one), and the partial reinforcement effect (where a group given reinforcement on all trials is generally less resistant to extinction than a group given reinforcement on fewer than 100% of all trials). For this paper, resistance to extinction can be seen as analogous to the number of repeated attempts at success in the face of failure.

While Sequential theory has been applied in a great many learning situations in a variety of species (Capaldi, 1994), this paper deals specifically with both the theoretical and practical applications of this theory to training situations and human error analysis. Evidence in support of Sequential theory (Capaldi, 1966, 1994; Miller, in preparation, b) suggests that, with limited training, a group with more NR transitions (NRNRNR, where N=failure and R=success) or which has received success on every trial (RRRRRR) will be more resistant to extinction than a long N-length group (NNNNNR). With greater levels of training, however, this contingency reverses, with the Long N-length group producing the most resistance to extinction followed by the NRNRNR group, with the group receiving success on every trial actually producing the least resistance to extinction.

In other words, with low levels of training (novice performance), a worker is more likely to make more repeated attempts in the face of failure at a task which has produced a high rate of success (RRRRRR) in the past than at a task which has produced a low rate of success (NRNRNR or NNNNNR). However, with higher levels of training (expert performance), a worker is more likely to make more repeated attempts in the face of failure at a task which has produced a low level of success in the past than one which has produced a. high level of success. This paradoxical result occurs because, according to sequential theory, later in training, a worker trained to the contingency RRRRRR, has learned that success leads to success. Thus, in the face of failure, the worker stops responding. However, later in training, a worker trained to the contingency NNNNNR has learned that, though repeated failures may occur, enough attempts will result in success. Thus, when repeated attempts at success are desired in the face of failure, it may be best to teach a novice a task in such a way that successes occur most often, but as training progresses, it may be best to train the worker in a way that they learn that repeated failures lead to eventual success.

2 Experiment 1

This experiment was designed to simulate sequential effects utilizing a simple computer task. While these effects have been found in animals (Capaldi, 1994) utilizing a runway apparatus with food reinforcement, this experiment sought to find similar effects using a pattern recognition task. The task will consist of responding to a series of either circles or squares which will, one at a time, appear on the screen. The use of N-length in the animal studies is replaced here with the concept of O-length or Outcome Length, which refers to the number of trials of one outcome preceding a trial of a different outcome. For example, an O-length of 3 could be seen as three circles followed by a square. The dependent variable was the number of errors made in responding in the extinction phase (where, for example, if circle was trained in acquisition, then only squares would be presented). Four types of O-lengths will be utilized here: CO (consistent outcome), which refers to only a single type of outcome occurring in acquisition, and SO (short O-length), MO (medium O-length), and LO (long O-length), which refer to a maximum of one, two and four outcome lengths, respectively. It should be noted that, for this experiment, while the schedules are different, the SO, MO, and LO schedules are each 50% schedules, meaning that the various stimuli are each presented on half the given trials. The independent variables to be manipulated are the acquisition schedules themselves (CO, SO, MO, and LO) and the amount of training to be used (2, 4, 10, and 30 training trials). Though the CO and SO schedules will be trained with 2, 4, 10, and 30 acquisition training trial, the MO and LO schedules (because each requires at least 10 training trials) will only be tested with 10 and 30 acquisition trials. This results in two distinct factorial structures for this experiment. The first factorial structure is also a 2 X 4 design with 2 levels of Schedule (CO and SO) combined factorially with 4 levels of Acquisition Trials (2, 4, 10, and 30). The second is a 2 X 4 factorial design with two levels of Acquisition Trials (10 and 30) combined factorially with 4 levels of Schedule (CO, SO, MO, and LO.

The following findings, if obtained, would offer support to sequential theory as applied to humans. It is assumed that the partial reinforcement effect of greater resistance to extinction in SO, MO, and LO schedules than in CO schedules will be found under the increased training regimen. It is also assumed that O-length effects will be found (progressively greater resistance from SO to LO event schedules) in human subjects with increased training. Training level effects, with less resistance to extinction as training level increases for the CO schedule and greater resistance to extinction as training level increases for the SO, MO, and LO schedules, are also anticipated.

2.1 Method

2.1.1 Participants

The participants were freshman college students taking Introductory Psychology at a mid-western university. The research was a required part of their grade. Each subject was randomly assigned to an experimental group and was used only once. There were a total of 8 groups of 12 subjects each for a total of 96 subjects.

2.1.2 Apparatus

An IBM computer with an 80286 processor and VGA monitor.

2.1.3 Experimental Training

The experimental program was designed to study the effects of event schedule, training level.

Participants were brought into the experimental room, which contained a desk with the computer on it and two chairs, one for the participant in front of the computer monitor and one for the experimenter on the side.

Each trial began with a white triangle appearing in the center of the left side of the otherwise black screen for 2 seconds. The subjects were instructed to press either the key marked "Circle" or the key marked "Square" on the computer keyboard while the white triangle was still on the screen to indicate their guess as to which they predicted would appear following the white triangle. Their choices were recorded only while the white triangle was present on the screen. When the white triangle disappeared, with no delay either a circle or a square appeared on the center of the right side of the screen for 2 seconds. When the circle or square disappeared, the participant was shown a completely black screen for 3 seconds during the inter-trial interval (ITI) until the next trial started with the appearance of the white triangle again.

The instructions were initially given by being displayed on the computer monitor and were subsequently verbally repeated by the experimenter in the following manner:

"You will be asked to guess what shape will follow the appearance of a white triangle. The choices—Circle or Square—are labelled on your keyboard.

Here is an example of the white triangle followed by a circle: (example shown)
Here is an example of the white triangle followed by a square: (example shown)
During the experiment please guess as quickly as possible when the white triangle appears. Please try to guess as accurately as possible.
Thank you for your participation!
Press the spacebar to continue"

To test event schedule effects, 4 patterns were used: CO, SO, MO, and LO. The patterns for each are shown below (the hyphen indicates the separation between acquisition and extinction phases):

CO Pattern: 1111111111-22222222222222222222
SO Pattern: 1212121212-22222222222222222222
MO Pattern: 1221211221-22222222222222222222
LO Pattern: 1121222211-22222222222222222222

where if 1 equals circle, 2 equals square and vice versa. Each group was counterbalanced for shape. For example, in the CO schedule, half the participants saw the circle in acquisition training and the square in extinction training, while the other half saw the square in acquisition and the circle in extinction. In each event pattern, the first stimulus presented in acquisition was the one selected to be extinguished in the extinction phase. Thus, if circle was the correct stimulus for the first acquisition trial, only squares would be presented in the extinction phase and vice versa. Resistance to extinction was measured by the number of response errors (guessing, for example, circle when square was correct) in extinction.

To investigate the effect of training level, each event pattern (CO, SO, MO, and LO) was tested with 10 and 30 acquisition trials. To achieve 30 acquisition trials, the CO and SO patterns were simply extended out to 30 trials while the MO and LO patterns were presented as shown for the first 10 trials, in reverse order for the next 10 trials, and then in the original order for the last 10 trials, for a total of 30 trials. The CO and SO patterns were additionally tested with 2 and 4 acquisition trials each. Regardless of training level, all groups received 20 extinction trials.

2.2 Results

The results of this experiment for the CO and SO schedules can be seen in Figure 1.

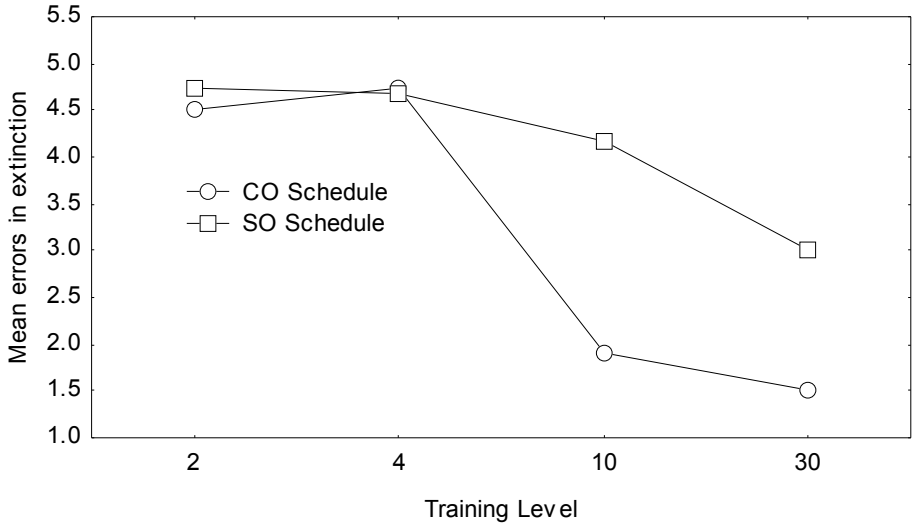

Figure 1. Mean errors in extinction for the CO and SO groups at each training level

A 2 (Schedule: CO, SO) X 4 (Training Level: 2, 4, 10, 30) ANOVA was performed on the data, where the number of errors made in extinction was the dependent variable. Significant effects included Schedule, $F(1,88) = 10.24$, $p < .01$, Training Level, $F(3,88) = 15.67$, $p < .01$, and the Schedule X Training Level interaction, $F(3,88) = 3.16$, $p < .05$. Newman Keuls post hoc tests on the significant interaction indicated that the 2 and 4 levels of training for CO and SO schedules did not differ in the number of errors produced, but that the 2 and 4 levels produced significantly more errors than did the 10 or 30 levels of training. Additionally, in the 10 and 30 levels of training, the SO schedule produced significantly more errors than did the CO schedule.

A 2 (Training Level: 10, 30) X 4 (Schedule: CO, SO, MO, LO) ANOVA was performed. The only significant effect was that of Schedule, $F(3,88) = 27.96$, $p < .01$. Both the Training Level and the Schedule X Training Level interaction were nonsignificant with $F(1,88) = 0.4$ and $F(3,88) = 0.79$, respectively. Significant Newman Keuls post hoc tests on Training Level indicated that the ordering of the schedules from the most to the least resistant in extinction was LO, MO, SO, CO.

2.3 Discussion

Overall, many of the hypotheses appear to be supported in this experiment. Though the CO and SO schedules did not differ as anticipated when trained with few trials (2 and 4), the SO schedule did produce greater errors in extinction than did the CO schedule with more (10 to 30) training trials. Similarly, though changing training levels from 10 to 30 trials did not increase the number of errors in extinction, the schedule effects postulated did occur, with greater O-lengths producing more errors in extinction.

3 Experiment 2

This experiment was designed to examine possible sequential effects in a less controlled, though more realistic user setting than Experiment 1. Here, participants were asked to perform a variety of tasks using Microsoft Word. The tasks consisted of either easy or medium difficulty tasks. In terms of sequential theory, the easy tasks were considered to be analogous to a continuously reinforced group, where all trial attempts to achieve a goal end in reinforcement or success, while the medium tasks were considered to be more closely related to partially reinforced schedules (where some but not all trial attempts end in success). To further the analogy, participants were either interrupted or not during their trials. For both groups, this allowed a comparison between short levels of training (interrupted) and longer levels of training (not interrupted). Extinction here consisted of subjects being asked to perform an impossible to complete task, with the number of tries they attempted before giving up being recorded. The sequential hypotheses being addressed are that the Easy Task group should make more attempts when interrupted (lower training), but should make fewer attempts when uninterrupted (greater training) as the greater training should make the memory of previous success more salient, and so the participant should cease sooner. The Medium Task group, however, should make few attempts when interrupted as only a brief pattern of failure leading to success has been encountered, while with greater training, the memory of failures leading to successes has been encountered more often and has become more salient.

3.1 Method

3.1.1 Participants

Participants were 20 undergraduates from a western university who were recruited from introductory level psychology classes and given extra credit for their participation.

3.1.2 Apparatus

The current study used Microsoft Office Word 2003 on a Dell Latitude X300 laptop computer equipped with a mouse. The laptop computer also had the “Mouse counter” program enabled and hidden from the view of the participants to monitor the number of clicks made by each participant. Microsoft Word 2003 was changed to only show the menu bar (i.e. no tool bars or shortcuts) as a list of options for the program. The program was also set so that the menu bars would not “Show Full Menus after a Short Delay” (This is done by going to “Tools,” then “Customize,” then to “Options,” and deselecting “Show Full Menus after a Short Delay”).

3.1.3 Procedure

Participants were taken into a room with the examiner who sat them down in front of the laptop with the aforementioned changes already made to the setup of Microsoft Word. Students were told that this was a study looking at college students' familiarity with Microsoft Office Word 2003. Each participant was given two tasks, one at a time.

The first task (depending on their learning schedule) was classified as either "Easy Uninterrupted," "Easy Interrupted," "Medium Uninterrupted," and "Medium Interrupted." If the participant was classified as "Uninterrupted" then they were given all the time necessary for them to complete the task. Those given the "Easy" tasks were given the same instructions from the examiner which went as follows: "Click on 'File' once to open the menu and then click on 'File' again to close the menu. Continue this procedure with each of the menu bar items starting with 'File', 'Edit', 'View' and so forth. If the participant was to be interrupted the examiner would interrupt when the participant closed the "format" menu bar and then said "I'm going to have to stop you here." Those given the "Medium" tasks were given the same instructions regardless of whether they were interrupted or uninterrupted which was as follows: "Without using the help functions, change the size of the document from letter size to legal size." For those participants that were classified as "Interrupted", the examiner stopped them after the participants eventually opened "Page Setup" and "Paper" (which were the correct steps to begin to solve this task) and would say "I am going to have to stop you here."

After the participants finished with one of these tasks the examiner took the computer from them and enabled the mouse counter program then gave the computer back to the participant for their final task. All participants were given an unlimited amount of time to finish the final task. The instructions were the following: "Without using the WordArt function or help menu, align the text upside-down. You can use the menu bar. After you either accomplish the task or decide to give up, please just let me know."

Participants were given no other instructions other than those listed above. When the participants finally gave up and left the room the examiner would note the number of mouse clicks done for that task. (2 clicks were taken off of the number listed by the mouse counter program due to the examiners use of the mouse before and after the task.)

3.2 Results

The results for this experiment can be seen in Figure 2.

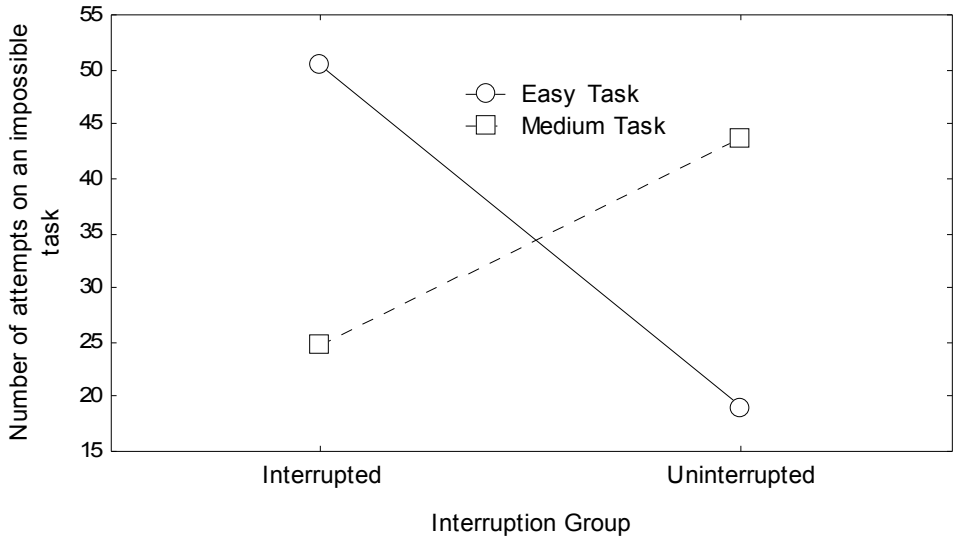

Figure 2. Interrupted and Uninterrupted Group performance for Easy and Medium tasks.

A 2 (Task Difficulty: Easy, Hard) X 2 (Interruption Group: Interrupted, Uninterrupted) ANOVA was performed on the number attempts to perform the impossible task for this experiment. Though the Task Difficulty, $F(1,16)=0.00$, $p>.05$, and Interruption Group, $F(1,16)=0.38$, $p>.05$, main effects were not significant, the Task Difficulty X Interruption Group interaction was significant, $F(1,16)=6.52$, $p<.05$. Though the Newman Keuls post hoc tests showed marginal significance for the differences, the mean differences for the significant Task Difficulty X Interruption Group interaction, as shown in Figure 2, were apparently due to the Easy Task participants making more attempts on the impossible task when interrupted (lower training) than when uninterrupted (greater training), with the Medium Task group making fewer attempts on the impossible task when interrupted (lower training) than when uninterrupted (greater training).

3.3 Discussion

The results are supportive of the effects which sequential theory suggests should be found. The easier task made more attempts on an impossible task when given less training and fewer attempts when given more training. The Medium Task group performed in the opposite manner, with fewer attempts on an impossible task when training was low and more attempts when training level was higher.

4 Experiment 3

This experiment was performed to determine if sequential effects would influence a simple cognitive learning task. While in Experiment 1, participants were able to see if their answers were correct or incorrect and in Experiment 2, participants were able (in acquisition) to create and experience success and failure, here we are testing to see if simply telling a participant that they are correct or incorrect will influence their persistence in a learning/teaching task. In order to properly test the sequential effects, we compared a group consistently told that they were correct with two other groups who were given different patterns of "successful" and "unsuccessful" trials, though in the latter two groups, we held the percent of correct and incorrect answered to 50% for each. As many software programs in use are implemented to teach or test a participant, it was thought that this test would serve as a means of identifying if sequential effects could be found in this or similar teaching/evaluation situations.

4.1 Method

4.1.1 Participants

Participants were 40 undergraduates from a western university who were recruited from introductory level psychology classes and given extra credit for their participation.

4.1.2 Apparatus

The study was hosted on a remote website which students were allowed to access in order to participate.

4.1.3 Procedure

This study was conducted online and was based largely on Martin L. Mehr's and Richard Videbeck's (1968) persistence study.

As in Mehr's et al., (1968) work, the instructions to the participants were as follows:

"The task we are going to ask you to do is part of an ongoing research project in psycholinguistics. Specifically, we are interested in determining whether or not a person who has no knowledge of a language can make better than chance guesses regarding the meaning of certain words in that language. In other words, the possibility exists that all languages have a common core, and that having learned one language you can make some good guesses about the meanings of words in a language which is from all appearances quite strange and foreign to you.

"We will list a series of words from the language spoken by the Mbundu, a primitive tribe in Central Africa. The words have been transcribed in such a way that they are readily pronounceable by someone who knows only English. Consider each word carefully. If you wish, attempt to pronounce it. Then take note of the English word which follows. If in your judgment the English word is at least a rough equivalent of the Mbundu word, click 'same'. If, in your judgment, the English word is not a reasonable equivalent then click 'different'. For example: Does this

Mbundu word and this English word have the same meaning? **mef-det club**.

"If you cannot decide whether the English word is or is not a reasonable equivalent, simply guess. After you click your answer we will tell you if you were correct or not.

"We want to emphasize that this is just a pilot study and we are primarily interested in perfecting research methods. Therefore, after you have made a number of judgments you may quit any time you wish. If you are ready then click 'start' to begin."

After each answer, participants were informed if they had answered "correctly" or "incorrectly" and were updated on the percent of questions they had correctly answered. Correct and incorrect responses from the online program were not made to be contingent upon actual answers of subjects. Rather it was predetermined that subjects would receive different patterns of "correct" and "incorrect" responses. The patterns for the short training (8 trials) condition were continuous reinforcement (CRF) - RRRRRRRR, N-length 1 - NRNRNRNR, and N-length 4 - NNNNRRRR, where R represents where the participants were told they were "correct" and N represent when the participants were told they were "incorrect". It should be noted that both the N-length 1 and the N-length 4 groups are each 50% schedules, meaning that in each of the schedules, the participants were informed that they were incorrect half the time and correct the other half. For the long training (40 trials) condition, the series were simply repeated five times. If participants tried to quit before they completed the required trials they were told, "Please answer more questions before you quit." Each pattern was followed by extinction consisting of only "incorrect" responses until the participant decided to quit or until the participant made 100 attempts in extinction. After the first question during extinction was answered, participants were told, "You have now completed your required trials - you may quit at any time."

In order to ensure that participants put forth their best effort, it was emphasized that the percent of questions they answered correctly would be the percent of extra credit they earned. However, because the quiz was rigged all participants were actually given the full amount of extra credit.

4.2 Results

A 2 (Training Level: 8, 40) X 3 (Schedule: CRF, N-length 1, N-length 4) ANOVA was performed. Significant main effects for Training Level, $F(1,85)=7.2$, $p<.01$, and Schedule, $F(2,85)=7.17$, $p<.01$, were found. The interaction was nonsignificant, $F(2,85)=0.76$, $p>.05$. Significant Newman Keuls post hoc tests indicated that greater levels of training led to fewer errors in all groups. Newman Keuls post hoc tests run on the significant Schedule effect indicated that the N-length 4 group produced significantly greater responding in extinction than was produced by the N length-1 or the CRF groups. While the CRF group produced the fewest responses of any group in extinction, it was not significantly lower than the N length-1 group. The results can be seen in Figure 3.

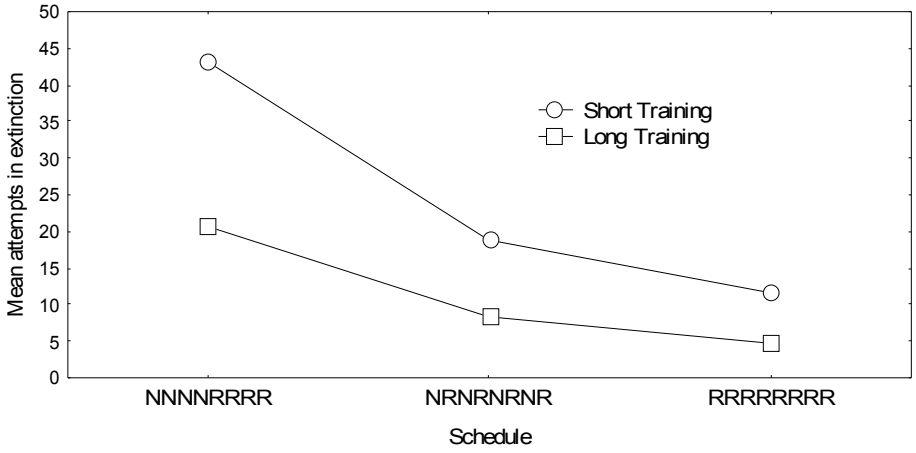

Figure 3. Mean attempts in extinction for each schedule and level of training.

4.3 Discussion

Here the greater training condition again led to less responding in extinction. The N-length findings overall, though, did support sequential theory in that the ordering of the schedules from the most to the least resistant to

extinction was N length-4, N-length-1, and CRF. The results do support the idea that simply informing a participant that they are correct or incorrect in a specific sequential pattern can influence their behaviour, even though the overall "error" rates may be the same.

5 Conclusion

The results of each experiment are suggestive that Sequential effects are present in, and affect training involving, human participants. In each experiment, the specific schedules which were trained in acquisition influenced the resulting resistance to extinction. In experiments 1 and 3, these schedule effects occurred even though each of the partial schedules was a precisely 50% schedule, indicating that overall percentage of error was less useful than knowledge of the sequence trained in predicting user persistence in the task. However, the training effects were less clear. In Experiment 1, more training did lead to greater differences between the CO and SO groups, but increasing training for the MO and LO groups did not. The results of Experiment 2 were in accordance with sequential theory's predictions, though in Experiment 3, training served only to reduce the overall number of errors. More investigations along these lines would be useful, especially in terms of novice versus expert performance. Overall, though, while our training level data may not be as clear as desired, it appears useful for training contingencies to incorporate sequential effects in order to better maximize or minimize worker efforts in the face of possible failure. Similarly, in analyses of error rates in the acquisition of a task, strict percentages of errors in acquisition may be less important than the analysis of actual sequences of success and failures.

References

Capaldi, E. J. (1966). Partial reinforcement: A hypothesis of sequential effects. *Psychological Review, 73,* 459-477.

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, 1,* 156-181.

Capaldi, E. J., and Miller, R. M. (2004). Serial learning in Rats: A test of three hypotheses. *Learning and Motivation, 35,* 71-81.

Maehr, M. L., Videbeck, R. (1968). Predisposition to risk and persistence under varying reinforcement-success schedules. *Journal of Personality and Social Psychology. 9,* 96-100.

Miller, R. M. (in preparation, a). An Analysis of Sequential Variables in Pavlovian Conditioning.

Miller, R. M. (in preparation, b). Sequential effects in the workplace: Novel views of workplace training and error rate analysis.