JOURNAL OF THE EXPERIMENTAL ANALYSIS OF BEHAVIOR 1972, 17, 363-368 NUMBER 3 (MAY)

# PROBABILITY LEARNING AS A FUNCTION OF MOMENTARY REINFORCEMENT PROBABILITY[1]

BEN A. WILLIAMS

THE COLORADO COLLEGE

Pigeons were trained on a probability learning task where the overall reinforcement probability was 0.50 for each response alternative but where the momentary reinforcement probability differed and depended upon the outcome of the preceding trial. In all cases, the maximum reinforcement occurred with a "win-stay, lose-shift" response pattern. When both position and color were relevant cues, the optimal response pattern was learned when the reinforcement probability for repeating the just-reinforced response was 0.80 but not when the probability was 0.65. When only color was relevant, learning occurred much more slowly, and only for subjects trained on large fixed ratio requirements.

Several investigators have suggested that various learning phenomena can be explained by the assumption that animals conditionally respond on the basis of their preceding response. For learning-set acquisition (Schusterman, 1964; Warren, 1966) and probability learning (Shimp, 1966) organisms have been assumed to ignore the overall reinforcement history for the response alternatives, even when such reinforcement history has been extensive, and instead choose their next response on the basis of the outcomes of the just preceding responses. In such cases, the major effect of reinforcement and nonreinforcement appears to be cueing differential behavior immediately after, instead of providing overall increments in the degree of excitation and inhibition.

Although the cue function of response outcomes has often been postulated as an explanatory principle, little is known about what is required for such cue function to develop. The present experiment explored some of the variables affecting such learning. Two response alternatives were presented in a probability learning situation. Each alternative had an overall reinforcement probability of 0.50, but the momentary probability varied as a function of the outcome of the preceding trial. If the pigeon can utilize outcome of the preceding trial as a cue, therefore, responding should be better predicted by the momentary probability than by the overall probability. Three factors affecting such learning were investigated: differences in momentary probability, cue saliency, and the response requirement. The last variable has been shown to facilitate several other varieties of complex learning (Williams, 1971*a*, 1971*b*, 1971*c*).

## EXPERIMENT I

### METHOD

*Subjects*

Six White Carneaux pigeons were maintained at 80% of their free-feeding body weights. Subjects 45, 46, and 56 were experimentally naive; Subjects 1, 2, and 3 had recently been trained on a series of color discrimination reversals with the stimuli red and green.

*Apparatus*

The test chamber consisted of a 12-in. (30.5-cm) cube enclosed within a larger chamber to dampen outside noise. Three inches (7 cm) below the top of the front panel of the inner chamber were two Gerbrands pigeon keys, 0.75 in. (1.9 cm) in diameter and 2.75 in. (7 cm) apart, center to center. The keys required at least 15 g (0.15 N) force for operation. Behind each key were two 7.5-w Christmas tree light

[1]This experiment was conducted when the author was a National Science Foundation Predoctoral Fellow and was supported by research grants NIMH 15494 and NIH-GM-15258 to Harvard University. Reprints may be obtained from the author, Department of Psychology, The Colorado College, Colorado Springs, Colorado 80903.

bulbs (red behind the left key, green behind the right key) that were illuminated except during the intertrial interval. Between and 3 in. (7.5 cm) below the keys was a 2- by 1.75-in. (5.0- by 4.5-cm) aperture through which the birds were fed when a grain hopper was activated.

*Procedure*

Both response keys were illuminated on each trial. For any given trial only one response key could produce reinforcement. The key designated correct was governed by a probability distribution constructed from a random number table. The overall probability of reinforcement for a peck to either key was 0.50. The momentary probability depended upon the outcome of the preceding trial. After a correct response, during phase one, the probability of reinforcement on the same key was 0.65 and the probability of reinforcement on the other key was 0.35. During phase two, these probabilities were 0.80 and 0.20, respectively. After an incorrect response, for both phases, the probability of reinforcement for the same response was zero and the probability of reinforcement on the other key was 1.0. The probability distributions were constructed such that probability x on trial n = probability x on trial n+1, with the limitation that sequences of reinforced trials for a given response could be no longer than eight for the 0.65 probability distribution, and no longer than 12 for the 0.80 probability distribution.

Whenever the correct key was pecked, a 2.8-sec access to the food hopper was provided, followed by 0.2 sec of complete darkness before the keys were illuminated for the start of the next trial. Pecks on the incorrect key resulted in 3.0 sec of darkness before the start of the next trial.

After key-peck training, two subjects were assigned to each of three response requirements. Subjects 45 and 3 were required to emit only one response per trial (FR 1); Subjects 46 and 2 were required to emit at least five pecks per trial (FR 5); and Subjects 56 and 1 were required to emit at least 15 pecks per trial (FR 15). The designated requirement had to be fulfilled entirely on one key. That is, correct and incorrect pecks were counted separately so that whichever first reached the FR requirement determined whether the trial was correct or incorrect.

For phase one of the experiment, the 0.65 probability distribution was presented for 60 sessions to Subjects 45, 46, and 56, and for 20 sessions to Subjects 1, 2, and 3. For phase two, the 0.80 probability distribution was presented for 50 sessions to Subjects 46, 56, 1, and 2, and for a total of 60 sessions to Subjects 45 and 3. For the latter two subjects, three sessions were interspersed between Sessions 35 and 36 of the 0.80 training in which reinforcement was scheduled on only one key position during the entire session, *i.e.*, a position discrimination. All sessions terminated after 48 reinforcements.

## Results

*Probability = 0.65*

Subjects shown in the left panel of Figure 1 were experimentally naive; subjects in the right panel had experience on a variety of complex learning tasks. Subjects in the upper, middle, and lower panels were trained on FR 1, FR 5, and FR 15, respectively. No consistent differences occurred as a function of either variable. Instead, five of the six subjects eventually adopted the same behavior pattern:

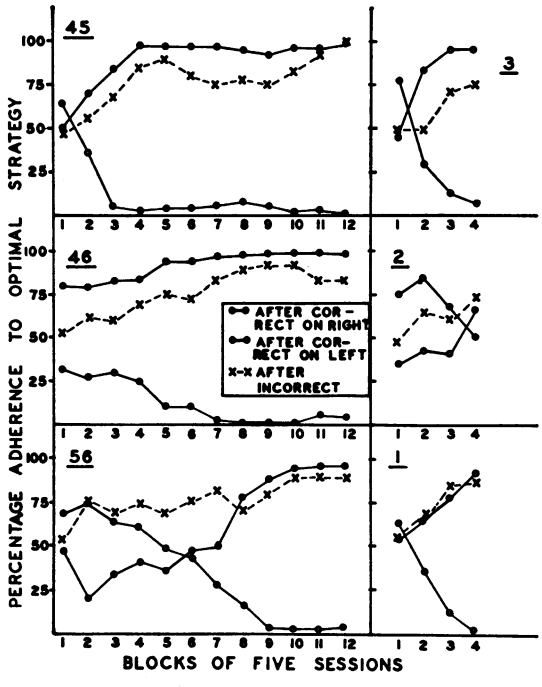


Fig. 1. Differential performance as a function of the outcome of the preceding response. Optimal behavior was to repeat a just reinforced response and to shift away from a just unreinforced response.

switching their response after incorrect trials and always choosing the same position (and color) after a correct trial. The behavior of the remaining subject, 2, had not stabilized, due to the small number of sessions and the fact that its position preference was apparently reversing. It is notable that the position preferences shown after a correct trial generally developed with training. With the possible exception of Subject 46, position preferences existing at the beginning of training did not predict the eventual stable behavior.

*Probability = 0.80*

Figure 2 shows the results of changing the probability that the response just reinforced would be reinforced on the next trial. Subjects trained on FR 5 and FR 15 gradually ceased their position preference after correct trials, and began repeating the response that had been reinforced on the preceding trial. For the three subjects for which this behavior was stabilized (2, 56, 1) adherence to the "win-stay, lose-shift" response pattern was 80 to 100%. In contrast, subjects trained on FR 1 showed no tendency to change their consistent

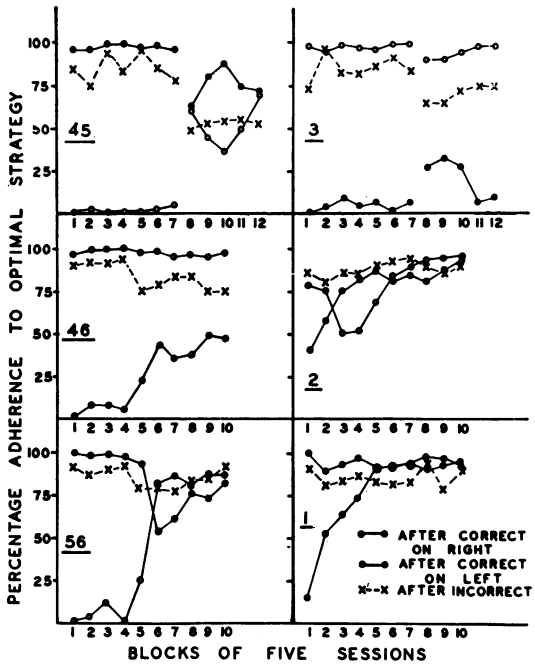


Fig. 2. Differential performance, with a probability of 0.80, as a function of the outcome of the preceding response. Optimal behavior was to repeat a just reinforced response and to shift away from a just unreinforced response.

position preference. After 35 sessions in which no change occurred, the FR 1 subjects had three sessions in which only a response to the nonpreferred position was reinforced, in order to determine if the elimination of the position preference would allow other response patterns to develop. Both subjects were responding only to their originally nonpreferred position at the end of the three days. When returned to the probability situation (see Figure 2), however, one subject, 3, quickly resumed its original preference. The second subject, 45, began gradually reverting to its original position preference, but then abandoned the position preference and began repeating, on 70% of the trials, the response just reinforced on the preceding trial.

### Discussion

The present experiment demonstrated that pigeons can learn to respond on the basis of their preceding response's outcome, but that such learning occurs only under a selected set of conditions. Learning did not occur for any subject when the probability of reinforcement for the response just reinforced was 0.65, and occurred only for subjects trained on larger response requirements when the probability was changed to 0.80. It is noteworthy that whenever the "win-stay, lose-shift" response pattern was not learned, a consistent response pattern nonetheless developed for all subjects. This pattern was to respond to the same position regardless of which response was reinforced on the preceding trial. The question posed is why less reinforcement was necessary to develop and maintain that response pattern than the pattern of repeating the response reinforced on the last trial.

## EXPERIMENT II

Since the "win-stay, lose-shift" response pattern occurred only under limited conditions, some exploration of the generality of the above findings would seem desirable. One question outstanding is whether such response patterns develop based upon cues other than position. That other cues might produce different results was suggested by the development in Experiment I of alternative response patterns, which presumably interfered with the "win-stay, lose-shift" learning. Cues other than position might be less susceptible to such inter-

ference. The present experiment explored this question by continuing the subjects in Experiment I on the same probability schedule but with position now irrelevant and color relevant.

## Method

The subjects, apparatus, and general procedure were unchanged from Experiment I. The only change was in which cues were relevant. Red and green each appeared on one of the keys each trial. The positions of the colors changed randomly after a correct trial but were unchanged after an incorrect trial. The probability of reinforcement for repeating a just reinforced response to color was 0.80; the probability of reinforcement for a response to the alternative color was 0.20. Since position was irrelevant, the probability of reinforcement for repeating the just reinforced response to position was 0.50. Since the positions of the colors changed only after correct trials, the probability of reinforcement for repeating a response nonreinforced on the preceding trial was 0.00, for both color and position. A total of 55 sessions were presented under this procedure. For Subjects 46, 2, 56, and 1, training was preceded by 20 sessions in which the conditions of Experiment I were continued but with varied intertrial intervals. For Subjects 45 and 3, the end of Experiment I was immediately followed by the start of the color-relevant training.

## Results

Figure 3 shows the results of the subjects trained on FR 1 (upper panel) and FR 5 (lower panel). Since all subjects continued to switch their response after an incorrect trial with 80 to 100% accuracy, only trials after correct trials are shown. The eventual performance for all four subjects was quite similar: performance with respect to color was random; performance with respect to position was essentially the same as that in Experiment I when the probability was 0.65, *i.e.*, the subjects adopted a position preference. This position preference was simply a continuation of the behavior at the end of Experiment I for FR 1 subjects (it is noteworthy that Subject 45 quickly abandoned the "win-stay, lose-shift" response pattern partly developed at the end of Experiment I) but gradually developed for the FR 5 subjects. The latter subjects, especially

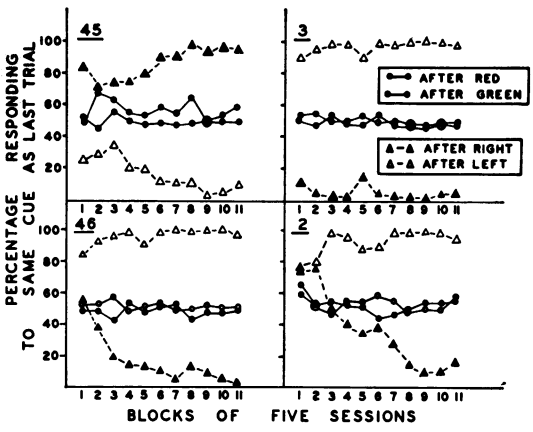


Fig. 3. Differential performance as a function of the preceding response for subjects trained on FR 1 and FR 5. Only color was a relevant cue.

2, initially repeated their response to the reinforcement position of the preceding trial, but this pattern gradually extinguished in favor of the position preference. At no time during the experiment did any of the four subjects respond differentially as a function of color associated with reinforcement on the preceding trial. It is noteworthy that during Experiment I, color and position were confounded, *i.e.*, subjects could have produced the data shown in Figure 2 by responding to position, color, or both. Apparently, that behavior was learned solely on the basis of position.

Figure 4 shows the two subjects trained on FR 15. By the end of training, both subjects repeated the color response reinforced on the last trial with a probability greater than chance (68% for Subject 56 and 74% for Subject 1). In addition, neither subject developed any substantial position preference. Instead, for one subject, 56, a substantial color preference was evident. The color preference gradually extinguished with training, however. Finally, it is noteworthy that the response pattern of repeating the color response reinforced on the preceding trial occurred more often than expected by chance at the beginning of training. Hence, the FR 15 subjects may have learned the "win-stay, lose-shift" response pattern during Experiment I for both position and color. When only color became relevant, the position response gradually extinguished and the color response was enhanced. As can be seen from Figure 4, however, such enhancement occurred extremely slowly.

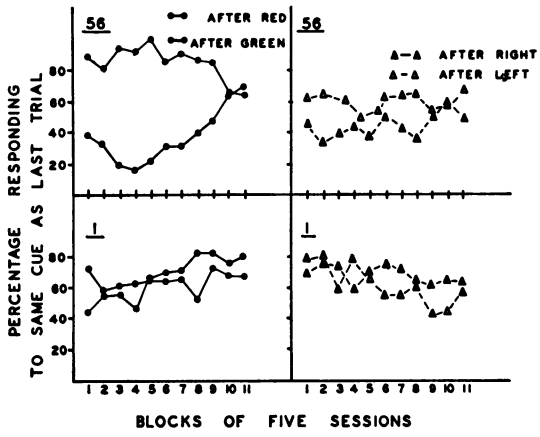


Fig. 4. Differential performance, for FR 15 subjects, as a function of the outcome of the preceding response. Only color was a relevant cue.

A final observation concerns the effects of the FR variable on accuracy for all pecks, not just those terminating a trial. In a previous experiment in which the FR variable was found to facilitate non-spatial delayed alternation learning (Williams, 1971*a*), a correlation was noted between the FR requirement and the number of changeovers between the two stimuli within a trial. That is, the pigeons often began responding to the incorrect stimulus and then corrected this tendency within a trial, with this correction more probable with larger FR requirements. Thus, the overall accuracy for pecks was slightly less than the overall accuracy for trials. In the present experiment, however, the two accuracy measures were not different. At no time during the experiment (or in Experiment I) did any subject show any meaningful tendency to changeover between colors within a trial.

## Discussion

The present results indicate that responding to the color associated with reinforcement on the preceding trial was a more difficult problem for pigeons to learn than responding to the position just associated with reinforcement. Only subjects trained on FR 15 learned to repeat the color response, and even they were responding with only 65 to 75% accuracy. The greater salience of the position cue is perhaps relevant to previous probability learning experiments that have found pigeons to be more likely to "match" when visual cues were used but to "maximize" when spatial cues were used (Bullock and Bitterman, 1962; Graf, Bullock, and Bitterman, 1964). It is noteworthy that experiments finding matching behavior (where response preference is proportional to reinforcement probability) have used some type of correction procedure and hence were similar in design to the present study. A major finding in the present study was that failure to learn the optimal "win-stay, lose-shift" response pattern was highly correlated with the adoption of other response patterns. Competition between response patterns must thus be considered a major ingredient of probability learning experiments, with the relative salience of relevant and irrelevant cues being an important determinant of which behavior pattern is eventually adopted. The poor learning with the visual cues in the present study suggests that previous matching behavior with visual cues need reflect nothing more than imperfect stimulus control (*cf.* Mackintosh, 1969).

A second finding of the present study was that learning was enhanced by larger fixed-ratio requirements. The failure of subjects trained with FR 1 to learn the "win-stay, lose-shift" response is consistent with the results of Williams (1971*a*), which found that non-spatial delayed alternation did not occur with fixed-ratio values less than five. The present experiment also suggests that it may be important for the ratio requirement to increase with the difficulty of the discrimination problem. Subjects trained with FR 5 learned the optimal response pattern with respect to position but not with respect to color, whereas subjects trained with FR 15 learned the response pattern with respect to both cues. This interaction between response requirement and problem difficulty is also consistent with previous work. Williams (1971*c*), using a successive discrimination reversal procedure, found that the difference in errors per reversal between subjects trained on different ratio values was increased as the difficulty of a visual discrimination was increased.

Although the present fixed-ratio effects are consistent with previous complex learning experiments, they differ from previous results in one significant respect. Williams (1971*a*) noted that the probability of an error decreased with the number of responses after the preceding reinforcement. Similar effects have been found by Nevin (1967) and Zeiler

(1968). On the basis of these data, and the additional observations that subjects trained on larger ratio values often switched their response between stimuli within trials, whereas subjects trained on smaller ratio values did not, Williams (1971*a*) proposed that larger FR values facilitated learning by allowing error tendencies present at the start of a trial to be corrected within the trial. The present results are inconsistent with this interpretation, however, in that larger FR values facilitated learning, even though changeovers between the two stimuli within a trial seldom occurred.

An alternative interpretation of the FR facilitation, which is consistent with both present and previous results, is that complex learning is facilitated by larger work requirements. Also supporting this hypothesis are the results of Ferster (1958, 1960) who found learning facilitation with larger ratio requirements in two different learning situations (matching to sample with pigeons and response "counting" with a chimpanzee). Whereas the present experiment required the ratio requirement to be completed within a trial, Ferster scheduled his ratio requirement across trials (reinforcement occurred after a fixed number of correct trials, each trial requiring only one response). Like the present experiment, however, he found increases in learning proficiency over the range of RF 1 to FR 15. The only apparent similarity between Ferster's method and the present procedure was the amount of effort required per reinforcement. Contrary to the effortfulness hypothesis, however, is a visual discrimination experiment by Elsmore (1971), who found similar facilitation of learning by varying the number of responses required per trial but found little effect of varying the amount of force required for a keypress. Elsmore's experiment suggests that effortfulness *per se* might not be a critical determinant of learning. Just what facet of the fixed-ratio requirement does facilitate complex learning remains open to question.

## REFERENCES


Bullock, D. H. and Bitterman, M. E. Probability-matching in the pigeon. *American Journal of Psychology,* 1962, **75,** 634-639.

Elsmore, T. F. Effects of response effort on discrimination performance. *Psychological Record,* 1971, **21,** 17-24.

Ferster, C. B. Intermittent reinforcement of a complex response in a chimpanzee. *Journal of the Experimental Analysis of Behavior,* 1958, **1,** 163-165.

Ferster, C. B. Intermittent reinforcement of matching to sample in the pigeon. *Journal of the Experimental Analysis of Behavior,* 1960, **3,** 259-272.

Graf, V., Bullock, D. H., and Bitterman M. E. Further experiments on probability-matching in the pigeon. *Journal of the Experimental Analysis of Behavior,* 1964, **7,** 151-157.

Mackintosh, N. J. Comparative studies of reversal and probability learning: rats, birds, and fish. In R. M. Gilbert and N. S. Sutherland (Eds.), *Animal discrimination learning.* New York: Academic Press, 1969. Pp. 137-162.

Nevin, J. A. Effects of reinforcement scheduling on simultaneous discrimination performance. *Journal of the Experimental Analysis of Behavior,* 1967, **10,** 251-260.

Schusterman, R. J. Successive discrimination-reversal training in one trial learning by chimpanzees. *Journal of Comparative and Physiological Psychology,* 1964, **58,** 153-156.

Shimp, C. P. Probabilistically reinforced choice behavior in pigeons. *Journal of the Experimental Analysis of Behavior,* 1966, **9,** 443-455.

Warren, J. M. Reversal learning and the formation of learning sets by cats and rhesus monkeys. *Journal of Comparative and Physiological Psychology,* 1966, **61,** 421-428.

Williams, B. A. Color alternation learning in the pigeon under fixed-ratio schedules of reinforcement. *Journal of the Experimental Analysis of Behavior,* 1971, **15,** 129-140. (*a*)

Williams, B. A. Non-spatial delayed alternation in the pigeon. *Journal of the Experimental Analysis of Behavior,* 1971, **16,** 15-21. (*b*)

Williams, B. A. Fixed-ratio schedule of reinforcement as a determinant of successive discrimination reversal learning in the pigeon. *Psychonomic Science,* 1971, **25,** 143-144. (*c*)

Zeiler, M. D. Stimulus control with fixed-ratio reinforcement. *Journal of the Experimental Analysis of Behavior,* 1968, **11,** 107-115.


*Received 2 August 1971.*
*(Final acceptance 22 December 1971.)*