SHORT-TERM MEMORY IN THE PIGEON: THE PREVIOUSLY REINFORCED RESPONSE[1]

CHARLES P. SHIMP

UNIVERSITY OF UTAH

Eighteen pigeons served in a discrete-trials short-term memory experiment in which the reinforcement probability for a peck on one of two keys depended on the response reinforced on the previous trial: either the probability of reinforcement on a trial was 0.8 for the same response reinforced on the previous trial and was 0.2 for the other response (Group A), or, it was 0 or 0.2 for the same response and 1.0 or 0.8 for the other response (Group B). A correction procedure ensured that over all trials reinforcement was distributed equally across the left and right keys. The optimal strategy was either a win-stay, lose-shift strategy (Group A) or a win-shift, lose-stay strategy (Group B). The retention interval, that is the intertrial interval, was varied. The average probability of choosing the optimal alternative reinforced 80% of the time was 0.96, 0.84, and 0.74 after delays of 2.5, 4.0, and 6.0 sec, respectively for Group A, and was 0.87, 0.81, and 0.55 after delays of 2.5, 4.0, and 6.0 sec, respectively, for Group B. This outcome is consistent with the view that behavior approximated the optimal response strategy but only to an extent permitted by a subject's short-term memory for the cue correlated with reinforcement, that is, its own most-recently reinforced response. More generally, this result is consistent with "molecular" analyses of operant behavior, but is inconsistent with traditional "molar" analyses holding that fundamental controlling relations may be discovered by routinely averaging over different local reinforcement contingencies. In the present experiment, the molar results were byproducts of local reinforcement contingencies involving an organism's own recent behavior.

Key words: short-term memory, choice, structure of behavior, molar *versus* molecular analysis, pigeons

Reinforcement contingencies allow one to study and control not only the function of behavior but also the structure of behavior. The structure, organization, or patterning of behavior can be controlled by "structural contingencies", in which the delivery of a reinforcing stimulus is contingent on a class of behaviors sharing some structural property (Shimp, 1976). Several common schedules of reinforcement, such as Sidman avoidance schedules and differential-reinforcement-of-low-rate schedules extend the traditional reinforcement contingency involving response-reinforcer contiguity to include behavior not coincident in time with the reinforcing stimulus. These schedules accordingly provide some control over the structure of behavior. More recently developed schedules of reinforcement permit an even greater control over quantitative features of the structure of behavior (Hawkes and Shimp, 1975; Shimp, 1969*b*, 1974).

The present experiment was designed to investigate further the general problem of the extent to which one may control the structure of behavior, that is, its sequential patterning. In particular, the present experiment examined the structure of behavior maintained by a discrete-trials method in which the local reinforcement contingency on a trial involved the behavior not only on that trial but the reinforced behavior on the previous trial as well. This experiment resembles one conducted by Williams (1972). In his experiment, overall reinforcement probability for a peck on one of two keys was 0.50, but the local reinforcement probability depended on the outcome of the previous trial in such a fashion that optimal behavior consisted of a "win-stay, lose-shift" response pattern: after a "win" *i.e.*, after a re-

[1]This work was supported in part by NIMH grants 16928 and 24537 and was begun when the author was a Fellow at the Center for Advanced Study in the Behavioral Sciences. The author wishes to thank Marilynne Moffitt, Robert Taylor, and Howard Rhoades for their assistance with the computer programming and for their help in conducting the experiment. Reprints may be obtained from the author, Department of Psychology, The University of Utah, Salt Lake City, Utah 84112.

inforced response, the probability of reinforcement for a peck on the *same* key was 0.80 (or 0.65 in some conditions) and was 0.20 for a peck on the other key. After a "loss", *i.e.,* after a nonreinforced response, the probability of reinforcement for a peck on the *other* key was 1.00 and was zero for a peck on the same key. The results indicated that the lose-shift component of the optimal behavioral pattern was learned rather quickly and well, but that the win-stay component was learned slowly at best.

The optimal behavioral pattern for one group in the present experiment was the same as in Williams's experiment, a win-stay, lose-shift strategy. For another group, the reinforced pattern was a win-shift, lose-stay strategy. We have noted that Williams found that a win-stay component was more difficult to learn than a lose-shift component. There appears, however, to have been a corresponding bias in the reinforcement of these components: first, the lose-shift component was reinforced with a probability of one, but the win-stay component was reinforced with a less-than-unity probability; second, the delay between successive responses in the win-stay component was 6 sec (3-sec blackout plus 3-sec reinforcement) but the delay in the lose-shift component was only 3 sec (3-sec blackout and no reinforcement). Accordingly, the present experiment was arranged so that each trial ended with the same stimulus, reinforcement: the cue for reinforcement on the next trial was the response reinforced on the previous trial. The intertrial-interval was manipulated in order to determine how memory for the previous response affects the approximation between a subject's behavior and the optimal behavioral pattern.

METHOD

Subjects

Eighteen experimentally naive White Carneaux pigeons were maintained at approximately 80% of their free-feeding weights.

Apparatus

Six standard two- and three-key Lehigh Valley Electronics pigeon chambers were interfaced to a Digital Equipment Corporation PDP-8e Computer that arranged contingencies, presented stimuli, and recorded data. Only the left and right keys were used in the three-key chambers. A minimum force of approximately 0.15 to 0.20 N was required to operate the keys.

Procedure

Key-peck training. From the beginning of magazine training and key-peck training an effort was made to prevent the onset of position biases. Thus, this early stage of pretraining was made to approximate single-alternation training, although the variability inherent in the early stages of training of course ensured that the approximation was only a very rough one.

Discrete-trials procedure. Each experimental session consisted of 100 discrete trials, each of which ended with the delivery of a reinforcer, 2.0-sec access to mixed grain. At the beginning of each trial, a houselight and two keys were illuminated. A single peck on the key to which reinforcement was arranged turned off the keylights and the houselight, turned on a light over the food hopper, delivered the hopper, and ended the trial. A single peck on the key to which reinforcement was not arranged turned off the keylights but not the houselight, and began a 5.0-sec correction interval (but see Pretraining for Bird 19), at the end of which the trial was recycled until the subject pecked the correct key and the reinforcer was delivered.

Intertrial interval. The intertrial interval, the blackout interval between the end of reinforcement for one trial and the illumination of the keys and houselight at the beginning of the next, was varied as shown in Table 1. For both groups, it was either 0.5 sec or 2.0 sec in Condition 1: the two intertrial intervals were selected randomly so that each appeared equally often. For Group A in Condition 2 and for Group B in Condition 3, the intertrial interval was either 0.5 sec or 4.0 sec. Since the reinforcer duration was 2.0 sec, there was a total delay of 2.5, 4.0, or 6.0 sec separating the onset of one trial from the response reinforced on the previous trial. This response determined the reinforcement probability as next described.

Arrangement of reinforcements. Over all trials, pecks on left and right keys were equally reinforced, but at the beginning of a trial after a reinforced left-key peck, the reinforcement probability on left and right keys was 0.8 and 0.2, respectively, for Group A. For Group B, the reinforcement probability on left and right keys was 0.0 and 1.0 (Condition 1) or 0.2 and 0.8 (Conditions 2 and 3). For Group A, at the

Table 1

	Group A (win-stay, lose-shift)			Group B (win-shift, lose-stay)		
Condition	*ITI (sec)*	*Reinforcement probability on trial n for the response reinforced on trial n-1*	*Number of days of training*	*ITI (sec)*	*Reinforcement probability on trial n for the response reinforced on trial n-1*	*Number of days of training*
1	0.5, 2.0	0.8	60	0.5, 2.0	0.0	20
2	0.5, 4.0	0.8	30	0.5, 2.0	0.2	20
3				0.5, 4.0	0.2	20

beginning of a trial after a reinforced right-key peck, the reinforcement probability on left and right keys was 0.2 and 0.8, respectively. For Group B, it was 1.0 and 0.0 (Condition 1) or 0.8 and 0.2, respectively. Thus, the response reinforced on the previous trial provided a better-than-chance basis for predicting the location of reinforcement. The optimal strategy was to repeat the response reinforced on the previous trial (Group A) or to switch and not repeat the response reinforced on the previous trial (Group B). If a subject's choice was incorrect, then the subject had to peck the other key after the correction interval to collect the reinforcement. The sequence of reinforced responses was experimentally controlled and it may help to understand the procedure to observe that this sequence could have been arranged in advance of an experimental session, because a subject's behavior did not affect it. In practice, the computer waited until the reinforcement arranged on a trial was collected and then determined the reinforced key for the next trial on the basis of which response was just reinforced. In this same manner, the response reinforced on the last trial of a session determined the response to be reinforced on the first trial of the next session. Experimental sessions were conducted five or six days per week.

Pretraining. After the initial key-peck training, subjects in Group A experienced approximately three months of training with a single intertrial interval of 0.5 sec. Condition 1 began immediately after the end of this pretraining. After the initial key-peck training, subjects in Group B experienced a more extended pretraining because severe position biases appeared immediately after the subjects were placed on the contingencies of Condition 1. Different experimenters manually shaped the key-pecking behaviors for Groups A and B, and this difference may have contributed to the different behaviors of the two groups when they were first exposed to the different contingencies in their respective versions of Condition 1. But for whatever reason, subjects in Group B required exposure to a modification of the basic procedure before their behavior was sufficiently free of obvious position biases. This modification involved more severe "punishment" of errors by means of longer correction intervals. This interval was lengthened for four weeks to 20 sec from its initial value of 5 sec. For three weeks thereafter, it was gradually reduced to its initial value of 5 sec without the reappearance of position biases, except for Bird 19, for which the interval had to remain at 20 sec because of a continued tendency for that subject to revert to position biases with any shorter correction interval. To this point in pretraining, the intertrial interval was always 0.5 sec. Eight additional weeks of training then followed with an intertrial interval of 2 sec and with the final value of the correction interval equal to 5 sec, except for Bird 19 for which this interval remained at 20 sec throughout the subsequent experiment. Finally, Condition 1 began with the intertrial intervals randomly distributed between values of 0.5 and 2 sec.

RESULTS

The objective of the present experiment requires us to focus on steady-state behavior. Also, acquisition data were subject to the idiosyncratic variables operating during pretraining and so their inclusion here is not justified. Thus, the results presented here are averages over the terminal five sessions of a condition. Table 2 shows the conditional probability of a choice of the key having the momentarily greater reinforcement probability. The conditional probability of a choice of the left key

Table 2

Conditional Choice Probability

Intertrial Interval (sec)	*Conditional Choice Probability*	*Group A Group Number* 1	3	4	6	7	8	9	10	11	12	13	14	15	*Average*
0.5	P(L|L)	0.984	0.985	1.000	0.956	0.990	0.986	0.870	0.986	0.960	0.991	0.986	0.948	0.969	0.970
	P(R|R)	0.958	0.919	1.000	0.959	0.917	0.975	0.897	0.995	0.847	0.941	0.979	0.962	0.921	0.944
2.0	P(L|L)	0.990	0.857	0.970	0.974	0.953	0.000	0.652	0.986	0.780	0.991	1.000	0.000	0.961	0.778
	P(R|R)	0.723	0.989	0.889	1.000	0.924	1.000	0.540	0.970	0.884	0.946	0.957	1.000	1.000	0.909
4.0	P(L|L)	0.988	0.801	0.941	0.405	0.953	0.922	0.638	0.612	0.922	0.623	0.990	0.760	0.717	0.790
	P(R|R)	0.850	0.595	0.914	0.914	0.735	0.869	0.675	0.952	0.324	0.430	0.194	0.785	0.737	0.690

		Group B Subject Number 16	17	18	19	20	*Average*
0.5	P(R|L)	0.873	0.974	0.986	0.829	0.920	0.916
	P(L|R)	0.858	0.893	0.976	0.950	0.829	0.901
2.0	P(R|L)	0.667	0.914	0.921	0.730	0.338	0.714
	P(L|R)	0.835	0.854	0.942	0.694	0.769	0.819
0.5	P(R|L)	0.887	0.925	1.000	0.791	0.985	0.918
	P(L|R)	0.818	0.872	0.921	0.956	0.909	0.895
2.0	P(R|L)	0.816	0.942	0.764	0.760	0.832	0.823
	P(L|R)	0.757	0.708	0.944	0.647	0.893	0.790
0.5	P(R|L)	0.816	0.943	0.842	0.602	0.727	0.786
	P(L|R)	0.882	0.937	0.928	0.829	0.901	0.895
4.0	P(R|L)	0.815	0.589	0.130	0.371	0.649	0.511
	P(L|R)	0.526	0.494	0.841	0.611	0.514	0.597

after a reinforced left-key peck is abbreviated as P(L|L), the conditional probability of a choice of the left key after a reinforced right-key peck is abbreviated as P(L|R) and so on. P(L|L) was computed by dividing the frequency of times a subject received reinforcement for a left-key peck and then chose left, by the total frequency of reinforced left-key pecks. The other choice probabilities were computed similarly. There was no meaningful difference in results for Group A between Conditions 1 and 2 on trials following a 0.5-sec intertrial interval, and Table 2 accordingly presents 0.5-sec data averaged over both conditions. Table 2 is based on only the first response in each trial: computations excluded responses under control of the separate correction contingency.

We asked first if there was any control over choice behavior by local reinforcement probability. That is, was the probability of choosing a given alternative dependent on the previously reinforced response? Table 2 clearly shows there was a difference for a 0.5-sec intertrial interval for all 18 subjects. For a 2-sec intertrial interval, there was a differerence for 15 of 18 subjects (for all but Subjects 8, 14, and perhaps 20, each of which developed a position bias). For a 4-sec intertrial interval, there was a difference for 12 of 13 subjects in Group A (all but Subject 12) but perhaps for only one of five subjects in Group B (Subject 16). Therefore, local reinforcement probability tended to control local choice probability, but to a lesser degree as the total delay between a choice and the stimulus correlated with local reinforcement probability increased.

Examine now in greater detail the extent of this dependency of local choice probability on the delay between a choice and the previously reinforced response. Table 2 shows that the longer the delay between choice and the cue correlated with local reinforcement probability, the poorer the control exerted by the latter on the former. First consider Group A. For a 2.5-sec delay between reinforced response and the subsequent choice, the average probability of choosing the key corresponding to the optimal win-stay, lose-shift strategy was 0.957. For all 13 subjects, this choice probability was greater than the probability-matching value of 0.800. For a 4-sec delay, the average probability of choosing the optimal key was less, 0.844, but still slightly greater than the probability-matching value. For eight of the 13 subjects, both P(L|L) and P(R|R) exceeded the probability-matching value. For a delay of 6 sec, the average probability of choosing the optimal key was 0.740 and for only three of the 13 subjects were both P(L|L) and P(R|R) greater than the probability matching value. However as noted previously, even after a 6-sec delay, local choice probability deviated from overall choice probability for 12 of 13 subjects, so that while the magnitude of control by local reinforcement probability was reduced, it was still measurable.

Now consider Group B. The average probability of choosing the key corresponding to the optimal win-shift, lose-stay strategy was 0.91 and 0.77 for intertrial intervals of 0.5 and 2 sec, respectively, in Condition 1. For each subject in this condition, the probability that a choice conformed to that required by the optimal strategy was greater after total delay of 2.5 sec than of 4 sec. Similarly, for each subject in Conditions 2 and 3, the probability of an optimal choice was less after a delay of 6 sec than after 4 sec. Thus, for both groups, the extent to which choice behavior approximated an optimal strategy depended inversely on the interval separating successive choices.

DISCUSSION

The present experiment provides information on the way in which the function of behavior and the structure of behavior are interrelated: local choice probability, and therefore local structure of behavior, was controlled by local reinforcement probability. It demonstrates also that the function relating these two probabilities is not fixed: local choice probability depends on the time separating it from the cue correlated with local reinforcement probability. The present results suggest further that local choice probability depends on the particular reinforced behavioral pattern. Stated differently, some kinds of reinforced behavioral organizations are more easily remembered than others. Thus, recall of the optimal response for the win-stay, lose-shift strategy generally was greater than for the win-shift, lose-stay strategy. Shimp (1966), Morgan (1974), and many others have found a tendency for a pigeon to perseverate on a response even in the absence of differential reinforcement for such perseveration. In the present experiment, such a tendency would assist performance in the win-stay, lose-shift condition but would interfere with performance in the win-shift, lose-stay condition. This natural tendency may therefore partially explain the better performance in the win-stay, lose-shift condition. The finding in the experiment by Williams (1972) that the win-stay pattern was more difficult than the lose-shift pattern apparently is attributable to features of the method of that experiment, as described above in the Introduction, that made recall of the win-stay pattern more difficult than recall of the lose-shift pattern.

The present results are consistent with the hypothesis that the structure, or organization of behavioral output, the temporal sequence of choices, approximates an optimal strategy but only to a degree permitted by a subject's short-term memory for recent events (Shimp, 1975; Silberberg and Williams, 1974). Put differently, behavior can approximate an optimal strategy only to an extent permitted by the precision with which short-term memory for recent events, such as a subject's own recent behavior, forms a component of the functional stimulus in the "presence" of which a subject chooses. The present results indicate that local reinforcement probability can play a somewhat more vital role than indicated by the results of the experiment by Williams (1972). Here, the "win-stay" component of the optimal strategy was much more closely approximated, for a local reinforcement probability of 0.8, than in the earlier experiment. The difference in results between the two experiments emphasizes that a general understanding of how reinforcement contingencies may control the structure of behavior requires an understanding of how an organism remembers its own recent behavior.

An interesting comparison can be made between the present results for the simple single-alternation condition with Group B and results obtained by Hearst (1962) with pigeons in a similar condition. Hearst's subjects performed markedly better than those here for similar retention intervals. His retention interval, like that here, was a blackout, but his reinforcement duration was half as long, 1 sec instead of the present 2 sec. This 1-sec difference may have contributed in two ways to the difference in results. First, the total retention interval was of course 1 sec shorter for a given intertrial interval in Hearst's study than here. But more importantly, a reinforcing stimulus is a very salient, powerful stimulus indeed. It can be assumed to have an interfering effect on recall similar to, but stronger than, that of other stimuli. Visual stimuli during a retention interval have been shown to interfere with subsequent recall (Moffitt, 1972; Zentall, 1973). Therefore, a response-reinforcer pairing may be assumed to be the occasion for retroactive interference: short-term memory for a response may be *reduced* by a reinforcer intervening between the response and subsequent recall of that response. This outcome is striking from the perspective of the Law of Effect that predicts only a simple "strengthening" effect on a response preceding a reinforcer.

The present reinforcement contingencies established various local behavioral patterns or structures. The present experiment therefore bears on a controversy over the appropriate "level of analysis" in our attempts to understand behavior established and maintained by reinforcement contingencies in general. The controversy is simply defined. A "molar" position emphasizes that there is some broad range of contexts in which basic laws of behavior are revealed only after one averages over different local reinforcement contingencies that have no controlling effects on behavior (Herrnstein, 1970; Herrnstein and Loveland, 1975). A "molecular" position emphasizes that there is some broad range of contexts in which basic laws involve local reinforcement contingencies and the behavioral patterns these contingencies establish, and that these laws are only obscured if one averages over the different contingencies (Hawkes and Shimp, 1975; Shimp, 1966, 1969*a*, 1975, 1976). One obvious research strategy to resolve this controversy is to determine empirically the range of contexts over which each level of analysis applies. Such a strategy over time builds up various categories of research methods including: a molar category including those methods for which basic laws are expressed in terms of averages over local reinforcement contingencies; a molecular category including those for which basic laws are expressed in terms of local reinforcement contingencies; an indeterminate category for those methods for which it is not yet certain which level of analysis, or indeed whether any level, is appropriate.

It would be easy, at least in principle, to resolve the molar-molecular controversy if an agreed-upon criterion were available to let one categorize a method as supporting either a molar or a molecular view. But no such criterion exists. Two criteria are frequently used, and they do not give the same categorization of methods. First, one could put into the molar category any method that produces elegant functions, or even just tolerably noise-free functions if these functions exclusively involve molar variables. This criterion of course places many methods and results into the molar category (Herrnstein, 1970). On the other hand, one could adopt a more conservative criterion and put methods producing molar functions, no matter how elegant, into the indeterminate category until it is shown that these results are not byproducts of, or heavily confounded with, the effects of uncontrolled local reinforcement contingencies. This more stringent criterion surprisingly results in an empty molar category: this category contains according to this criterion no method of which the present author is aware. The molecular category includes, among others, synthetic variable-interval schedules (Shimp, 1973), compound pacing schedules in general (Shimp, 1975) and probability-learning experiments such as the present one and others (Shimp, 1975; Shimp, 1966): in each of these and many other experiments, local reinforcement contingencies incontestably determine the local structure of behavior. But perhaps the largest category is the indeterminate category. This category contains many standard free-operant methods, such as variable-interval schedules, concurrent and multiple schedules with variable-interval components, and a great many others. These schedules do not provide for the rejection of the hypothesis that the molar results they produce are byproducts of, or importantly confounded with the effects of, subject-controlled local reinforcement contingencies (Hale and Shimp, 1975; Shimp, 1973, 1975). Thus, according to the more conservative criterion for categorizing methods, these schedules must be labelled indeterminate.

Any survey, ranging from the most cursory to the most exhaustive, of the experiments described in this journal will reveal a strong tendency for experimenters to average over local reinforcement contingencies for free-operant data but to refrain from doing so for discrete-trials data. The fact that according to one criterion the molar category described above is empty raises the serious and difficult question of the nature of the justification for averaging over the various local reinforcement contingencies that prevail so often in free-operant methodology.

REFERENCES

Hale, J. M. and Shimp, C. P. Molecular contingencies: reinforcement probability. *Journal of the Experimental Analysis of Behavior,* 1975, **24,** 315-321.

Hawkes, L. and Shimp, C. P. Reinforcement of behavioral patterns: shaping a scallop. *Journal of the Experimental Analysis of Behavior,* 1975, **23,** 3-16.

Hearst, E. Delayed alternation in the pigeon. *Journal of the Experimental Analysis of Behavior,* 1962, **5,** 225-228.

Herrnstein, R. J. On the law of effect. *Journal of the Experimental Analysis of Behavior,* 1970, **13,** 243-266.

Herrnstein, R. J. and Loveland, D. H. Maximizing and matching on concurrent ratio schedules. *Journal of the Experimental Analysis of Behavior,* 1975, **24,** 107-116.

Moffitt, M. *Short-term recognition memory in the pigeon.* Unpublished doctoral dissertation, University of Utah, 1972.

Morgan, M. J. Effects of random reinforcement sequences. *Journal of the Experimental Analysis of Behavior,* 1974, **22,** 301-310.

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

Shimp, C. P. Optimal behavior in free-operant experiments. *Psychological Review,* 1969, **76,** 97-112. (*a*)

Shimp, C. P. The concurrent reinforcement of two IRTs: the relative frequency of an IRT equals its relative harmonic length. *Journal of the Experimental Analysis of Behavior,* 1969, **12,** 403-411. (*b*)

Shimp, C. P. Synthetic variable-interval schedules of reinforcement. *Journal of the Experimental Analysis of Behavior,* 1973, **19,** 311-330.

Shimp, C. P. Time allocation and response rate. *Journal of the Experimental Analysis of Behavior,* 1974, **21,** 491-499.

Shimp, C. P. Perspective on the behavioral unit: choice behavior in animals. In W. K. Estes (Ed.), *Handbook of learning and cognitive processes,* Vol. 2. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1975. Pp. 225-268.

Shimp, C. P. Organization in memory and behavior. *Journal of the Experimental Analysis of Behavior,* 1976, **26,** 113-130.

Silberberg, A. and Williams, D. R. Choice behavior on discrete trials: a demonstration of the occurrence of a response strategy. *Journal of the Experimental Analysis of Behavior,* 1974, **21,** 315-322.

Williams, B. Probability learning as a function of momentary reinforcement probability. *Journal of the Experimental Analysis of Behavior,* 1972, **17,** 363-368.

Zentall, T. R. Memory in the pigeon: retroactive inhibition in a delayed matching task. *Bulletin of the Psychonomic Society,* 1973, **1,** 126-128.

Received 13 January 1976.
(Final Acceptance 24 May 1976.)