PROBABILISTICALLY REINFORCED CHOICE BEHAVIOR IN PIGEONS¹

CHARLES P. SHIMP²

BROWN UNIVERSITY

A single principle, "momentary maximizing", may account for much of a pigeon's steady-state behavior in both probability learning and concurrent variable interval experiments. The principle states that a pigeon tends to choose the alternative that momentarily has the higher probability of reinforcement. A successive discrimination procedure, which produced matching in an earlier experiment, produced here a tendency to maximize if training were adequately extended. Maximizing was produced also by other procedures, in which no reinforcing event was presented on some trials: one procedure did and two did not provide a bird with information about the availability of reinforcement on a key after an unreinforced response on the other key. The latter two procedures were analogous to concurrent variable interval schedules in two respects: the reinforcement probability on one key increased while a bird responded on the other key; and they produced matching. But sequential statistics suggested that matching resulted from momentary maximizing. Depending on the procedure, the tendency to maximize produced different relative frequencies of pecking a key for a fixed relative frequency of reinforcement. Computer simulation of maximizing behavior in several concurrent variable interval schedules produced matching and sequential statistics similar to those produced by a real bird.

The present sequence of experiments on two-choice behavior of pigeons reveals that a single principle may account for the apparent inconsistency of behavior produced by two kinds of experiments: discrete trial probability learning experiments and concurrent variable interval reinforcement schedules. This principle, "momentary maximizing", states that pigeons tend to choose the alternative that momentarily has the higher probability of reinforcement.

Graf, Bullock, and Bitterman (1964) performed several probability learning experiments, two of which produced choice proportions that most closely approximated matching for the greatest number of trials; that is, the relative frequency of a choice response ap-

proximated the relative frequency with which that response was reinforced. These two experiments employed either a successive or a simultaneous visual discrimination, a center key and a correction procedure. Other visual problems as well as non-visual problems lead to maximizing (Bullock and Bitterman, 1962; Graf *et al.*, 1964). In these experiments, maximizing was said to occur when the bird almost always chose the more frequently reinforced alternative. In short, Bitterman and his associates found that pigeons may match in visual problems but tend to maximize in non-visual problems.

Others (Catania, 1963; Herrnstein, 1961, 1964; Reynolds, 1963) have described the results of several concurrent variable interval experiments. The results show that pigeons produce a close match between the relative frequency of responding on a key and the relative frequency of reinforcement on it.

To summarize, pigeons maximize in all but one class of discrete trial probability learning experiments; in other words, in all but certain visual discriminations. Yet they consistently match in concurrent variable interval experiments. The purpose of the following experiments was to account in terms of a maximizing principle for many steady-state properties

¹This paper is based on a thesis submitted to the Department of Psychology, Brown University, in partial fulfillment of the requirements for the degree of Doctor of Philosophy. The research was partially supported by USPHS Grant MH-02456 (D.S. Blough, principal investigator) and by the author's NSF Graduate and Postdoctoral Fellowships. The author wishes to acknowledge his deep indebtedness to his adviser, Donald S. Blough.

²Reprints may be obtained from the author, Institute for Mathematical Studies in the Social Sciences, Stanford University, Ventura Hall, Stanford, California 94305.

of the results of both probability learning and concurrent variable interval experiments.

METHOD FOR EXPERIMENTS I TO IV

Subjects

Six male White Carneaux pigeons, experimentally naive and approximately one year old at the beginning of training, were kept at 80% of their free-feeding weights. Two additional birds were used in Exp IV. One was experimentally naive and the other had responded for several months on various fixed interval reinforcement schedules. Bird 6 died between Exp I and II.

Apparatus

A Lehigh Valley Electronics two-key pigeon chamber was modified by inserting a third key $2\frac{3}{4}$ in. above the two choice keys on the vertical line midway between them. The floor was raised 2 in. to permit easy reach of the center key. White noise masked most exterior sounds.

Solid state random coincidence gates programmed random events so that sequences of stimuli and of reinforcing events were unrestricted. All responses, and the corresponding stimulus and reinforcement conditions, were recorded sequentially and automatically by a Friden auxiliary paper tape punch. The data were then transferred to magnetic tape and analyzed on a LINC computer (Clark and Molnar, 1964).

Terminology

A choice of the lighted left or right key is denoted by A_0 or A_1 respectively. Three different definitions of "choice response" are used. The need for different definitions arises from the different experimental procedures. The definition appropriate to an experiment is given in the last paragraph of the Procedure section of that experiment. A distinction is made between "unreinforced" and "incorrect" responses. In Exp II, III, and IV, reinforcement was not always set up. When this was the case, responses were neither correct nor incorrect, but only *unreinforced*. However, where reinforcement was set up on the key not chosen, the response, which was unreinforced, was also *incorrect*. Thus, not all unreinforced responses were incorrect responses.

There is no need to distinguish between reinforced and correct responses.

Presentation of the feeder after a peck at the lighted left or right key is called E_0 or E_1 , respectively.

EXPERIMENT I

This was similar to one of the two visual probability learning experiments that have produced matching. The terminal choice probabilities obtained after prolonged training in Exp I generally deviated from matching towards maximizing. The results are interpreted in terms of maximizing together with the incomplete stimulus control that may be present in experiments which combine visual problems with correction procedures.

Procedure

Each of the 100 sessions was a sequence of 200 discrete trials. Each trial began by illuminating the center key with white light. A single peck to the lighted center key illuminated the two choice keys and darkened the center key. With probability $1/2$, both choice keys were red (denoted by S_0); with probability $1/2$, both choice keys were green (denoted by S_1). In the presence of red keys the probability of reinforcement for a choice of the right key was $1/4$, of the left key, $3/4$. In the presence of green keys the probabilities of reinforcement on the two keys were reversed. On each trial, reinforcement was on only one key.

A correct response, *i.e.*, a response to the key on which reinforcement was set up, darkened the choice keys and produced reinforcement, which was access to grain for $1\frac{1}{8}$ sec. The center key was lighted and the next trial began if a minimum inter-trial interval of $1\frac{1}{2}$ sec elapsed and if a bird did not peck on any of the darkened keys. Each response during the inter-trial interval reset the inter-trial interval timer.

An incorrect response, which in Exp I was the same as an unreinforced response, darkened the choice keys. Then the inter-trial interval had to elapse before the next onset of the center key. The same key color and the same designation of the key on which reinforcement was set up remained programmed on successive presentations of the lighted choice keys until the reinforced response was made and the trial ended. By definition, a trial be-

gan with the first lighted center key presentation after reinforcement and ended with the next reinforcement. The choice response was the first response within a trial to a lighted left or right key.

Results

The learning curves shown in Fig. 1 are representative. On the basis of a visual inspection of the learning curves, the experimenter decided that behavior over the last 20 sessions was representative of terminal performance for all six birds. Choice probabilities diverged from matching towards maximizing over the last 20 sessions in 11 out of 12 cases (see columns 1 and 2 of Table 1). There was a statistically significant correlation ($r = .88$, $t = 3.70$, $p < .05$; see formula 15.26.1 on page 529 in Hays, 1963) between the frequency of within-trial unreinforced responses, not including unreinforced choices, and the mean relative frequency of the less frequently reinforced choice; that is, there was a significant correlation between column 3 of Table 1 and the mean of $\Pr(A_1|S_0)$ and $\Pr(A_0|S_1)$. In particular, birds 1, 4, and 6 more often chose the less frequently reinforced alternative and, within trials, made more incorrect responses after an incorrect choice response.

Table 2 presents sequential statistics averaged over all six birds for six successive blocks of early sessions, and for individual birds. The statistics for an individual bird were based on trials over which choice probabilities appeared fairly stable from a visual inspection of the bird's learning curve. The blocks of trials were chosen so that these probabilities were not so close to maximizing that frequen-

cies of the less frequently reinforced response were too low to be reliable. Sequential statistics are not reported for the first session because the probabilities within the first session presumably were unstable.

Trial subscripts are often suppressed in sequential statistics in this paper. The trial order is always the same as in the following expression:

$$\Pr(A_{1,n}|A_{j,n-4}A_{k,n-3}A_{l,n-2}A_{m,n-1}),$$

which is to be read as the relative frequency of choice A_1 on trial n given choices A_j , A_k , A_l , and A_m (where i , j , k , l , and $m = 0$ or 1), on trials $n-4$, $n-3$, $n-2$, and $n-1$, respectively.

Fig. 1. The learning curves for birds 4 and 5. These representative curves show that early in training the birds approximately matched but that later they tended to maximize. The solid horizontal lines give the matching values.

Table 1
Relative Frequencies of Choices and Frequencies of Repetitive Unreinforced Responses Within Trials

	Relative frequencies of left key choices (A_0) given that the keys were red (S_0) or green (S_1) (sessions 80-100)		Frequency of repetitive unreinforced responses within trials (excluding unreinforced choices) cumulated over sessions 86 to 95
	$\Pr(A_0 S_0)$	$\Pr(A_0 S_1)$	
Bird 1	.77	.22	270
Bird 2	.89	.08	191
Bird 3	.88	.04	190
Bird 4	.82	.07	261
Bird 5	.86	.06	201
Bird 6	.90	.29	273
Mean	.85	.13	
$\Pr(E_0)$	.75	.25	

The effects on choice probability of several preceding response and event sequences were consistent among birds. In every case, $\Pr(A_0) < \Pr(A_0|A_0)$. In all instances but one, $\Pr(A_0|A_0) < \Pr(A_0|A_0A_0)$. Similarly, in all instances $\Pr(A_0|A_1A_1) < \Pr(A_0|A_1)$. All of these statistics reveal response perseveration; in other words, there was a greater probability of a given choice on trial $n + 1$ if that choice occurred on trial n , whether that choice was the one more likely or less likely to be reinforced.

In all cases, $\Pr(A_0) < \Pr(A_0|E_0)$. The second order reinforcing event sequences contained several reversals but it is generally true that $\Pr(A_0|E_0) < \Pr(A_0|E_0E_0)$, $\Pr(A_0|E_1E_1) < \Pr(A_0|E_1)$, $\Pr(A_0|E_1E_0) < \Pr(A_0|E_0E_0)$, and

$\Pr(A_0|E_1E_1) < \Pr(A_0|E_0E_1)$. In summary, the event statistics reveal positive recency; that is, there was a greater probability of a given choice on trial $n + 1$ if that choice were reinforced on trial n .

It should be noted that the sequential statistics are based on trial data for a given color and successive trials of a given color were separated by a random number of trials of the other color.

Discussion

The values of $\Pr(A_0|S_i)$ at the end of training deviated from matching towards maximizing, but clearly maximizing was not perfect. In fact, the mean probabilities, .85 and .13, were about midway between the correspond-

Table 2
Sequential Statistics

Sessions Birds	2-4 1-6	5-7 1-6	8-10 1-6	11-13 1-6	14-16 1-6	17-19 1-6
S ₀ trials						
Pr(A ₀)	.758	.814	.832	.817	.800	.868
Pr(A ₀ A ₀)	.785	.826	.843	.834	.824	.877
Pr(A ₀ A ₁)	.673	.760	.775	.743	.704	.810
Pr(A ₀ A ₀ A ₀)	.795	.844	.858	.849	.841	.884
Pr(A ₀ A ₀ A ₁)	.673	.770	.786	.766	.740	.813
Pr(A ₀ A ₁ A ₀)	.744	.736	.759	.752	.737	.816
Pr(A ₀ A ₁ A ₁)	.651	.678	.679	.634	.585	.696
Pr(A ₀ E ₀)	.800	.836	.853	.833	.818	.881
Pr(A ₀ E ₁)	.633	.753	.773	.769	.753	.832
Pr(A ₀ E ₀ E ₀)	.824	.846	.859	.844	.835	.889
Pr(A ₀ E ₀ E ₁)	.669	.776	.783	.781	.757	.836
Pr(A ₀ E ₁ E ₀)	.722	.801	.822	.787	.763	.846
Pr(A ₀ E ₁ E ₁)	.517	.664	.721	.701	.714	.782
Pr(A ₀ A ₀ E ₀)	.828	.850	.867	.857	.842	.891
Pr(A ₀ A ₀ E ₁)	.657	.759	.767	.754	.770	.832
Pr(A ₀ A ₁ E ₀)	.708	.767	.760	.710	.706	.799
Pr(A ₀ A ₁ E ₁)	.541	.691	.755	.777	.664	.754
S ₁ trials						
Pr(A ₀)	.226	.170	.161	.171	.150	.149
Pr(A ₀ A ₀)	.298	.203	.210	.234	.254	.214
Pr(A ₀ A ₁)	.205	.164	.152	.158	.131	.138
Pr(A ₀ A ₀ A ₀)	.386	.292	.238	.243	.243	.300
Pr(A ₀ A ₀ A ₁)	.304	.245	.213	.222	.209	.176
Pr(A ₀ A ₁ A ₀)	.262	.183	.208	.235	.262	.194
Pr(A ₀ A ₁ A ₁)	.182	.152	.145	.150	.123	.136
Pr(A ₀ E ₀)	.298	.211	.199	.218	.202	.195
Pr(A ₀ E ₁)	.201	.156	.148	.155	.131	.133
Pr(A ₀ E ₀ E ₀)	.346	.241	.210	.210	.268	.192
Pr(A ₀ E ₀ E ₁)	.292	.215	.194	.201	.168	.172
Pr(A ₀ E ₁ E ₀)	.284	.205	.202	.227	.186	.203
Pr(A ₀ E ₁ E ₁)	.173	.140	.137	.143	.122	.124
Pr(A ₀ A ₀ E ₀)	.373	.114	.183	.193	.287	.250
Pr(A ₀ A ₀ E ₁)	.276	.234	.226	.254	.244	.210
Pr(A ₀ A ₁ E ₀)	.282	.234	.209	.230	.189	.194
Pr(A ₀ A ₁ E ₁)	.181	.142	.137	.139	.116	.123

ing pairs of maximizing and matching, that is, (1.00, .75) and (.00, .25).

Inadequate experimental control over the following two kinds of behavior may be the reason why the terminal choice proportions were not closer to maximizing. (1) *Inattention to key color*. Imagine the following three conditions: first, a bird attends to, or observes, the key color on only 160 of the 200 daily trials; second, if key color is unobserved, a bird's choices are randomly split .50-.50 over the two alternatives; and third, if key color is observed a bird maximizes. Under these hypothetical conditions, $\Pr(A_0|S_0)$ would be expected to be about .90 and $\Pr(A_0|S_1)$ about .10. Any one of a large number of similar examples could produce the terminal choice proportions observed in Exp I. No existing data

prove that the visual problems which have produced non-maximizing ensure an observing response to the key color on each trial.

(2) *Repetitive incorrect responses within trials*. According to the correlation revealed in Table 1, it is possible that the correction procedure did not adequately control unreinforced responding and thereby caused in part the non-maximizing behavior of Exp I. To clarify point (2), consider the following one of several possible hypothetical and qualitative accounts. Repetitive unreinforced responding within a trial may reduce $\Pr(A_i|S_i)$ in the following way. Assume that in the presence of S_i , birds tend to repeat unreinforced A_i 's more often than unreinforced A_j 's, $i \neq j$. Further assume that unreinforced pecks on a key reduce the probability of choices of that

Table 2 (cont.)

Sessions Birds	2-10 1	2-15 2	2-15 3	2-20 6	77-95 1
S ₀ trials					
$\Pr(A_0)$	.793	.777	.777		.805
$\Pr(A_0 A_0)$	.830	.797	.792		.817
$\Pr(A_0 A_1)$	.649	.707	.724		.759
$\Pr(A_0 A_0A_0)$	.847	.814	.818		.825
$\Pr(A_0 A_0A_1)$	.694	.722	.722		.763
$\Pr(A_0 A_1A_0)$	.730	.726	.691		.772
$\Pr(A_0 A_1A_1)$	.529	.637	.684		.692
$\Pr(A_0 E_0)$	.825	.818	.795		.808
$\Pr(A_0 E_1)$	.705	.660	.724		.797
$\Pr(A_0 E_0E_0)$	.841	.831	.810		.808
$\Pr(A_0 E_0E_1)$	.740	.691	.771		.806
$\Pr(A_0 E_1E_0)$	.762	.765	.736		.799
$\Pr(A_0 E_1E_1)$	.574	.559	.593		.745
$\Pr(A_0 A_0E_0)$	.857	.846	.811		.812
$\Pr(A_0 A_0E_1)$	.744	.648	.730		.823
$\Pr(A_0 A_1E_0)$	.679	.708	.727		.780
$\Pr(A_0 A_1E_1)$	.526	.663	.670		.664
S ₁ trials					
$\Pr(A_0)$		.191		.230	.199
$\Pr(A_0 A_0)$		.226		.308	.229
$\Pr(A_0 A_1)$		.183		.206	.192
$\Pr(A_0 A_0A_0)$		.273		.387	.278
$\Pr(A_0 A_0A_1)$		.233		.246	.210
$\Pr(A_0 A_1A_0)$		.216		.275	.217
$\Pr(A_0 A_1A_1)$		.176		.199	.191
$\Pr(A_0 E_0)$		.224		.304	.274
$\Pr(A_0 E_1)$		.180		.205	.173
$\Pr(A_0 E_0E_0)$		.262		.317	.273
$\Pr(A_0 E_0E_1)$		.278		.235	.201
$\Pr(A_0 E_1E_0)$		.217		.303	.278
$\Pr(A_0 E_1E_1)$		.150		.198	.167
$\Pr(A_0 A_0E_0)$		.213		.393	.264
$\Pr(A_0 A_0E_1)$		.236		.280	.218
$\Pr(A_0 A_1E_0)$		.233		.279	.280
$\Pr(A_0 A_1E_1)$		.171		.186	.166

key. Under these conditions, birds that more often repeat incorrect responses would tend to have lower values of $\Pr(A_1|S_1)$. These or other sets of assumptions cannot be confirmed or rejected because the frequencies of the relevant statistics are too low.

Graf *et al.* (1964) did not obtain a tendency to maximize, perhaps because they did not present an adequately large number of trials. They presented only 1400 trials over which the probability of reinforcement for a given alternative was constant, as compared to the 20,000 trials of Exp I. Table 2 shows that $\Pr(A_0|S_0)$ was .76 and .81 while $\Pr(A_0|S_1)$ was .23 and .17 over sessions 2 to 4 and 5 to 7 respectively. On the first session $\Pr(A_0|S_0)$ was .61 and $\Pr(A_0|S_1)$ was .25. Thus approximate matching was obtained over those trials of Exp I that correspond to the 1400 trials run by Graf *et al.*

The positive recency obtained in Exp I conflicts with the statement (Graf *et al.*, 1964) that "the probability that the pigeon will choose a given alternative on any trial is independent of which of the two has been reinforced on the preceding trial." This difference in sequential statistics may be attributable to different methods of programming colors and reinforcement. The procedure of Exp I did not restrict sequences either of colors or of reinforcing events. Graf *et al.* used Gellermann orders for the color sequences and included at least one minority reinforcing event in each block of five trials of one color. It would be valuable to know whether or not pigeons can discriminate dependencies such as those programmed by Graf *et al.* If the pigeons did discriminate them in Graf's experiment, then perfect maximizing might have differed by some unknown amount from the exclusive choice of the more frequently reinforced alternative.

Consider an implication of the results of Exp I. First, remember that the best evidence for matching by pigeons in probability learning experiments has been collected from simultaneous and successive visual problems in which there is a center key and the subject makes a correction response after an unreinforced choice. Because Exp I was just such a successive visual problem, and yet produced a tendency to maximize after extended training, it may be tentatively concluded that the data obtained by Graf *et al.* were not asymptotic. It is natural to suspect that Graf's other evi-

dence for matching is also based on non-asymptotic behavior. Presently, there does not seem to be any convincing evidence of asymptotic matching by pigeons in probability learning experiments of the type discussed above.

EXPERIMENT II

Experiment II, and Exp III and IV, attempted to relate the maximizing tendency in probability learning experiments to the matching in concurrent variable interval experiments. One difference between the probability learning procedures discussed above and concurrent variable interval procedures is that, in the former, reinforcement is always set up on some key. Consequently, a probability learning experiment was run to study the effects of relatively large numbers of blank trials. By definition, a blank trial is one on which no reinforcing event can occur.

Procedure

The procedure of Exp I was modified in several ways. First, the inter-trial interval was reduced to 0 sec; that is, a peck to a lighted left or right key turned off those keys and simultaneously turned on the center key. As in Exp I, each choice key peck was preceded by a center key peck. Second, only one position problem was presented instead of two in succession. The probabilities that the left and right keys were correct were 1/4 and 3/4 (called $1-\pi$ and π respectively). When reinforcement was set up, it was set up on only one key. The designation of the correct key did not change until the correct response was made. Third, reinforcement was not always set up on some choice key as in Exp I; instead, there was only a probability r that a peck to the lighted center key would set up reinforcement. The probability that reinforcement was set up was independent of the designation of the correct key. Once reinforcement was set up it remained so until presented for a peck to the correct choice key. Fourth, the consequences of a peck at either key, given that reinforcement was not set up, were made to be discriminably different from the consequences of an incorrect choice, given that reinforcement was set up. If an incorrect response occurred after reinforcement was set up, then on their successive presentations the lighted center key and choice keys appeared yellow in-

stead of white and remained thus until the bird pecked the other key, *i.e.*, made the correction response, and collected reinforcement.

As long as the keys were white, the relative probabilities of reinforcement on the left and right keys were $1/4$ and $3/4$ respectively. When the keys were yellow, the probability of reinforcement on the choice key previously pecked was 0 and the probability of reinforcement on the other key was 1.

The choice response was defined to be any peck to a white choice key. The value of r , which was the probability that reinforcement was set up when the keys were white, was progressively reduced over sessions from 1 to $1/2$ to $1/4$. In all, 11 sessions were run.

Results and Discussion

For all three values of r , the birds successively chose the key on the right until a reinforcing stimulus occurred or the keys turned yellow. The birds pecked the left key only after the key color changed to yellow. The mean left choice probabilities for r equal to 1, $1/2$, and $1/4$ were respectively .01, .03, and .02, where theoretical matching was .25. Thus, the birds maximized.

When $r = 1$, the procedure of Exp II reduced to a spatial probability learning task with a center key and a correction procedure. Graf *et al.* performed a similar experiment but with $\pi = .30$ and without the stimulus change after an incorrect choice. They found that the terminal probability of the less frequently reinforced alternative ranged over the six birds from .30 to .08. Experiment II suggests that the tendency to maximize becomes much more pronounced after extended training. It may be tentatively concluded that the difference in absolute reinforcement probability does not produce the difference in relative frequency of responding observed between probability learning experiments and concurrent variable interval experiments.

EXPERIMENT III

Concurrent variable interval procedures that produce matching do not distinguish the consequences of a choice of either key, given that reinforcement is not set up, from the consequences of an incorrect choice, given that reinforcement is set up on the other key. Experiments III and IV explored the possibility that

the ambiguous consequences of non-reinforced responses would produce matching in a probability learning situation.

The key color change in the procedure of Exp II was omitted from the procedures of Exp III and IV. Consider the effects of this omission: the relative probability of reinforcement on a key increased with each unreinforced peck at the other key. An analogous increase in the relative probability of reinforcement occurs in concurrent variable interval experiments; that is, reinforcement is more likely to be set up on a key the longer a bird pecks at the other key (Skinner, 1950; Findley, 1958). Omitting the key color change also prevented the experimenter from determining which, if any, responses correspond to the correction responses of Exp I and II. All responses to lighted choice keys were therefore called choice responses.

For $\pi = 3/4$ and $r = 1/4$, there exists a unique sequence of choices such that each choice is of the alternative that momentarily has the greater probability of reinforcement (see Appendix A). That unique sequence is called the maximizing sequence. The maximizing sequence contains the less frequently reinforced response, or in other words, the less frequently reinforced choice sometimes has the greater probability of reinforcement. Some computation shows that for $\pi = 3/4$ and $r = 1/4$, the maximizing sequence is the following sequence of responses: right, right, left, right, right, left, *etc.*, on the first, second, third, fourth, fifth, sixth, *etc.*, choices between successive reinforcements. The sequence may end with reinforcement after any choice.

It is possible to show that if a bird momentarily maximizes, *i.e.*, if a bird chooses according to the above maximizing sequence, then the expected relative frequency of a response approximates the expected relative frequency of reinforcement of that response (see Appendix B). Momentary maximizing produces overall matching. Experiment III was run to determine whether pigeons match under the conditions described above, and if they do, whether it is because they momentarily maximize.

Procedure

The procedure of the 13 sessions of Exp III was identical to that of Exp II except that the change of key color was omitted. Each of the

three keys, when lit, was white. The values of π and r were $3/4$ and $1/4$ respectively. Each response to a lighted choice key was called a choice response.

Results and Discussion

Over each of the last two sessions the mean probability of a left key choice was .26. This value approximated the matching value of .25 and was representative of the individual probabilities. Of the 10 probabilities produced by the five birds over the last two sessions, the lowest was .19 and the highest was .34.

The maximizing sequence begins anew with each reinforcement. Therefore in Exp III and IV sequential statistics do not include sequences in which any but the terminal response was reinforced. The expression

$$\Pr(A_{1,n}|A_{j,n-2}A_{k,n-1})$$

represents the conditional probability of response A_1 on the n -th *post-reinforcement* choice, given that choices A_j and A_k occurred on post-reinforcement choices $n-2$ and $n-1$ respectively. The theoretical sequential statistics, which depend only on π and r and not on estimated parameters, compared reasonably well in several instances with the obtained sequential statistics shown in Table 3. For example, $\Pr(A_1|A_0)$ was .97 instead of 1.00, $\Pr(A_1|A_1)$ was .57 instead of .50, and $\Pr(A_1|A_1A_0)$ was .97 instead of 1.00.

The sequential statistics reveal that switching did not exactly follow the maximizing pattern. Sometimes birds switched too soon (e.g., $\Pr(A_1|A_0A_1)$ is too low) but at other times they switched too late (e.g., $\Pr(A_1|A_1A_1)$

is too high). There was a distribution of choice sequences instead of a single sequence. The existence of switching patterns other than the maximizing pattern is not strong evidence against the hypothesis that the birds were controlled by, and tended to maximize, the overall time rate of reinforcement. Consider two non-maximizing sequences that sometimes occurred: $A_1A_0A_1A_0 \dots$ and $A_1A_1A_1A_0A_1A_1A_0 \dots$. These two non-maximizing sequences produce respectively more switching ($\Pr(A_1|A_0A_1) = .00$ instead of 1.00) and less switching ($\Pr(A_1|A_1A_1) = .50$ instead of .00) than the maximizing sequence. But these two sequences do not produce large differences in the absolute reinforcement rate. Indeed, if to maximize were to earn N reinforcements in an average of 1 hr, then it would be estimated that the two non-maximizing sequences given above would produce N reinforcements in an average time interval different from 1 hr by less than 1 min.

Thus, the momentary maximizing hypothesis accounts nicely for the gross features of the data. And, although some nonmaximizing sequences produce reinforcement rates that may be indiscriminable from the maximizing rate, these sequences nevertheless reduce the closeness of the fit between the maximizing predictions and the data.

EXPERIMENT IV

Too much single alternation produced some of the discrepancies between predictions and data in Exp III. For example, $\Pr(A_1|A_0A_1)$ was too low. Experiment IV was designed to

Table 3
Obtained and Predicted Sequential Statistics (three-day average)

	<i>bird</i> <i>1</i>	<i>bird</i> <i>2</i>	<i>bird</i> <i>3</i>	<i>bird</i> <i>4</i>	<i>bird</i> <i>5</i>	<i>mean</i>	<i>predicted</i>
$\Pr(A_1)$	.75	.73	.80	.65	.74	.73	.75
$\Pr(A_1 A_0)$	.97	.98	.99	.91	.98	.97	1.00
$\Pr(A_1 A_1)$	.61	.55	.70	.40	.57	.57	.50
$\Pr(A_1 A_0A_1)$	.79	.62	.73	.38	.62	.63	1.00
$\Pr(A_1 A_1A_0)$	.97	.98	.99	.91	.98	.97	1.00
$\Pr(A_1 A_1A_1)$	.52	.52	.63	.37	.51	.51	.00
$\Pr(A_1 A_0A_1A_1)$	.58	.50	.59	.37	.50	.51	.00
$\Pr(A_1 A_1A_0A_1)$	.78	.63	.73	.39	.62	.63	1.00
$\Pr(A_1 A_1A_1A_0)$	.98	.98	.99	.94	.99	.98	1.00

The statistics listed in the table are based, according to the maximizing sequence, on non-zero frequencies. The actual relative frequencies of some of the theoretically non-occurring sequences were the following: A_0A_0 , .011; $A_0A_0A_0$, .001; $A_0A_0A_1$, .013; $A_0A_1A_0$, .093; $A_1A_0A_0$, .013; $A_1A_1A_1$, .235.

reduce superstitious reinforcement of switching and thereby to decrease the frequency of alternating. The procedural modification was similar to the change-over delay sometimes added to concurrent variable interval schedules to reduce the frequency of switching (Herrnstein, 1961).

Procedure

The procedure was identical to that of Exp III except that only every second successive A_1 could produce reinforcement. Thus, switching was never directly reinforced. Every center key peck set up reinforcement with probability $1/8$ and as in Exp III, $\pi = 3/4$. The maximizing sequence consisted of the repeated pattern of four pecks to the right key followed by two pecks to the left key. Again it was determined that momentary maximizing would result in approximate overall matching. As in Exp III, each peck at a lighted choice key was called a choice response.

Results and Discussion

Figure 2 shows that the birds matched. The mean choice probability is representative of individual data; the daily range of the individual choice probabilities approximated the daily range of the reinforcement probabilities. Of the 42 probabilities produced by the seven birds over the last six sessions, the lowest was .16 and the highest was .41. The mean choice probability was little changed by the non-reinforcement of switching but the sequential statistics, as shown in Table 4, were drastically changed; this was predicted on the basis of the new maximizing sequence. For example, $\Pr(A_1|A_0)$ was .47 (as opposed to .97 in Exp III) compared to the predicted .50 and

$\Pr(A_1|A_1)$ was .81 (as opposed to .57 in Exp III) compared to the predicted .75. Numerous statistics showed that the frequency of single alternation was greatly reduced.

The birds did not adequately learn to emit in pairs their successive choices of a given key. For example, $\Pr(A_1|A_0A_0)$ was only .69 instead of 1.00.

Another way to measure the extent to which the birds momentarily maximized was to examine the A_1 probability on particular post-reinforcement choices. The maximizing sequence for Exp IV began with four successive A_1 's followed by two A_0 's; hence, it was expected that the obtained A_1 probability would be high for the first four and low for the next two post-reinforcement choices. As shown in Table 5, the mean A_1 probability on each of the first four, the fifth, and the sixth post-reinforcement choice was respectively greater than, equal to, and less than the overall mean A_1 probability; that is, the first switch to the left key tended to occur on the specific post-reinforcement choice predicted by momentary maximizing. In Exp III the first switch tended to occur on the second and not on the predicted third post-reinforcement choice; this discrepancy may be related to the strong tendency to alternate.

The first and later switches did not always occur on the predicted post-reinforcement choice. Rather, there were built up, over many post-reinforcement sequences, distributions of locations for the first and later switches. Consider Exp III: the n -th post-reinforcement response, over the aggregate of all post-reinforcement sequences for a session, would appear at each of the three locations in the sequence $A_1A_1A_0$. If successive instances

Fig. 2. The relative frequencies of responding on the left key (A_0) and of reinforcement on the left key (E_0) over sessions. The two curves are essentially identical; i.e. the birds matched.

of the n -th post-reinforcement response were to occur with equal probability on each of the locations in the sequence $A_1A_1A_0$, then $\text{Pr}(A_1)$ would be expected to approximate $2/3$ on the n -th post-reinforcement choice. A similar argument applies to Exp IV. Table 5 shows that for both Exp III and IV, $\text{Pr}(A_1)$ did in fact approximate $2/3$ on later post-reinforcement choices, where it is plausible to assume equal probability.

The behavior in Exp III and IV was consistent with that in Exp I and II inasmuch as

the birds tended to maximize in all four experiments. The notion of momentary maximizing accounted for a good part of the data of Exp III and IV, *e.g.*, matching, all first order sequential statistics, and many higher order statistics.

COMPUTER SIMULATION
AND DISCUSSION

The similarity between the procedures of Exp III and IV on the one hand and concur-

Table 4
Obtained and Predicted Sequential Statistics (three-day average)

	<i>bird</i> 1	<i>bird</i> 2	<i>bird</i> 3	<i>bird</i> 4	<i>bird</i> 5	<i>bird</i> 7	<i>bird</i> 8	<i>Mean</i>	<i>Pre-</i> <i>dicted</i>
$\text{Pr}(A_1)$	.77	.79	.77	.61	.73	.71	.79	.74	.75
$\text{Pr}(A_1A_0)$	.45	.43	.56	.43	.48	.49	.45	.47	.50
$\text{Pr}(A_1A_1)$	.84	.86	.82	.77	.76	.77	.87	.81	.75
$\text{Pr}(A_1A_0A_0)$	.61	.75	.94	.60	.85	.54	.56	.69	1.00
$\text{Pr}(A_1A_0A_1)$	.90	.95	.95	.97	.97	.68	.92	.91	1.00
$\text{Pr}(A_1A_1A_0)$	.30	.20	.35	.19	.19	.46	.35	.29	.00
$\text{Pr}(A_1A_1A_1)$	.81	.83	.77	.65	.66	.78	.85	.76	.67
$\text{Pr}(A_1A_0A_0A_1)$	.94	.98	.98	.98	.98	.71	.93	.93	1.00
$\text{Pr}(A_1A_0A_1A_1)$	.91	.88	.85	.61	.65	.71	.87	.78	1.00
$\text{Pr}(A_1A_1A_0A_0)$	.60	.77	.94	.55	.85	.56	.53	.69	1.00
$\text{Pr}(A_1A_1A_1A_0)$	.28	.19	.34	.18	.18	.46	.34	.28	.00
$\text{Pr}(A_1A_1A_1A_1)$	.77	.79	.72	.63	.63	.78	.83	.74	.50
$\text{Pr}(A_1A_0A_0A_1A_1)$	.93	.90	.88	.62	.65	.78	.87	.80	1.00
$\text{Pr}(A_1A_0A_1A_1A_1)$	.77	.85	.70	.58	.71	.67	.83	.73	1.00
$\text{Pr}(A_1A_1A_0A_0A_1)$	.94	.97	.98	.98	.98	.66	.91	.92	1.00
$\text{Pr}(A_1A_1A_1A_0A_0)$	.59	.76	.94	.54	.85	.55	.53	.68	1.00
$\text{Pr}(A_1A_1A_1A_1A_0)$	.28	.19	.32	.17	.18	.45	.34	.28	.00
$\text{Pr}(A_1A_1A_1A_1A_1)$	.72	.75	.71	.62	.55	.78	.83	.71	.00

The statistics listed in this table are based, according to the maximizing sequence, on non-zero frequencies. The actual relative frequencies of the theoretically non-occurring three-tuples were the following: $A_0A_0A_0$, .043; $A_0A_1A_0$, .017; $A_1A_0A_1$, .048.

Table 5

<i>Post-reinforcement</i> <i>Choice number</i>	<i>Relative frequency</i> <i>of A₁ over the last</i> <i>three sessions of</i> <i>Exp III</i> <i>Mean of five birds</i>	<i>Relative frequency</i> <i>of A₁ over the last</i> <i>three sessions of</i> <i>Exp IV</i> <i>Mean of seven birds</i>
1	.95	.87
2	.65	.89
3	.70	.86
4	.69	.80
5	.70	.75
6	.67	.70
7	.65	.68
8	.66	.69
9	.69	.71
10	.69	.69
11	.65	.68
The mean of all choices	.73	.74

rent variable interval schedules on the other suggested that momentary maximizing produces the matching observed in the latter. A LINC computer was programmed to simulate the performance of a momentarily maximizing subject on several concurrent variable interval schedules. The inter-reinforcement intervals were approximately arithmetically distributed and the absolute response rates were governed by geometric interresponse time distributions.

The resulting stat-bird data are informative in two ways. First, Fig. 3 shows that for all the concurrent variable interval schedules, momentary maximizing did in fact produce matching. Second, Fig. 4 shows that the obtained first order sequential statistics closely correspond to those for one of three real birds, as reported by Reynolds (1963). The correspondence for the other two birds is not as close as that shown for the one bird. Figures 3 and 4 suggest that the matching in concurrent variable interval experiments may result from a tendency to maximize.

Considered together, Exp I to IV and the computer simulation show that the relative frequency of reinforcement on a key and the relative frequency of responding on that key are not invariably equal. Maximizing in different procedures leads to different relative

frequencies of responding on a key. In particular, maximizing may produce a relative frequency equal to 1 (with typical discrete trial probability learning experiments and Exp I and II) or equal to π (with the simulated concurrent variable interval schedules and Exp III and IV).

Concurrent fixed ratio data (Herrnstein, 1958) may be consistent with the maximizing notion. If two independent fixed ratio schedules are programmed on two keys, then maximizing consists of the exclusive choice of the key associated with the smaller ratio. Herrnstein obtained such behavior if the ratios differed by factors of 3 or 4 to 1. Other data collected by Herrnstein on dependent concurrent fixed ratio schedules indicated that the birds closely duplicated the ratio requirements on the two keys, to a large extent independently of the relative frequencies of reinforcement on the keys. For a given absolute response rate, the near duplication of the programmed ratios on the two keys approximated maximizing. Herrnstein's data agree with the present data in that the relative frequency of reinforcement apparently played a minor role in controlling choice behavior.

The following is suggested: a bird's tendency to maximize the absolute rate of reinforcement is a process that may affect the form

Fig. 3. The relative frequency of left key pecks (A_0) as a function of the relative frequency of reinforcement for left key pecks (E_0). Each point represents one stat-bird session. Matching corresponds to the solid line. These data show that simulated maximizing on various concurrent variable interval schedules produced matching.

Fig. 4. The relative frequency of two successive pecks on the same key as a function of the relative frequency of reinforcement on that key. Each point represents one stat-bird session. The stat-data appear to follow the same function as the data for bird 111 in a study by Reynolds (1963).

of the curve that describes the absolute rate of pecking a key as a function of the absolute rate of reinforcement on that key. This function has been shown to be curvilinear in one-key experiments (Reynolds, 1961) and linear in two-key concurrent variable interval experiments (Reynolds, 1963). We agree with Reynolds (1963) that the curve becomes linear as a by-product of matching. It is noted in addition that a pigeon has the opportunity to maximize only if the experimenter provides a second key, and that maximizing may produce approximate matching. In short, a maximizing tendency may produce the linear relation between the absolute response rates and reinforcement rates in concurrent variable interval experiments.

APPENDIX A

Reinforcement Probabilities and the Maximizing Sequence

Let $E_{0,n}$ and $E_{1,n}$ be the absolute probabilities of reinforcement for A_0 and A_1 on the n -th choice after the preceding reinforcement. Then $E_{0,1}$ equals r times $1-\pi$; that is, the probability that reinforcement sets up times the probability that A_0 is correct. Similarly, $E_{1,1}$ equals $r\pi$. The first row of Table 6 gives the values of $E_{0,1}$ and $E_{1,1}$ in Exp III. The maximizing response on the first post-reinforcement choice is A_1 because $E_{1,1} > E_{0,1}$. Reinforcement may be set up on A_0 for the second choice in two ways, assuming an unreinforced A_1 on the first choice: first, with probability $r(1-\pi)$ reinforcement was available for an A_0 on the first choice and thus still available for an A_0 ; second, with probability $(1-r)r(1-\pi)$ reinforcement may set up on A_0 on the second choice. Table 6 shows the continuation of this process up to the fourth post-reinforcement choice. The resulting maximizing sequence is $A_1A_1A_0A_1 \dots$

APPENDIX B

The Expected Relative Frequency of A_1 Assuming Maximizing

For each n , the larger of the two values of $E_{1,n}$ defined in Appendix A equals the proportion of all choice sequences, assuming maximizing, that end with reinforcement for response A_1 on the n -th choice. The ratio $w_1/(w_0 + w_1)$ is the expected relative frequency of A_1 choices where

$$w_0 = \sum_{n=1}^{\infty} E_{1,n}N_{0,n}$$
 and $N_{0,n}$ is the number of A_0 's

in the maximizing sequence up to the n -th choice and where

$$w_1 = \sum_{n=1}^{\infty} E_{1,n}N_{1,n}$$
 and $N_{1,n}$ is the number of A_1 's

in the maximizing sequence up to the n -th choice. Each $E_{1,n}$ is the larger of $E_{0,n}$ and $E_{1,n}$. An approximation to $w_1/(w_0 + w_1)$ was obtained by employing only the first 13 terms of each of the two infinite series corresponding to w_0 and w_1 . This approximation accounted for 97% of the sequences and resulted in a ratio of $.747 \approx .75 = \pi$.

It is emphasized that this approximate equality does not always hold. For example if $\pi = .75$ but $r = 1$, then the maximizing sequence is A_1A_0 . The expected relative frequency of A_1 is $w_1/(w_0 + w_1) = 1/(1 + 1/4) = .8 \neq .75$. This case corresponds to the usual discrete trials probability learning experiment with a correction procedure.

REFERENCES

Bullock, D. H. and Bitterman, M. E. Probability-matching in the pigeon. *Amer. J. Psychol.*, 1962, 75, 634-639.

Table 6
Reinforcement Probabilities Assuming Momentary Maximizing and $r = 1/4$, $\pi = 3/4$

Post-reinforcement choice number	$E_{0,n}$	$E_{1,n}$
1	$r(1-\pi) = 1/16$	$< r\pi = 3/16$
2	$r(1-\pi) + (1-r)[r(1-\pi)] = 7/64$	$< (1-r)r\pi = 9/64$
3	$r(1-\pi) + (1-r)[r(1-\pi)] + (1-r)^2[r(1-\pi)] = 37/256$	$> (1-r)^2r\pi = 27/256$
4	$(1-r)^3[r(1-\pi)] = 27/1024$	$< (1-r)^3r\pi + (1-r)^4r\pi = 189/1024$

- Catania, A. C. Concurrent performances: reinforcement interaction and response independence. *J. exp. Anal. Behav.*, 1963, 6, 253-263.
- Clark, W. A. and Molnar, C. E. The LINC: a description of the Laboratory Instrument Computer. *Ann. N. Y. Acad. Sci.*, 1964, 115, 653-668.
- Estes, W. K. Probability learning. In A. W. Melton (Ed.), *Categories of human learning*. New York; Academic Press, 1964, pp. 89-128.
- Findley, J. D. Preference and switching under concurrent scheduling. *J. exp. Anal. Behav.*, 1958, 1, 123-144.
- Graf, V., Bullock, D. H., and Bitterman, M. E. Further experiments on probability-matching in the pigeon. *J. exp. Anal. Behav.*, 1964, 7, 151-157.
- Hays, W. L. *Statistics for psychologists*. New York; Holt, Rinehart and Winston, 1963.
- Herrnstein, R. J. Some factors influencing behavior in a two-response situation. *Trans. N. Y. Acad. Sci.*, 1958, 21, 35-45.
- Herrnstein, R. J. Relative and absolute strength of response as a function of frequency of reinforcement. *J. exp. Anal. Behav.*, 1961, 4, 179-184.
- Herrnstein, R. J. Secondary reinforcement and rate of primary reinforcement. *J. exp. Anal. Behav.*, 1964, 7, 27-36.
- Reynolds, G. S. Relativity of response rate and reinforcement frequency in a multiple schedule. *J. exp. Anal. Behav.*, 1961, 4, 179-184.
- Reynolds, G. S. On some determinants of choice in pigeons. *J. exp. Anal. Behav.*, 1963, 6, 53-59.
- Skinner, B. F. Are theories of learning necessary? *Psychol. Rev.*, 1950, 57, 193-216.

Received November 11, 1965