OPTIMAL CHOICE

J. E. R. STADDON, JOHN M. HINSON, AND RICHARD KRAM

DUKE UNIVERSITY

We present a classification and theoretical analysis of discrete-trial and free-operant choice procedures in which reinforcement is assigned to one alternative only, or independently to both, is either always available or conditionally available, and is either "held" or not from trial to trial. Momentary-maximizing and (globally) optimal choice sequences are defined in terms of *initializing* and *marker* events. Free-operant choice is analyzed in terms of a *clock space* whose axes are the times since the last A and B choices. The analysis shows that most molar matching data are derivable from momentary maximizing, and that the momentary-maximizing hypothesis has not been adequately tested in either discrete-trial or free-operant situations.

Key words: discrete-trials, free-operant, choice, matching, momentary-maximizing hypothesis

Some years ago, Shimp (1966, 1969) proposed that the well-known matching relation between response and reinforcement ratios on concurrent variable-interval variable-interval (concurrent VI-VI) schedules (Herrnstein, 1961) could be explained by a response-by-response process that he termed *momentary maximizing*. This process requires that the animal learn something about the sequence of choices that maximizes the immediate probability of success and, by implication, the overall rate of reinforcement.

There has recently been a resurgence of interest in this view of choice, with new experiments both favoring (Silberberg, Hamilton, Zirias, & Casey, 1978) and attacking (Heyman, 1979; Nevin, 1979) Shimp's position. In going over these studies some gaps in our theoretical understanding of choice procedures are apparent. Procedures that appear intuitively similar, such as concurrent variable-interval schedules programmed with one or two interval timers, are treated as equivalent in the absence of formal proof. Few distinctions are drawn between free-operant and discrete-trial procedures, even though their properties are in some respects quite different.

The debate between advocates of molar

matching and momentary maximizing is not purely empirical: without a clear theoretical understanding of what is implied by momentary maximizing, adequate experimental tests for it cannot be devised. To evaluate maximizing as a total or partial explanation for matching, it is important to have exact information on (1) the effect of procedural variations on the optimal choice sequence, (2) the effect of variations in the sequence on the proportions of reinforcements obtained for each choice, and (3) the difference between momentary-maximizing and overall-maximizing (optimal) strategies. In this paper we attempt to answer these questions for some commonly used procedures. Our intention is not to resolve the matching-maximizing controversy, but rather to lay the theoretical groundwork for empirical investigation of it.

The schedules used in many choice experiments are quite complex. It is not difficult to program an experiment that will provide a year's work for an applied mathematician. We are not mathematicians, so we have been forced to restrict our attention to relatively simple arrangements. We provide a taxonomy of two-choice procedures beginning with discrete-trial probability learning and progressing through other discrete-trial procedures. We then extend the analysis to concurrent VI-VI and VI-VR schedules. This scheme suggests a few new procedures, some of which have interesting properties.

The discrete-trial procedures we consider are

This research was supported by grants from the National Science Foundation to Duke University, J. E. R. Staddon, Principal Investigator. Reprints may be obtained from J. E. R. Staddon, Department of Psychology, Duke University, Durham, North Carolina 27706.

Table 1
Summary of Discrete-Trial Procedures

<i>Class of Procedure</i>	<i>Procedure</i>	<i>Approximate Terms</i>	<i>Momentary-Maximizing Sequence</i>	<i>Is M. M. Seq. Optimal?</i>	<i>Init. Event, (Marker Event)</i>
SINGLE ASSIGNMENT	Single Assignment (SA)	Probability learning	<i>AA...</i>	YES	Choice
	SA with HOLD	P. L. with correction	<i>AB</i>	YES	Reinforce
	Single Assignment Conditional Availability (SACA)		<i>AⁿBAB...</i>	YES	Reinforce
DUAL ASSIGNMENT	SACA with HOLD	SHIMP procedure	<i>AⁿBAⁿB...</i>	NO	Reinforce (Switch)
	Dual Assignment (DA)		<i>AA...</i>	YES	Session (Choice)
	DA with HOLD	NEVIN procedure	<i>AⁿBAⁿB...</i>	NO	Session (Switch)

summarized in Table 1. The table is organized in terms of three properties: (a) whether reinforcement is assigned to only one choice at a time (*single assignment*) or independently to both (*dual assignment*); (b) whether a reinforcement once assigned is always available or only available with some probability (*conditional availability*); (c) whether once available on trial *n*, reinforcement remains available on subsequent trials until a correct response occurs (*hold*).

Theoretical discussion of each procedure deals with three questions: (a) what is the rule for momentary-maximizing behavior? (b) Is the momentary-maximizing sequence the same as the sequence that maximizes overall reinforcement rate (the optimal sequence)? (c) Do these sequences yield matching of relative response and reinforcement rates, and if not, how large are the deviations? How sensitive is the matching result to variations in the choice sequence?

The initial sections of the paper summarize the essential theoretical arguments for each procedure and provide illustrative examples. The latter part goes into additional quantitative detail for selected procedures.

DISCRETE-TRIAL EXPERIMENTS

Single Assignment

The simplest two-choice discrete trial procedure is so-called probability learning in which reinforcement is available for only one of the two responses on each trial. The to-be-reinforced alternative is determined by sam-

pling on each trial from a random distribution with probability *p* that alternative *A* will be selected and (usually) complementary probability $q = 1 - p$ that alternative *B* will be selected. In terms of the taxonomy of procedures outlined in Table 1, probability learning is a single assignment (SA) procedure. Note that the *availability* of a reinforcer, given *assignment*, is equal to unity.

As is well known, the optimal (global maximizing) strategy here (which is also the momentary-maximizing sequence) is exclusive responding to the higher-probability alternative. The closest free-operant parallel is the concurrent variable-ratio variable-ratio schedule.

Single Assignment with Hold

Animals of several species trained on the simple SA procedure will usually fixate on the richer alternative (cf. Sutherland & Mackintosh, 1971, for a review). To ensure sampling of both alternatives (a curious requirement, since reliable fixation on the richer alternative implies that the animal sampled both) a so-called correction procedure is often added. In the usual version, this means that once reinforcement is assigned to an alternative, it remains assigned, and no other assignment is made, until the correct response has occurred. For continuity with free-operant usage, we term this contingency a *hold*.

The single-assignment-with-hold procedure provides a convenient introduction to three concepts that will be useful in subsequent analysis. For every discrete-trial procedure, it is

possible to devise a sequence of choices such that each choice is the best (i.e., most likely to be reinforced) one available on that trial. We term this the *momentary-maximizing* sequence. This sequence may or may not be the same as the sequence that yields the highest overall rate of reinforcement, which we term the *optimal* sequence. In either case, the event that signals the beginning of the sequence is termed an *initializing event* (IE).

If, by convention, we always take A to be the majority alternative (i.e., probability $p \geq q$), then it is obvious that for the single-assignment-with-hold procedure the optimal and momentary-maximizing sequences are just AB (a lose-shift strategy) and the initializing event is the occurrence of reinforcement (F).

Matching

The relative frequency of A -choices, $A/(A+B)$, in the SA-with-hold-procedure predicted by either momentary or global maximizing is $1/(2-p)$, e.g., .80 for $p = .75$. This result is closer to matching between choice and reinforcement proportions than the exclusive majority choice to be expected in the no-hold procedure. As we shall see, this shift away from exclusive choice is a feature of optimal performance on any procedure in which assigned reinforcements are retained until collected (hold).

Single Assignment Conditional Availability

The next logical step suggested by the taxonomy of Table 1 is to relax the condition that the availability is equal to unity. Figure 1 illustrates this *single assignment, conditional availability* (SACA) procedure using a two-stage decision tree.¹ An assignment, once made, remains until reinforcement is collected, and no other assignment is made until this has occurred. Hence, probability p is sampled only once for each reinforcement delivery, whereas the availability probability, r , is sampled on every trial. Thus, availability on trial n does not imply availability on trial $n + 1$. The clos-

Fig. 1. Single-assignment, conditional availability (SACA) choice procedure.

est free-operant parallel to this procedure is the single-assignment random-reinforcement schedule studied by Morgan (1974).

Consider now the optimal choice sequence for the SACA procedure. Since the procedure is derived from SA with hold, the IE is the occurrence of reinforcement (F). Consequently, in deriving the momentary-maximizing sequence we need consider only negative ($\bar{F}$) outcomes for each choice. Clearly, the animal's task is to test two *hypotheses*, namely, $H1$: that reinforcement has been assigned to alternative A , and $H2$: that reinforcement has been assigned to B . These two hypotheses are mutually exclusive and exhaustive, since it is certain that reinforcement has been assigned to one alternative. The optimal sequence is then derived by choosing A or B on each trial according as $H1$ or $H2$ has the higher probability.

This is a problem in Bayesian inference: given a prior probability for each hypothesis, the outcome of each choice is then used to compute a revised, posterior probability. Since $p \geq 1 - p$, the first choice is A . If the outcome is F , then the optimal sequence begins again with A . If the outcome is $\bar{F}$, we compute the probability that $\bar{F}$ could have occurred, given $H1$, multiply this by the prior probability of $H1$ and divide the result by the total probability of $\bar{F}$. This quantity is simply the ratio of the ways that the event "no reinforcement for an A response" could occur and $H1$ be true, divided by all the ways that this event could occur. For any pair of successive trials:

$$P_{n+1}(H1) = \frac{P_n(H1)(1-r)}{P_n(H1)(1-r) + 1 - P_n(H1)} \\ = \frac{P_n(H1)(1-r)}{1 - rP_n(H1)}, \quad (1)$$

¹In actually implementing this procedure, it is sometimes more convenient to sample the assignment probability, p , after the availability probability has yielded a positive outcome. From the animal's point of view, however, this procedure is indistinguishable from the one described by Figure 1, in which the assignment is decided first. We treat the latter case, since it is easier to analyze.

where $P_n(H1)$ is the probability of $H1$ on the n th trial, and r is the availability probability in Figure 1. This is just an adaptation of the usual Bayesian formula,

$$P_{\text{post}}(H1) = \frac{P(H)P(E|H)}{P(E)}, \tag{2}$$

where $P(H)$ is the prior probability of H and E is the outcome event (cf. Feller, 1957, or a similar reference). On trial one, $P1(H1) = p$, so that equation 1 reduces to

$$P_2(H1) = \frac{p(1-r)}{1-pr}. \tag{3}$$

Since it is certain that either $H1$ or $H2$ is true, $P_n(H2) = 1 - P_n(H1)$, so that

$$P_2(H2) = \frac{1-p}{1-pr}. \tag{4}$$

To continue the sequence, we need to know exact values for p and r . If $p = .75$ and $r = .25$ continuation of this process yields the numbers shown in Table 2. Columns two and three show the best estimate of assignment probabilities, $P(H1)$ and $P(H2)$, the fourth column shows the optimal choice, and the last column shows the probability that food will be obtained only on postfood trial n .

The momentary-maximizing sequence (which is also the optimal sequence) divides naturally into two parts: an initial exclusive preference for the majority alternative, followed by simple alternation. Because each reinforcement is assigned only once, the Bayesian analysis indicates that the initial advantage of the majority alternative is discounted by a fixed number (here four) of initial, unreinforced majority choices. Thereafter, the probabilities of $H1$ and $H2$ alternate around .50 as choices alternate.

Data Analysis

The optimal sequence for the SACA procedure is not simply repetitive. Hence, the appropriate method of data analysis is one that records events as a function of trial number since the last initializing event. This technique is illustrated in Figure 2. The dotted contour shows the probability of the majority alternative, $P(A)$, on each post-IE trial predicted from the optimal sequence. The solid line shows hypothetical data consistent with approximately optimal choice: The majority alternative is chosen almost exclusively on the first four or so post-IE trials; thereafter choice prob-

Table 2 Optimal Choice Sequence for SACA Procedure				
Post-IE Trial No.	Probability of:		Choice	$P_n(F)$
	$H1$	$H2$		
1	.750	.250	A	.188
2	.692	.308	A	.141
3	.628	.372	A	.105
4	.559	.441	A	.079
5	.487	.513	B	.062
6	.559	.441	A	.059
7	.487	.513	B	.047
8	.559	.441	A	.045

abilities approximate .5. If the animal is responding optimally, more fine-grain analysis should show that the probability of alternation increases to a maximum after four post-IE trials or so.

Matching

The relative frequencies of A and B responses and reinforcements can be computed from the probabilities in Table 2. The probability that an A or B response will be reinforced on a given trial is:

$$P_n(F|A) = rP_n(H1) \tag{5a}$$

$$P_n(F|B) = rP_n(H2). \tag{5b}$$

Hence, the probability that the first post-IE trial will not terminate in reinforcement is given by

$$P(2) = 1 - P_1(F|A), \tag{6}$$

which is also the probability of reaching the second trial. In turn, the probability that the first post-IE reinforcement will occur on the second trial is equal to the probability of reach-

Fig. 2. Hypothetical results from the SACA procedure; dotted line indicates the optimal sequence (for $p = .75$, $r = .25$), solid line hypothetical data.

ing trial two, multiplied by the probability of reinforcement given response A , that is

$$P_2(F) = P(2)P_2(F|A) = P(2)rP_2(H1). \quad (7)$$

For the values shown in Table 1, Equation 6 yields $P_1(F) = P_1(F|A) = .75 \times .25 = .188$ and $P(2) = 1 - P_1(F) = .812$. Similarly, from Equation 7, $P_2(F) = .812 \times .25 \times .692 = .141$ and $P(3) = 1 - P_2(F) = .859$. The remaining probabilities were computed in the same way, from the general formula:

$$P_n(F) = (1 - P_{n-1}(F))rP_n(H), \quad (8)$$

where H is either $H1$ or $H2$, depending on whether response A or B occurred. The values of $P_n(F)$ for the first eight trials are shown in column four of Table 2. The relative frequency of reinforcement for responses A and B can be computed by summing the reinforcement probabilities for A and B responses across a sufficient number of post-IE trials. For 100 trials, which yields a cumulative probability of reinforcement within 10^{-5} of unity, the total probability of reinforcement for an A response sums to .75 and for a B response to .25 (i.e., equal to the .75/.25 value fixed by the single-assignment procedure). Similarly, the relative frequency of A and B responses can be computed by summing over all sequences the product of the number of A s (or B s) in the sequence times the probability of a sequence of exactly that length, that is $\sum_{n=1}^{\infty} K_n P_n(F)$, where n is the sequence length and K_n is the number of A s (or B s) in a sequence of that length. For $n = 1$ to 100 this yields the proportions A : .709 and B : .291. Consequently the proportions of responses and reinforcements correspond reasonably closely to matching, using the momentary-maximizing sequence and these values of p and r .

We are not aware of any published studies in which this procedure has been used. The distinctive, nonrepetitive nature of the momentary-maximizing sequence suggests that SACA may be a useful way to explore further the relations between molar matching and momentary maximizing. If, for example, good matching is found in the absence of anything approaching the momentary-maximizing sequence, the momentary-maximizing interpretation for matching would be correspondingly called into question.

Single Assignment Conditional Availability with Hold

In the simple SACA procedure, reinforcement is assigned after each reinforced trial, but reinforcement availability is determined independently on every trial. Hence, a reinforcement available for response A on trial n will not be obtained if A is not made, and may not be available for response A on the next trial. Thus, reinforcements "set up" (made available) on trial n may be "missed." The version of SACA most commonly used is one in which a reinforcement once available is held until the correct response occurs. This is the procedure used by Shimp (1966, Experiment III) and others (Silberberg et al., 1978, Experiment I, Shimp replication) as a discrete-trial analogue to concurrent VI VI schedules with a single timer (sometimes termed *interdependent* schedules).

The probabilities of $H1$ and $H2$ on each trial, and the associated probabilities of reinforcement, can be computed in exactly the same way as before (Equations 3 and 4), with the single difference that effective availability on the unsampled alternative, $r(n)$, increases as long as it remains unsampled: on the n th trial $r_A(n) = 1 - (1 - r)^m$, where m is the number of trials since choice A was last made. For the case $p = .75$ and $q = .25$, the relevant probabilities appear in Table 3. The first four columns are familiar from Table 2. The fifth and sixth columns show the probabilities that determine the momentary-maximizing sequence: the probabilities of food on trial n for choices A and B . The right-most column, as in Table 2, shows the probability that a sequence of choices will terminate with reinforcement on the indicated trial, i.e., the probability the post-IE choice

Table 3
Reinforcement Probabilities for SACA with Hold

Trial No.	Probability of:		(Momentary-Maximizing)			
	$H1$	$H2$	Choice	$P_n(F A)$	$P_n(F B)$	$P_n(F)$
1	.750	.250	A	.188	.063	.188
2	.692	.308	A	.173	.135	.141
3	.628	.372	B	.157	.215	.145
4	.800	.200	A	.350	.050	.185
5	.692	.308	A	.173	.135	.059
or						
5	.692	.308	B	.173	.135	.046
6	.800	.200	A	.350	.050	.104

sequence will be exactly n trials in length. These are the probabilities given by Shimp (1966, Appendix A).

Several features of Table 3 are of interest. First, as the conditional probabilities in columns 2 and 3 show, it is obvious that after the first post-IE trial, the momentary-maximizing sequence is a repeating one: $P_n(F|A) = P_{n+3}(F|A)$, $n > 1$. Second, the first post-IE trial is always unique. Third, the momentary-maximizing sequence does not maximize overall reinforcement rate. For example, consider the sequence shown in the last two rows of Table 3: AABABAB . . . It is easy to show that in terms of overall payoff this sequence is slightly better than the momentary-maximizing sequence. Taken out to 75 post-IE trials (within 10^{-6} of all cases), the mean probability of reinforcement per trial for the momentary-maximizing sequence is .223228, and for the alternative it is .223552. Computer simulation (200,000 trials) confirms the superiority of the alternative sequence. We do not know if this alternative is the optimal (best possible) sequence, but it is clear both here and in the dual-assignment-with-hold procedure to be discussed shortly, that the growth in availability of the unsampled alternative in hold procedures means that a "hill-climbing" algorithm such as momentary-maximizing will not always lead to the global optimum—although the shortfall is generally small.

Changeovers reset the objective probabilities to fixed values. Table 3 shows that for $p = .75$ and $r = .25$, an AB switch resets $P(F|A)$ to .350 and $P(F|B)$ to .050, and a BA switch resets them to .173 and .135. In the cause of terminological hygiene, we term (unreinforced) changeovers *marker* events rather than IEs, because a changeover does not initiate a unique sequence, and because reinforcement always indicates that the next choice should be A , no matter which choice was reinforced. Thus, reinforcement is the true IE.

Matching

The relative frequency of A and B choices to be expected from the momentary-maximizing and alternative sequences can be computed in the way just described for the SACA procedure. As Shimp showed, for the momentary-maximizing sequence with $p = .75$ and $r = .25$, a result close to matching is expected: for 75 post-IE trials, $P(A) = .75$. The alternative sequence

(AABABAB . . .) yields $P(A) = .67$, a considerable deviation from matching.

The momentary-maximizing sequence does not lead to matching at all values of p and r . For example, Figure 3 shows the ratio of $P(A)$ to $P_F(A)$, the proportion of reinforcements for A , for $p = .75$ and a range of r values from 0 to 1. The momentary-maximizing sequence changes from AAB to AB (alternation) at $r = .5$, and for some values of r , such as .05 or .55, the deviation from matching is substantial. Overall, the majority choice proportion tends to be below the proportion of reinforcements for that choice (undermatching). The value of r (.25) chosen by Shimp happens to be one for which matching and momentary-maximizing are consistent; for other p and r values, the difference between the two predictions may be substantial.

DUAL ASSIGNMENT

The SACA procedure assigns reinforcement to only one alternative at a time. A discrete-trials procedure closer to the usual concurrent VI VI schedule is one in which reinforcements are assigned independently to each alternative on each trial, with probabilities p and q that need not be complementary. We term this procedure *dual assignment* (DA). As with SA, response alternatives are presented concurrently in discrete trials, with only one response per

Fig. 3. Effect of availability (r) on the ratio of response and obtained reinforcer proportions for the momentary-maximizing sequence in SACA with hold. Assignment value (p) is .75. Left half of curve represents momentary-maximizing sequence AAB ; right half represents AB . The line parallel to the abscissa represents perfect matching.

trial. Again, as in the single assignment cases, reinforcement made available may be held until collected, or the reinforcement may be collected only on the trial in which it becomes available. The latter case is similar to free-operant conc VR VR, the former to (dual-tape) conc VI VI. Nevin (1969), most notably, has used DA with hold.²

In DA without hold, the probabilities of reinforcement for each alternative are always the initial reinforcement probabilities, regardless of the animal's behavior. Hence, the optimal strategy is simply to respond always to the majority alternative. In DA with hold, the situation is quite different. If the initial probability of reinforcement is p , then the probability increases to $1 - (1 - p)^n$, where n is the number of trials since the alternative was last sampled. Formally, for each alternative, the probability of an assignment on the n th trial since a response to alternative A is,

$$P_A(F|n_A) = 1 - (1 - p)^n \quad (9)$$

and similarly for alternative B . Each switch resets reinforcement probability for that alternative to its original value. Thus, the momentary-maximizing strategy is to respond to the majority alternative until the increasing probability on the unsampled side exceeds the initial majority probability. For example, with two alternatives with initial probabilities of reinforcement $p = .75$ and $q = .25$, the momentary-maximizing choice sequence is displayed in Table 4. After a transient effect of initial conditions, the probabilities settle into a repeating cycle associated with the choice sequence $AAAAB \dots$ (abbreviated A^4B), as shown.

As with SACA with hold, the momentary-maximizing sequence is not always the sequence that maximizes overall reinforcement probability. This can be seen by computing reinforcement probability to each choice in the sequence, as shown in column 4 of Table 4. In this example, the mean reinforcement probability per trial of the repeating sequence (e.g., Trials 2 to 6) is .7902. This value is not the maximum possible. For these initial probabili-

Table 4
Optimal Choice Sequence for DA Procedure

Trial No.	$P_A(F n_A)$	$P_B(F n_B)$	Momentary-Maximizing	
			Choice	$P_n(F)$
1	.750	.250	<i>A</i>	.750
2	.750	.438	<i>A</i>	.750
3	.750	.578	<i>A</i>	.750
4	.750	.684	<i>A</i>	.750
5	.750	.763	<i>B</i>	.763
6	.938	.250	<i>A</i>	.938
7	.750	.438	<i>A</i>	.750

ties, the best sequence of the general form $A \dots AB \dots$ is six responses to A followed by a response to B , which yields an average reinforcement probability per trial of .7935.

On all the symmetrical procedures we discuss, free-operant as well as discrete-trial, optimal behavior requires a return to the majority alternative after just a single minority choice. Hence, the best sequence of the form A^NB (e.g., A^6B for $p = .75$) is probably the optimal DA sequence (Houston & McNamara, 1981, show this analytically).

The true initializing event for the DA-with-hold procedure is the beginning of the experimental session, when the assignment probabilities are p and q . No subsequent event restores the probabilities to these values, and the momentary-maximizing sequence is fully determined from that point. As a practical matter, however, the marker event, a changeover, is of more interest. As Table 4 shows, the assignment probabilities are fully determined after either an AB or a BA switch: following an AB switch, they are $P(F|A) = 1 - (1 - p)^2$ and $P(F|B) = q$, and following a BA switch they are $P(F|A) = p$ and $P(F|B) = 1 - (1 - q)^2$.

Data Analysis

DA-with-hold experiments have usually been analyzed in terms of changeover probability as a function of majority run length. The present analysis suggests that separate plots of choice probability as a function of postswitch trial, for both AB and BA switches, would provide a better test for momentary-maximizing: since the momentary-maximizing sequence following an AB switch is $AAAAB$ and following a BA switch is $AAABA$, momentary-maximizing predicts a clear difference between these two plots. The clock-space analysis of free-operant DA procedures to be discussed shortly can also

²Nevin used a discrete-trial procedure in which reinforcements were set up by independent VI schedules rather than by probability generators. This procedure is equivalent to DA if the time between responses on successive trials is constant and the VI schedules are constant probability.

easily be adapted to the analysis of the discrete-trials case.

Matching

DA-with-hold procedures have often been used to try to distinguish between matching vs. maximizing interpretations of choice. From the example in Table 4, for the momentary maximizing sequence of A^4B , the relative frequency of A reinforcement is .8069, compared to the relative frequency of A responses of .80. For the alternative sequence A^8B , the relative frequency of A reinforcement is .8440, compared to a relative frequency of A choices of .8571. In both cases, there is good matching between relative frequencies of responses and obtained reinforcement. We show in a moment that this consistency between matching and momentary maximizing is a general characteristic of the DA procedure.

A final DA variant is one in which reinforcers are held for one choice, but not for the other, a discrete-trial parallel to conc VI VR. One can show by the techniques used earlier that, if the "ratio" alternative, A , has an initial reinforcement probability of .75 and the "interval" alternative, B , a probability of .25, then the momentary maximizing sequence is the same as that derived earlier for this set of probabilities, but the best sequence of the form $A^N B$ is different. The matching relations are slightly different also, because the value of $P(F|A)$ is not affected by the nonoccurrence of an A response. For the momentary-maximizing sequence A^4B , the relative frequency of A reinforcement is .797, a good approximation to the matching value of .80. For the optimal sequence, A^8B , the relative frequency of A reinforcement is .866 vs. the relative frequency of A of .889, again, a tolerable approximation to matching.

FREE-OPERANT EXPERIMENTS

Neither the SACA nor the DA procedure is a valid model for the usual two-timer conc VI VI schedule. The SACA procedure fails because reinforcements are assigned independently to each alternative in the concurrent schedule. Both fail because animals on concurrent schedules do not respond at fixed intervals of time.

A better discrete-trials model is a modification of the DA procedure in which the animals

are not required to respond on every trial. As trial duration is made smaller and smaller (or, equivalently, the magnitudes of p and q diminish to negligible values) the procedure converges on the conc VI VI schedule.

When trial duration is finite, the probability of reinforcement for choice A on the n th trial since the preceding A is given by Equation 9. As trial duration tends to zero, however, this probability is increasingly well approximated by

$$P_A(F|t_A) = 1 - \exp(-\lambda_A t_A), \quad (10)$$

where $P_A(F|t_A)$ is the probability of reinforcement for an A choice at time t_A since the last A choice and λ_A is the average rate of reinforcement programmed by a constant-probability VI schedule (cf. Feller, 1957; Staddon, 1977). The same analysis applies to the probability of reinforcement for choice B .

Optimal choice can now be defined in a way that is independent of exactly *when* a choice is made. Since the probability of reinforcement is defined by Equation 10 for any point in time, the momentary-maximizing choice is always to pick the alternative for which $P(F|t)$ is highest.

The momentary-maximizing condition is illustrated in Figure 4. The top panel shows a sequence of two A and two B choices. At the time of the second A choice, $t_A = a$ and $t_B = b$; at the time of the second B choice, $t_A = a'$ and $t_B = b'$. These values of t_A and t_B are plotted in the bottom panel of Figure 4. For convenience we refer to this as a *clock space*, since both axes measure times. Thus, in Figure 4 the coordinates for the hypothetical choice A are (a, b) , and for choice B are (a', b') . The line through the origin that divides this space into two regions is derived from Equation 10 as follows. Indifference between choices occurs when $P_A(F|t_A) = P_B(F|t_B)$ which, from Equation 10, after rearrangement yields the relation

$$t_B = \frac{\lambda_A t_A}{\lambda_B}, \quad (11)$$

where λ_A and λ_B are, as before, the average rates of programmed VI reinforcement for the two alternatives. Equation 11 defines a *switching line* of slope λ_A/λ_B which divides the clock space. With this form of representation, choice conforms to momentary maximizing if all responses lie between the switching line and the time axis of the response chosen. The choices in Figure 4 are optimal in this sense.

Fig. 4. Switching-line analysis of optimal choice on the concurrent variable-interval variable-interval schedule. Top panel: a sequence of *A* and *B* choices. Bottom panel: representation of each choice as a point in t_A - t_B clock space; t_A , t_B are the times since the last *A* or *B* choice. Other details are in the text.

It is obvious that this method of analysis can readily be applied as a test of the momentary-maximizing idea to data from free-operant choice studies. Moreover, the distribution of choices in the clock space is likely to provide clues to the process involved. For example, "errors" (non-optimal choices) might cluster around the switching line, i.e., in the region of the space where the probabilities of reinforcement for each choice are similar. Time-discrimination error is proportional to the interval, which implies that t_A and t_B may need to differ by a larger absolute value when both are large; hence, momentary-maximizing "errors" may increase with the absolute value of interresponse time.

The diagram in Figure 4 can be used to compute the optimal sequence of choices for any temporal distribution of choices. If choices are made at fixed intervals (which was approximated in Experiment 1 of Nevin's, 1969, study),

the sequence of optimal choices will converge on a closed path, indicating a fixed sequence. Figure 5 shows an example, the trajectory of optimal choices when $\lambda_A = 3\lambda_B$. The initial response occurs by convention at $t_A = t_B = 1$, and subsequent responses occur at unit time intervals. The horizontal and vertical dotted lines indicate that the occurrence of an *A* or a *B* choice sets t_A or t_B to zero; the 45° dotted lines indicate that during the time between choices, t_A and t_B grow at the same rate. The line of slope 3:1 through the origin is the switching line that determines the momentary-maximizing choice at any instant. The solid line forming a triangle shows the stable trajectory through the clock space for strictly periodic momentary-maximizing choice. This example illustrates an ambiguity characteristic of all cases where λ_A/λ_B is equal to an integer. If $\lambda_A/\lambda_B = N$, then after $n - 1$ occurrences of the majority response, the value of $P(F|t)$ is the

Fig. 5. Momentary-maximizing trajectory for periodic responding on conc VI VI. Line through the origin is the switching line. Successive *A* and *B* choices are indicated. Dotted line: path of representative point; solid line: closed path of optimal choices.

same for both choices; hence there are two possible momentary-maximizing sequences, depending on how this ambiguity is resolved. For the example in Figure 5, either *AAB* or *AAAB* is a momentary-maximizing sequence.

If choices are made at irregular intervals, the trajectory in the $t_A - t_B$ plane is not closed and there is no unique maximizing sequence.

The matching prediction can be derived directly from the maximizing sequence if this is known. The value of t_A and t_B , hence $P(F|t)$, is known for each choice in the sequence, so the relative frequency of reinforcement for each alternative can be computed and compared with the relative frequency of *A* and *B* choices. For the example in Figure 5, if $\lambda_A = .3$ and $\lambda_B = .1$, these values are .733 and .667 for the *AAB* sequence, and .746 and .750 for the *AAAB* sequence, both reasonable approximations to matching. The latter sequence is the overall maximum, with a mean probability of reinforcement per response of .325.

The momentary-maximizing sequence does not depend on the absolute values of λ , but the matching prediction does. For example, the relative frequency of *A* reinforcement is .7479 for the *AAB* sequence and .7497 for the *AAAB* sequence when $\lambda_A = .03$ and $\lambda_B = .01$: both values are closer to each other, and to the scheduled proportion of .75, than when $\lambda_A = .3$ and $\lambda_B = .1$, even though the ratio of λ_A/λ_B is the same in both cases. The overall-maximum sequence also varies slightly as a function of the absolute values of λ , but the difference between all sequences diminishes as λ decreases (i.e., as the ratio of response rate to reinforcement rate increases). This is illustrated in Table 5, which shows the expected probability of reinforcement per trial for eight sequences

of the form $A \dots AB$ for the two cases $\lambda_A = .3$ or $.03$. At the lower λ_A value, the difference in overall payoff probability per trial for the various sequences is negligible.

These neat calculations cannot, of course, be made for the realistic case in which responses occur at irregular intervals. Here the payoff probability for each choice must be individually computed and the totals summed to estimate the effectiveness of various choice strategies and their conformity to matching.

The switching-line momentary-maximizing analysis can readily be applied to concurrent VI VR schedules. In this case, the probability of reinforcement on the ratio alternative is constant and equal to the reciprocal of the ratio value, M . The switching line can be derived from Equation 10, thus

$$1/M = 1 - \exp(-\lambda_A t_A), \quad (12)$$

if *A* is the interval alternative. By conventional manipulations, this yields the relation

$$t_A = \frac{\ln(M/(M-1))}{\lambda_A} \quad (13)$$

as the equation for the switching line. Since λ_A and M are constants, this describes a straight line parallel to the t_B axis.

For the simple case where choices are made at fixed intervals, momentary maximizing permits only three possible allocations of responding. If the interresponse time is long, so that $1 - \exp(-\lambda_A t_A) > 1/M$, only *A* (VI) choices should occur. If the IRT is shorter, so that $1 - \exp(-\lambda_A t_A)$ is slightly less than $1/M$, choices should alternate; for still shorter IRTs, the momentary-maximizing sequence should be of the form $B^N A$. Thus the proportion of *A* choices is restricted to values between zero and .5, unless choices are all to the interval alternative.

It is clear that even if choices are made aperiodically, if the average rate at which choices are made is high (i.e., usually $1/M \gg 1 - \exp(-\lambda_A t_A)$), interresponse times to the ratio alternative will generally be shorter than IRTs to the interval alternative.

Matching

The clock-space switching-line analysis specifies only *which* choice should be made, not *when* a choice should occur. Hence, there is no simple analytic answer to the question, does momentary maximizing on conc VI VI imply

Table 5

Probability of reinforcement for periodic-choice sequences on concurrent VI VI ($\lambda_A/\lambda_B = 3$).

Sequence	$\lambda_A =$	
	.3	.03
<i>AB</i>	.3162	.03900
<i>AAB</i> *	.3232	.03913
<i>A²B</i> *	.3248	.03915
<i>A³B</i>	.3245	.03916
<i>A⁴B</i>	.3232	.03913
<i>A⁵B</i>	.3215	.03911
<i>A⁶B</i>	.3196	.03908
<i>A⁷B</i>	.3177	.03905

*momentary-maximizing sequences

matching? Once some response rule, such as random or strictly periodic choice, is given, this question can be answered. We consider the case of random choice in the following section.

Given a particular response rule, it is possible to see whether momentary maximizing leads to overall maximizing. For the cases we have examined, the two are not generally equivalent—momentary maximizing is not optimal—but the discrepancy (in terms of overall reinforcement rate) is generally small.

One response rule that has intuitive appeal is the idea that choices should be made so that either all, or most choices occur on the switching line. Because of the geometry of the clock space, this is impossible unless the time between responses is allowed sometimes to be infinitesimal. However, if choice rate is high relative to the time between VI reinforcements, a simple rule of this sort can be followed in concurrent interval-ratio schedules. Many ratio responses will occur between each interval response and because the temporal distribution of choices affects reinforcement probability for the VI alternative only, it is possible for all VI choices to lie on the switching line. In this case, the probability of reinforcement for each VI choice equals the fixed reinforcement probability for a VR choice, $1/M$. We also know for the VI choice that the obtained rate of reinforcement for that choice, R_A , equals $P_A(F|t_A)/t_A$, which equals $P_A(F|t_A)x_A$, where x_A is the interval response rate. Moreover, from the properties of ratio schedules, $1/M = R_B/x_B$. Since ex hyp. $P_A(F|t_A) = 1/M$, it follows at once that $R_A/x_A = R_B/x_B$, that is, perfect matching.

The assumption that animals on conc VI VR switch to the VI the instant that $P(F|t_A) > 1/M$ is unrealistic in two ways: VI choices may not occur at fixed intervals of time, and even if they do, there is likely to be some kind of switching threshold, so that $P(F|t_A) - 1/M$ must exceed some value, Δ , before a VI choice occurs. Thus, even for the simple case of periodic VI responding, the switching condition may be $\Delta + 1/M = P(F|t_A)$. If Fechner's law applies to the animal's estimation of probability, then Δ should be proportional to $1/M$, so we can write $\alpha/M + 1/M = P(F|t_A)$, $0 < \alpha < 1$. Substituting and rearranging yields the biased matching relation $x_B/x_A = (1 + \alpha)R_B/R_A$ which is in fact the usual empirical result on this procedure (Herrnstein & Heyman, 1979).

SIMULATIONS

In this section we look in more detail at the quantitative properties of some of these procedures. In most cases, the results were obtained analytically and checked by computer simulation. When we were not able to obtain analytic results, simulations alone were used.

As in all computer simulation, the size of the parameter space to be explored must be restricted in some way. With one exception, we consider only cases where the ratio of reinforcement probabilities is 3:1 and free-operant choice is either periodic or random. Houston and McNamara (1981) have independently proved for the free-operant case that optimal behavior almost always requires only a single choice of the minority alternative before returning to the majority. Our own work on concurrent VI VI schedules (Hinson & Staddon, 1981) shows that when there is no changeover delay pigeons approximate this pattern, responding either once or with a brief "burst" to the minority choice before switching back. Consequently, in the simulations we consider variations only in the pattern of majority choice.

The optimal choice sequence in the SACA procedure is of the form $A^N BABAB \dots$, where A is the majority choice and N is an integer. Figure 6 shows the effect on mean reinforcement probability, $P(F)$, and obtained majority-choice proportion, $P(A)$, of variation in N , the number of initial majority choices preceding alternation. These curves were obtained by the theoretical means described earlier (Table 2), considering post-IE sequences up to 100 trials in length. They were checked by direct Monte Carlo simulations with 100,000 trials per data point.

The top panel in Figure 6 shows the effect of majority run length, N , on mean reinforcement probability, $P(F)$, for an assignment probability, p , of .75 and four different availability probabilities: $r = .75, .5, .25$, and $.1$. It shows two features that appear to be characteristic of this procedure: The sensitivity of $P(F)$ to variation in N is highest at high values of r (top curve), negligible at low r values, where N can vary widely with little effect on average payoff. The optimal sequence (crosses) and the sequence that yields the proportion of majority choice, $P(A)$, closest to matching (unfilled circles) are similar or identical at all r values. The

Fig. 6. Effect of majority run length on overall reinforcement probability [$P(F)$] and majority choice proportion [$P(A)$] for the SACA procedure. Top two panels: effect of variation in availability probability, r . Bottom two panels: effect of variation in assignment probability, p . +: optimal sequence; 0: best matching sequence.

second panel of Figure 6 shows the value of $P(A)$ as a function of N for the same four values of r . Matching is indicated by the horizontal line at $P(A) = p = .75$. As r decreases, the optimal sequence shows some tendency to depart from the best-matching sequence, but as the function also grows increasingly shallow, the value of $P(A)$ associated with the optimal sequence remains throughout close to matching.

The third and fourth panels of Figure 6 show $P(F)$ and $P(A)$ as a function of N for an r value of .25 and four different p values from .5 to .95. As in the top two panels, the optimal and best matching sequences are similar and even when different in terms of the value of N , differ little in the value of $P(A)$ associated with them.

Figure 7 shows the effect of availability, r , on the matching relation for SACA with hold.

Each point shows the ratio of majority choice proportion, $P(A)$, to majority reinforcement proportion, $P_F(A)$ for the momentary-maximizing sequence for r values from 0 to 1.0. The function has a sawtooth form in which each break signals a change in the value of N , the majority run length. Three points are noteworthy about this and other similar plots: (a) The choice proportion never departs substantially from matching (indicated by the horizontal line through 1.0 on the ordinate), and each change in N causes the ratio $P(A)/P_F(A)$ to oscillate around the matching line; (b) for most r values the momentary-maximizing sequence leads to undermatching (majority choice proportion less than majority reinforcement proportion); but (c) the extent of possible deviations from matching decreases with decreasing r value. This is because the large value of N at low r values allows the single minority choice to occur when $P(F|A)$ is only slightly less than $P(F|B)$; at high r values, within a single trial $P(F|B)$ may change from less than to substantially larger than $P(F|A)$ before a switch can occur. It is relatively straightforward to prove that if a switch always takes place when reinforcement probabilities are precisely equal—i.e., $P(F|A) = P(F|B)$,—and if $P(F|A)$ is approximately the same for all A choices, then the overall result must be perfect matching. The practical requirement that the probabilities must differ by some amount be-

Fig. 7. Effect of availability (r) on the ratio of response and obtained reinforcer proportions for the momentary-maximizing sequence in SACA with hold. Assignment value (p) is .9. Each continuous curve represents a different majority run length; run lengths range from 1 (far right) to 8 (far left). The line parallel to the abscissa represents perfect matching.

Fig. 8. Effect of majority run length on overall reinforcement probability $[P(F)]$ and majority choice proportion $[P(A)]$ for the DA procedure. 0: optimal sequences; +: momentary-maximizing sequences.

fore a switch can occur, together with the fact that the majority probability grows faster than the minority one, imposes the overall constraint that leads to undermatching.

Figure 8 shows comparable results for the dual-assignment (DA) procedure. The curves shown were derived theoretically, but as with Figure 6, were also checked by Monte Carlo simulations. The top panel shows $P(F)$ as a function of majority run length, N , for sequences of the form $A^N B A^N B \dots$, where A is the majority choice. Curves are shown for four values of p from .075 to .75 with $p/q = 3$ throughout. Comparison with Figure 6 shows that $P(F)$ is much less sensitive to the value of N in the DA than in the SACA procedure. The overall maximizing (optimal, unfilled circles) and momentary-maximizing (crosses) sequences are slightly different at high p values, identical at low values. The fit of these sequences to matching cannot be simply represented in this figure because the proportion of reinforcements (e.g., $P_F(A)$ = reinforcements for A /all reinforcements) is not fixed by the values of p and q (as in the SACA procedure) but depends on the choice sequence. The bottom panel of Figure 7 shows $P(F)$ as a function of N for $p = .6$ and four different values of q . The same general pattern is evident: the momentary-maximizing sequence is at the same or a slightly smaller value of N than the optimal sequence. As the value of N increases above one or two, there is little further effect of N value on the value of $P(F)$.

Figure 9 illustrates the relative insensitivity of matching to the value of N (hence to the proportion of A choices) in the DA procedure. The four curves are for four values of p with $p/q = 3$ throughout. Each point on a curve corresponds to a different value for N , ranging from unity (alternation) through 99 (almost exclusive A choice). The ordinate is the proportion of A choices, $P(A)$ which is equal to $N/(N + 1)$, and the abscissa is the proportion of A reinforcements, $P_F(A)$. It is clear that as N increases beyond unity, the feedbacks in the procedure ensure good approximations to matching between $P(A)$ and $P_F(A)$, with the permissible deviations increasing as the value of p decreases. For example, even at the lowest p value, the majority run length can vary between 2 and 6 without the choice and reinforcement proportions deviating by more than 10% from matching. At higher p values, the function never deviates by more than 10% no matter what the upper limit of majority run length. Optimal and momentary-maximizing sequences (unfilled circles and crosses) show good approximations to matching at all p values. (Note that Figure 9 shows only one quarter of the usual unit square, since the remainder can be generated by symmetry.)

We explored the implications of the clock-space switching-line analysis of concurrent VI VI and VI VR schedules by assuming a random temporal distribution of choices. Thus, during each program cycle,³ the probability of a choice was P_c , which varied from .15 through .75. The probability of a reinforcement being "set up" for each alternative was p_A and p_B , where these two values were constrained such that $p_A + p_B = .003$; seven values of p_A , from .0001 through .0018, were examined at each P_c value. When a choice occurred, it was allocated to A or B according to the switching-line analysis, where t_A and t_B (see Figure 4) were measured in program-cycle (discrete-time) units. Results reflect 100,000 choices at each p_A value.

For each value of P_c , the number of A and B choices, and the number of A and B reinforcements, R_A and R_B , obtained were counted and the values of the ratios A/B and R_A/R_B at each value of p_A were noted. In every case, these pairs of ratios were well fitted by power func-

³We are simulating continuous time using a discrete-time machine; "program cycle" refers to the computations performed in each discrete-time period.

mal behavior is not consistent with simple matching (Staddon & Motheral, 1978). Heyman and Luce (1979), with slightly different assumptions, come to a similar conclusion about concurrent VI VI. We show here that providing choice rate is high momentary-maximizing leads to matching on concurrent VI VR no matter what the choice distribution.

5. Both in the simulations presented here and in others we have done, the usual deviation from matching is in the direction of undermatching. For example, in SACA-with-hold, at most r values, momentary-maximizing entails some undermatching (Figures 3 and 7) and on concurrent VI VI and VI VR, random choice entails undermatching that is inversely related to the rate of responding (i.e., the higher the rate of responding in relation to the rate of reinforcement, the less the undermatching, cf. Table 6). Undermatching is by far the commonest deviation found empirically (cf. Baum, 1974; de Villiers, 1977; Herrnstein & Heyman, 1979).

6. Sequential analysis (with respect to initializing or marker events) is appropriate as a test for momentary-maximizing in discrete-trial procedures, but apparently trivial procedural differences, such as the introduction of a hold in the SACA procedure, can make a substantial difference to the momentary-maximizing sequence. Analysis of choice sequences is inadequate as a test for momentary maximizing in free-operant procedures, however. For the dual-assignment (concurrent VI VI and VI VR) cases we discuss, the clock-space switching-line analysis is the only appropriate test.

7. If animals are behaving according to some kind of momentary-maximizing rule, then our initializing and marker events have the status of (trace) discriminative stimuli. Data analysis should therefore focus on the aftereffects of these events, rather than on behavior preceding them. Some studies have looked at data from this point of view (e.g., Morgan, 1974; Silberberg et al., 1978) but others have not (e.g., Nevin, 1979; Shimp, 1966).

8. Our analysis suggests a wealth of experimental tests. For example, if animals are behaving according to momentary-maximizing on SACA with hold, then parametric variation in availability, r , should yield systematic changes in the matching relation corresponding to the sawtooth plots in Figures 3 and 7. The SACA (no hold) procedure has never been used in an

experiment; it should be relatively easy to see if animals do in fact develop two-part choice sequences as predicted by momentary-maximizing. Measurement of individual choice inter-response times (t_A and t_B) should allow direct clock-space tests of momentary-maximizing in dual-assignment free-operant procedures (cf. Hinson & Staddon, 1981).

REFERENCES

- Baum, W. M. On two types of deviation from the matching law: Bias and undermatching. *Journal of the Experimental Analysis of Behavior*, 1974, **22**, 231-242.
- de Villiers, P. Choice in concurrent schedules and a quantitative formulation of the law of effect. In W. K. Honig & J. E. R. Staddon (Eds.), *Handbook of operant behavior*. Englewood Cliffs, N.J.: Prentice-Hall, 1977.
- Feller, W. *An introduction to probability theory and its applications* (2nd ed.). New York: Wiley, 1957.
- Herrnstein, R. J. Relative and absolute strength of response as a function of frequency of reinforcement. *Journal of the Experimental Analysis of Behavior*, 1961, **4**, 267-272.
- Herrnstein, R. J., & Heyman, G. M. Is matching compatible with reinforcement maximization on concurrent variable interval, variable ratio? *Journal of the Experimental Analysis of Behavior*, 1979, **31**, 209-223.
- Heyman, G. M. Matching and maximizing in concurrent schedules. *Psychological Review*, 1979, **86**, 496-500.
- Heyman, G. M., & Luce, R. D. Operant matching is not a logical consequence of maximizing reinforcement rate. *Animal Learning and Behavior*, 1979, **7**, 133-140.
- Hinson, J. M., & Staddon, J. E. R. Maximizing on interval schedules. In C. M. Bradshaw, E. Szabadi & F. Lowe (Eds.), *Recent developments in the quantification of steady-state operant behavior*. Amsterdam: Elsevier/North-Holland, 1981.
- Houston, A., & McNamara, J. How to maximize reward rate on two variable-interval paradigms. *Journal of the Experimental Analysis of Behavior*, 1981, **35**, 363-392.
- Morgan, M. J. Effects of random reinforcement sequences. *Journal of the Experimental Analysis of Behavior*, 1974, **22**, 301-310.
- Nevin, J. A. Interval reinforcement of choice behavior in discrete trials. *Journal of the Experimental Analysis of Behavior*, 1969, **12**, 875-885.
- Nevin, J. A. Overall matching versus momentary maximizing: Nevin (1969) revisited. *Journal of Experimental Psychology: Animal Behavior Processes*, 1979, **5**, 300-306.
- 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.
- Silberberg, A., Hamilton, B., Zirias, J. M., & Casey, J.

- The structure of choice. *Journal of Experimental Psychology: Animal Behavior Processes*, 1978, 4, 368-398.
- Staddon, J. E. R. Spaced responding and choice: A preliminary analysis. *Journal of the Experimental Analysis of Behavior*, 1968, 11, 669-682.
- Staddon, J. E. R. On Herrnstein's equation and related forms. *Journal of the Experimental Analysis of Behavior*, 1977, 28, 163-170.
- Staddon, J. E. R., & Motheral, S. On matching and maximizing in operant choice experiments. *Psychological Review*, 1978, 85, 436-444.
- Sutherland, N. S., & Mackintosh, N. J. *Mechanisms of animal discrimination learning*. New York: Academic Press, 1971.

Received May 7, 1980

Final acceptance January 14, 1981