

Dynamics of time matching: Arousal makes better seem worse

JOHN GIBBON

*New York State Psychiatric Institute and Columbia University
New York, New York*

Matching of time allocation across alternatives in proportion to relative reinforcement rates is a ubiquitous finding in the animal-learning literature on choice. The dynamics of the underlying mechanism, however, remain poorly understood. A recent finding by Belke (1992) profoundly challenges scalar expectancy theory (SET; Gibbon et al., 1988) and other accounts of matching in concurrent variable interval (VI) schedules. He studied concurrent probe tests of stimuli associated with equal VIs but trained in alternative concurrent pairs. In training, one was preferred and the other not. Unreinforced probes revealed a strong preference for the alternative preferred in training. An experiment is reported replicating this result and showing that it is not due to generalization of preference levels from training. When the probe is between the two preferred training stimuli, the richer schedule is unpreferred. A SET account of these results is presented which implicates two processes in time allocation: (1) the choice between alternatives based on memory for delays to reinforcement, and (2) the times at which such choices are made. The former process is sensitive to reinforcement scheduling; the latter is sensitive to arousal levels induced by overall reinforcement rates in training.

A now venerable tradition in operant schedule research shows that animals choosing between concurrently available variable schedules of reward allocate time spent at each alternative in proportion to their rates of payoff (the matching law; Herrnstein, 1970; see Williams, 1988, for a review). The matching law is ubiquitous throughout the animal kingdom. A form of matching is seen at the level of groups of animals distributing individuals across resource sites where it is known as the "ideal free distribution" (see Gallistel, 1990, for a review). At the level of individuals, where it has been studied extensively in the animal-learning literature, accounts of matching have engendered an enduring controversy about its source. Matching of behavior proportionate to payoff implies more choices for the less profitable resource than strictly necessary to maximize overall payoff rates. Several "molecular" or "quasi-molecular" accounts of matching, which posit maximizing payoffs at a local or molecular level, have been proposed. For example, melioration (Herrnstein & Vaughan, 1980; Vaughan, 1985) and "momentary maximizing" (Shimp, 1969) argue that subjects choose the better of two local rates or probabilities of reinforcement, and this results in overall matching. Rather than explaining it, other, "molar" matching accounts simply assume matching (e.g., Baum & Rachlin, 1969; Gallistel, 1990; Mark & Gallistel, 1994).

The author is grateful for extensive discussion of these ideas with C. R. Gallistel, who lent formative insights into an understanding of switching rates. Thanks are due S. Fairhurst for data collection and analysis. The work was supported by NIMH Grant MH41649-09. Address correspondence to John Gibbon, Biopsychology Unit 50, 722 W. 168th St., New York, NY 10032 (e-mail: jg34@columbia.edu).

Our account of matching (scalar expectancy theory [SET]; Gibbon, Church, Fairhurst, & Kacelnik, 1988) explained matching as a consequence of choice based on memory of the delays to food associated with each alternative. We argued that subjects remember a variety of delays for each alternative in a variable interval (VI) schedule. Each of these delays is represented in memory by a noisy perturbation governed by the scalar property. The standard deviation of the remembered times is proportional to the size of the interval being remembered.

Figure 1 shows a mixture of memories for a variable schedule of delays delivered at a constant probability, the Poisson process common in nature, and in the laboratory for scheduling variable delays to reinforcement. The geometric distribution of intervals is spaced evenly in time, and the relative frequency with which each is experienced decreases exponentially, proportional to the height of the spikes. The smooth undulating curve represents a mixture of gaussian distributions associated with each interval weighted by its probability of occurrence. The scalar property on variance of remembered times enhances the skew in the inducing, mixing distribution of the VI so that the result is a severely skewed distribution with an exponential right wing. We approximated these memory distributions by exponentials (shown as the heavy line function in the figure). This is a parameter-free approximation since the exponential mixing distribution dominates the mixture so that subjects with very different sensitivity to time nevertheless are all expected to represent a VI schedule exponentially with a common mean. We proposed that choice between two alternatives was governed by sampling a delay from each and simply choosing the alternative with the smaller

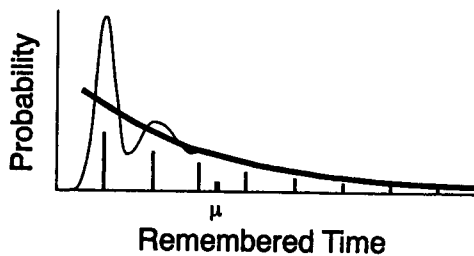


Figure 1. Probability density for a memory mixture of gaussian densities and for the exponential with the same mean (heavy-line function). The gaussian distributions have the scalar property on variance, and the mixing distribution, indicated by the spikes, is a constant-probability geometric series. (From "Scalar Expectancy Theory and Choice Between Delayed Rewards," by J. Gibbon, R. M. Church, S. Fairhurst, and A. Kacelnik, 1988, *Psychological Review*, 95, 102-114. Copyright 1988 by American Psychological Association. Adapted by permission.)

sample. That is, imagine a sample, a , from an exponential, VI:A, with rate parameter $\lambda_A = 1/A$ and a sample, b , from a comparable VI with mean B. We calculated the probability that $a < b$. For a fixed a , this is well known to be given by the "tail integral" of B, $\exp(-a\lambda_B)$, and integrating over a we have immediately

$$P(\text{"A"}) = \frac{\lambda_A}{\lambda_A + \lambda_B} = \frac{1}{1 + A/B}. \quad (1)$$

Our account thus required that choice probability between two alternatives match their relative payoff rates, but, in concert with other molar accounts, the dynamics of how such choices are made on a moment-to-moment basis was left unspecified.

On the other hand, while the dynamics of choice were not specified, it is clear from this account that the mnemonic representation is associated with the scheduled and delivered values of delays to reinforcement, not the local probability of reinforcement obtained in a concurrent choice situation. In concurrent VI schedules, both alternatives elapse toward reinforcement independently of the subject's behavior, and hence, for example, if reinforcement rates are scheduled in a 4:1 ratio across the two alternatives and responses (and times) match this 4:1 ratio, more responses, but also more reinforcers, occur on the majority choice. The result is that local probability of reinforcement on each alternative is equal. This feature is critical to "momentary maximizing" (Shimp, 1969), melioration (Herrnstein & Vaughan, 1980), and the "cumulative effects model" (Davies, Staddon, Machado, & Palmer, 1993), which rely upon an appreciation of local probability of reinforcement given a choice of either alternative (see Williams, 1994, for a review).

Local reinforcement rate does not enter into our account, and strong evidence has been adduced to argue against local probability of reinforcement as a controlling variable in choice. Williams and Royalty (1989) and Williams (1993, 1994) have shown that when stimuli

trained in a concurrent choice context are later probed in discrete choice trials against an alternative trained separately, it is the scheduled values of the concurrent stimuli, rather than the obtained local probability of reinforcement, that control choice in the probe trial. Preference appears to be determined by the scheduled values of delay to reinforcement, as SET claims, independently of the pattern of reinforced and nonreinforced responses actually experienced by the subject during concurrent choice.

A recent result by Belke (1992), however, casts doubt on the SET account as well as on other molar and molecular accounts of time matching in concurrent schedules. I describe this study in some detail because it provides the strongest challenge to extant accounts of matching. Belke's result makes abundantly clear the need for understanding the dynamics of concurrent choice.

In Belke's study, pigeons chose between two pairs of concurrent VI schedules in a four-key chamber. Two keys on one side were devoted to one pair of VIs; the remaining two were devoted to the other pair. Baseline training consisted of concurrent VI:20/VI:40 associated with white and red keys on the left and concurrent VI:40/VI:80 associated with green and yellow keys on the right (all VI designations are in seconds). Subjects were exposed to each pair of schedules for 1 min, followed by a 10-sec blackout, followed by exposure to the alternate pair, and so forth. Belke found matching to relative reinforcement rates, that is, a 2/3 time-allocation preference for white over red (twice as much time spent on white) and a 2/3 preference for green over yellow, as would be expected. Preference was expressed in response rates as well, but local rates were roughly constant in the two situations; hence, response matching was the result of time matching.

Unreinforced probe tests between red and green, however, showed a 4:1 time-allocation preference for green! Since local and scheduled probabilities of reinforcement were equal for the two keys, indifference between them would be expected according to molar matching accounts.

This kind of challenge is also damaging for the SET version of matching. If subjects sample a memory for the VI:40 in red and a memory for the VI:40 in green, these two samples are, on average, equally likely to favor either alternative and, hence, the matching level should be indifference. On the other hand, as Williams (1994) and Mark and Gallistel (1994) have pointed out, it may be that it is the probability of switching out of a schedule that is the controlling variable for preference here. Subjects may respond in the probe tests as though the alternative were the original training alternative. Nevertheless, even in this case, the preference actually obtained (4:1) is too extreme to be explained by the probabilities of switching out of a given probe schedule alone. That is, even if subjects were to preserve their training probability of switching out of red into white (2/3) and out of green into yellow (1/3), these switching probabilities

would still generate, at most, a 2:1 rather than a 4:1 preference for green. For then,

$$P(G) = \frac{P(R \rightarrow W)}{P(R \rightarrow W) + P(G \rightarrow Y)} = \frac{(2/3)}{(2/3) + (1/3)} = 2/3.$$

A qualitative account of Belke's result might, however, be ascribed to generalization of preference. Subjects in the probe tests prefer (perhaps too strongly) the schedule that was preferred in the training situation. They acted in favor of the previous majority choice, that is, green, while red was unpreferred.

I report below an experimental examination of the generalized preference question, and propose an expanded account of the dynamics of concurrent choice. The account depends importantly on subjects' maintaining, in probe tests, the rate at which they sampled from memory and made choices during training. The account adapts an important proposal by Myerson and Miezen (1980) on the kinetics of matching. These authors argued that it is the switching *rate*, not probability, out of the two signals which is the controlling variable in concurrent matching. In the following quantitative account, I elaborate on their Markov chain analysis and show experimentally that the switching rate out of, and back into, a given signal may, when properly understood, predict not only Belke's result but an even more striking case in which probes between two separately trained concurrent alternatives can result in preference for the *minority* alternative. That is, that better is worse in such probe preference tests.

Dynamics of Time Allocation

Myerson and Miezen (1980) showed that the dynamics of temporal switching between alternatives in concurrent VI schedules was well described by a two-state Markov chain with just two transitions, those from one to the other state. I adapt their account below to include four transitions. Two are hidden transitions from a state back to itself, and the remaining two are the switching transitions to the alternative. In Figure 2, the modified Markov chain is shown schematically. State A corre-

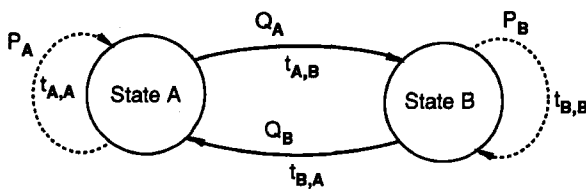
sponds to responding on VI:A, and State B corresponds to responding on VI:B. The transition from A to B is indicated with a time, $t_{A,B}$ and a probability Q_A , and similarly for the transition from B to A. What is added here are hidden transitions within a state (dotted arrows), again associated with times and probabilities. This is a stationary Markov chain, with probabilities $P_A = 1 - Q_A = Q_B = \lambda_A / (\lambda_A + \lambda_B)$, where λ_j = the mean reward rate in VI:J. That is, the transition out of B into A occurs with a probability equal to the relative reinforcement rate for A, as described above (Equation 1). The four transition times are exponentially distributed and reflect the times at which successive choices for one or the other schedule are made. Myerson and Miezen identified the between-state transitions as exponential, as expected with a two-state, two-transition Markov chain, but, more importantly, they also showed that the mean transition *rates* were proportional to the overall rate of reinforcement. No principled reason underlying this finding was given; rather, it was found as an empirical fact. I analyze this finding below as a consequence of memory sampling ideas.

Memory Sampling Rate and Preference

The major analytic result proposed here is that time allocation as a measure of preference in concurrent choice is really a composite of two factors: (1) the "true" preference for one or the other alternative based upon the memory representation of delays to food in that alternative, and (2) the rate at which that preference is accessed. It is the latter that provides a number of rather surprising predictions, including Belke's result described above. Adapting Myerson and Miezen's finding, we will argue that the average rate of choice, to either stay or switch, is proportional to the overall rate of reinforcement. This may be thought of as the effect of "arousal" in the sense espoused by Killeen and his colleagues (Killeen, Hanson, & Osborne, 1978), except that here arousal modulates the frequency of choice, rather than the frequency of a pacemaker controlling timing.

The overall rate of reward is the sum of the two independent rates, $\lambda = \lambda_A + \lambda_B$. Memory sampling governs both the hidden transitions within a state and the observed transitions between states. The mean interchoice interval, the inverse of the memory sampling rate (MSR), is proportional to $1/\lambda$, so $MSR = K\lambda$. Subjects sample from their memories and decide to either stay or switch, at this sampling rate, with $E(t_{i,j}) = \mu_{i,j} = 1/MSR = 1/K\lambda$, all $i, j = A, B$.

The MSR idea has its roots in some very early ideas on extinction, the partial reinforcement extinction effect and the "reflex reserve" (Skinner, 1938). It is exemplified in earlier work of ours on extinction of autoshaping conditioning (Gibbon, Farrell, Locurto, Duncan, & Terrace, 1980). Extinction occurs at a rate approximately proportional to the rate at which reinforcers were previously delivered. That is, if one counts "expected reinforcers omitted" in extinction, the rate of decay of an association is about the same no matter what the prior absolute rate of reinforcement. This is not a new idea, but the present



$$f(t_{i,j}) = MSR \exp(-MSR t_{i,j}), \quad i, j = A, B$$

$$E(t_{i,j}) = \mu_{i,j} = 1/MSR, \quad MSR = K\lambda.$$

Figure 2. Schematic of a stationary two-state Markov chain. The observed transitions between states A and B are shown with heavy line arrows and hidden transitions from a state back to itself with dotted line arrows. The times associated with these transitions are exponential, with means equal to the memory sampling rate, proportional to the overall rate of reinforcement (as indicated below the state diagram).

application is novel. I argue that adjustment to new levels of preference in the concurrent situation ought to depend on the previously experienced overall reward rate.

Time allocation. I now show how this mechanism would produce matching of time allocation in concurrent VI schedules. To be concrete, we use VI:20 and VI:40 in white and red, respectively, as in Belke's (1992) experiment. The overall food rate $\lambda_c = 3/40$ sec, hence on average approximately every 13 sec/ K subjects sample from memory and choose to either stay or switch, with probabilities appropriate to the matching level, that is, 2:1 in favor of VI:20. When the comparison favors white, then subjects switch to it if they are not already there, and stay if they are, and vice versa for red. This situation defines the asymptotic two-state chain, Markov in both time and space, shown in Figure 2. The fixed vector equals the transition probabilities to white and red, P_A, P_B , where $P_A = 2/3 = Q_B = 1 - P_B$.

The central feature to be added to this analysis is the dwell time (DT) in each state. The DT is the sum of the number of (hidden) transitions from a state back to itself, plus a final transition out. DTs in state J are then

$$DT(J) = \sum_{k=1}^n t_{J,J,k} + t_{J,I}, \quad J \neq I = A, B, \quad (2)$$

where $t_{J,J,k}$ is the k th transition time from state J back to itself, n is the number of hidden transitions, and $t_{J,I}$ is the final transition time out of J into I . Taking expectations on t [with $E(t_{i,j}) = 1/MSR = 1/K\lambda_c$] and on n [with $P(n = N) = (P_c)^N Q_c$], we have expected dwell time, EDT, given by

$$EDT(J) = \frac{1}{K\lambda_c Q_J}. \quad (3)$$

An immediate result is that the relative-time-allocation measure induces matching:

$$\begin{aligned} \frac{EDT(A)}{EDT(A) + EDT(B)} &= \frac{K\lambda_c Q_B}{K\lambda_c Q_A + K\lambda_c Q_B} \\ &= P_A = \frac{\lambda_A}{\lambda_A + \lambda_B}. \end{aligned} \quad (4)$$

Exponential dwell times. A number of results, beginning with Heyman (1979) and extending through some recent work of the Myerson group (Buckner, Green, & Myerson, 1993; Myerson & Hale, 1988; Real, 1983; Real & Dreyfus, 1985), confirm the exponential character of the DTs. Our analysis may be seen to encompass this result as follows: Assuming that the intersampling times, $t_{i,j}$, are themselves exponentially distributed, a pleasing result from stochastic process analysis (cf. McGill, 1963) applies. A sum of exponentially distributed times is well known to be gamma distributed. However, when a sum of exponentials is ended with a random stopping rule, that is, with some probability, the distribution of the sum returns to exponential form. Hence, from Equation 2, the dis-

tribution of DTs is itself exponential, with rate parameter MSR, weighted by the exit probability, Q (Equation 3).

Probe preference. The DT analysis immediately gives a prediction of a 4:1 rather than a 2:1 preference for the VI:40 that was paired with the VI:80 over the VI:40 paired with the VI:20 in Belke's experiment. Overall reinforcement rate in the white (VI:20), red (VI:40) pair is $\lambda_c(W,R) = 3/40$ sec. The overall reinforcement rate for the leaner pair is $\lambda_c(G,Y) = 3/80$ sec. The corresponding memory sampling rates are proportional to these values. EDT, then, for red in the 20/40 concurrent is

$$EDT(R) = \frac{1}{K\lambda_c(W,R) Q_R} = 20 \text{ sec}/K.$$

Similarly, the EDT in green is

$$EDT(G) = \frac{1}{K\lambda_c(G,Y) Q_G} = 80 \text{ sec}/K.$$

Hence, relative time allocation between green and red is 4:1 in favor of green, as Belke found. This is essentially the argument advanced by Mark and Gallistel (1994), but without the stochastic Markov analysis and assuming, rather than deriving, matching.

The argument may be summarized qualitatively as follows: (1) The time-allocation preference between two schedules is determined by the ratio of their DTs, that is, the time spent in the presence of each signal. (2) Distributions of DTs are determined by the probability of switching or remaining on a given alternative multiplied by the time between decision epochs. That is, DTs (Equation 2) are composed of the number of hidden transitions from a state back to itself (with probability P) plus the final transition to the other alternative (with probability Q). (3) The probability of staying or switching is determined by sampling from the distribution of remembered times for each schedule and choosing that with the shorter time. This results in choice proportions that obey the matching law. (4) The time between decisions is determined by overall reinforcement rate from both choice alternatives in training. (5) In the probe tests, subjects ignore the new alternative and, instead, maintain the switching rates established in training. The result is that a choice alternative with a rich schedule may nevertheless have a relatively short dwell time, because a high memory sampling rate can reduce the time spent on that alternative. In the Belke experiment, the memory sampling rate in the VI:20/VI:40 pair was twice that for the VI:40/VI:80 pair, resulting in exits that were more rapid from the VI:40 trained with the high reinforcement rate than from the VI:40 trained with the lower reinforcement rate. We will see below that an even more striking anomalous preference is entailed by these ideas.

BETTER IS WORSE: AN EXPERIMENT

The experiment reported below replicates Belke's finding in a two-key procedure with a changeover (CO) re-

sponse key and a main key (Findley, 1958; Findley concurrent). Two pairs of concurrent schedules, as in Belke's procedure, were trained during 1-min intervals with brief time-out periods between them. The main key was either white (VI:20) or red (VI:40) for one pair and either green (VI:40) or yellow (VI:80) for the other. After training, unreinforced probe tests were introduced. One probe test replicated Belke's procedure. The main key in the probe was either red (VI:40 from the white/red pair) or green (VI:40 from the green/yellow pair). CO responses switched the main key between red and green during the probe. The second, critical, probe test was conducted between white and green. Here white was the richer schedule and green, the leaner schedule; if subjects do not adjust their dwell times in the probe tests, however, time allocation should favor green 2:1 even though its relative density of reinforcement is half that of white! That is, dwell time in green = 80 sec/K, while dwell time in white = 40 sec/K (twice that in red, as shown above), and so relative time allocation should favor green 2:1, because here memory is sampled at a lower rate.

Method

Subjects. The subjects were 6 white Carneaux pigeons that had previously been used in timing tasks unrelated to concurrent choice.

Apparatus. The subjects were trained in a two-key pigeon chamber enclosed in a sound-attenuating box. Further acoustical isolation was provided by low-volume white noise broadcast through a speaker in the chamber. Each key was located 21.5 cm above the chamber floor and was transilluminated by an IEE projector. A 5 × 5 cm aperture centered 10 cm above the floor and located symmetrically between the two keys provided access to a solenoid-operated grain hopper. A dim houselight provided general illumination, except during reinforcement, when it was replaced by a light over the hopper. The procedure was implemented and data collected by a computer located in an adjacent room.

Procedure. The subjects were placed directly on the changeover-key concurrent schedules of VI:20/VI:40 or VI:40/VI:80 for 5 weeks of one session per day, consisting of 30 1-min periods during which one or the other concurrent pair was in force. These periods were followed by a 10-sec time-out, as in Belke's procedure, but successive pairs were chosen randomly after each time-out. Thus, on average, 15 1-min periods were devoted to the VI:20/VI:40 pair and 15 were devoted to the VI:40/VI:80 pair. During the 1-min periods, the main key was illuminated either white or red (chosen randomly at the start of each 1-min period) or green or yellow, and responses to the CO key changed the main key color to its alternate. A changeover delay (COD) of 2 sec was in effect such that a reinforcer set up for a given alternative could not be delivered until 2 sec had elapsed after a switch. Three of the 6 subjects were assigned white and red for VI:20 and VI:40 and green and yellow for VI:40 and VI:80, respectively. The other three were assigned red and white for VI:20 and VI:40 and green and yellow for VI:40 and VI:80.

Probe testing. Probe testing was then introduced for 2½ weeks, with probe days occurring variably but, on average, every 3rd day. Baseline training continued in the remaining sessions. On probe days, 10 of the 30 1-min periods were probe periods. During probe periods, no reinforcement was scheduled. In the probe trials of the first four probe sessions, the main key was illuminated white or green for all subjects. This meant that, for 3 of the subjects, the probe test was between the VI:20 of the previously trained VI:20/VI:40 pair and VI:40 of the previously trained VI:40/VI:80 pair. For the other 3 subjects, the probe test was between VI:40 of the 20/40 pair and VI:40 of the 40/80 pair (Belke replication). After four probe

sessions, which occurred in an average of 12 days, another four probe sessions were run between the red and green stimuli. For the subjects previously probed between the two majority choices, 20 versus 40, the second set of probes examined 40 versus 40 (Belke replication). For the other subjects, previously probed on the Belke replication, the second set of probes was between the two majority choices. In total, probe testing lasted approximately 5 weeks, four sessions of the first probe test interspersed with eight or nine sessions of baseline, followed by four sessions of the second probe test interspersed with eight or nine sessions of baseline.

Results

The proportion of time spent on the richer of the two training schedules during the baseline trials (taken from the probe testing phases) is shown in the left two histogram bars in Figure 3. Birds slightly undermatched the proportion of reinforcements (.67) in both pairs. For the 20/40 pair, there was no significant difference at the .05 level (used throughout) between matching and the level achieved. For the 40/80 pair, undermatching was reliable. Time allocation for both pairs was reliably different from indifference. Thus, birds matched, or slightly undermatched, relative reinforcement rate.

Time allocation to the majority choice during the probe tests is shown in the two histogram bars on the right. The first bar replicates Belke's result, with the VI:40 from the 40/80 pair preferred about 4:1 over the VI:40 from the 20/40 pair. The matching level (indifference) is shown for reference. This proportion was significantly different from .5 and not different from .8.

The critical observation is shown in the far right bar when the probe was between VI:40 from the 40/80 pair and VI:20 from the 20/40 pair. Here the matching level based

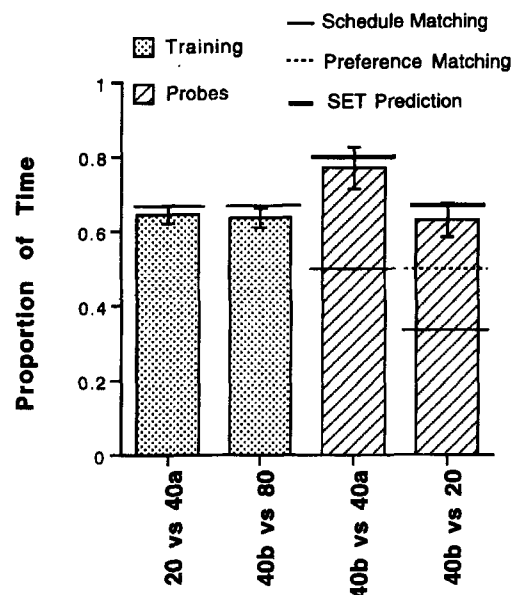


Figure 3. Relative time allocation to the majority choice in baseline (left two bars) and probe testing (right two bars). Matching levels are indicated with light horizontal lines and the scalar expectancy theory predictions with heavy horizontal lines. The generalized preference prediction is indicated with a dashed horizontal line.

on scheduled reinforcers would be a 2:1 preference for the VI:20, while the obtained level is about a 2:1 preference for VI:40! Preference is slightly below 2:1, but the result is not significantly different from .67, while it is different from .5 (and, hence, from the matching level of .33).

These surprising preferences are a direct result of unchanging DT distributions between training and probe tests. Figure 4 shows DT distributions averaged over these 6 subjects in semilog form. The straight lines in the figure are regression fits to data from training and probe trials restricted to intervals greater than 3 sec and less than 23 sec. This range was chosen because all subjects show a short rise in relative frequency, to a peak at about second 4 or 5, and a subsequent linear decline. The upper limit was chosen to exclude any zeros in the 1-sec relative frequency bins. The short rise may reflect a tendency to stay at an alternative beyond the COD (2 sec). Over the

middle range, linearity is well approximated for all four schedules. Linearity in semilog coordinates means that the distributions beyond the peak are exponential, replicating previous findings of exponentially distributed dwell times. The close correspondence between slopes of regressions on the training and probe data mean that subjects are performing similarly in training and testing.

The two distributions on the left are for the 20/40 training pair, and those on the right are for the 40/80 training pair. Notice that the slopes on the left are steeper than those on the right and that, in particular, the slopes for the 40-sec paired with the 80-sec schedule (40b) in the upper right are considerably shallower than those for the 40-sec paired with the 20-sec VI (40a). That is, exits from 40b occur at a much lower rate than do exits from 40a, and hence the strong preference for 40b in the probe test. Similarly, exits from 40b occur more slowly than

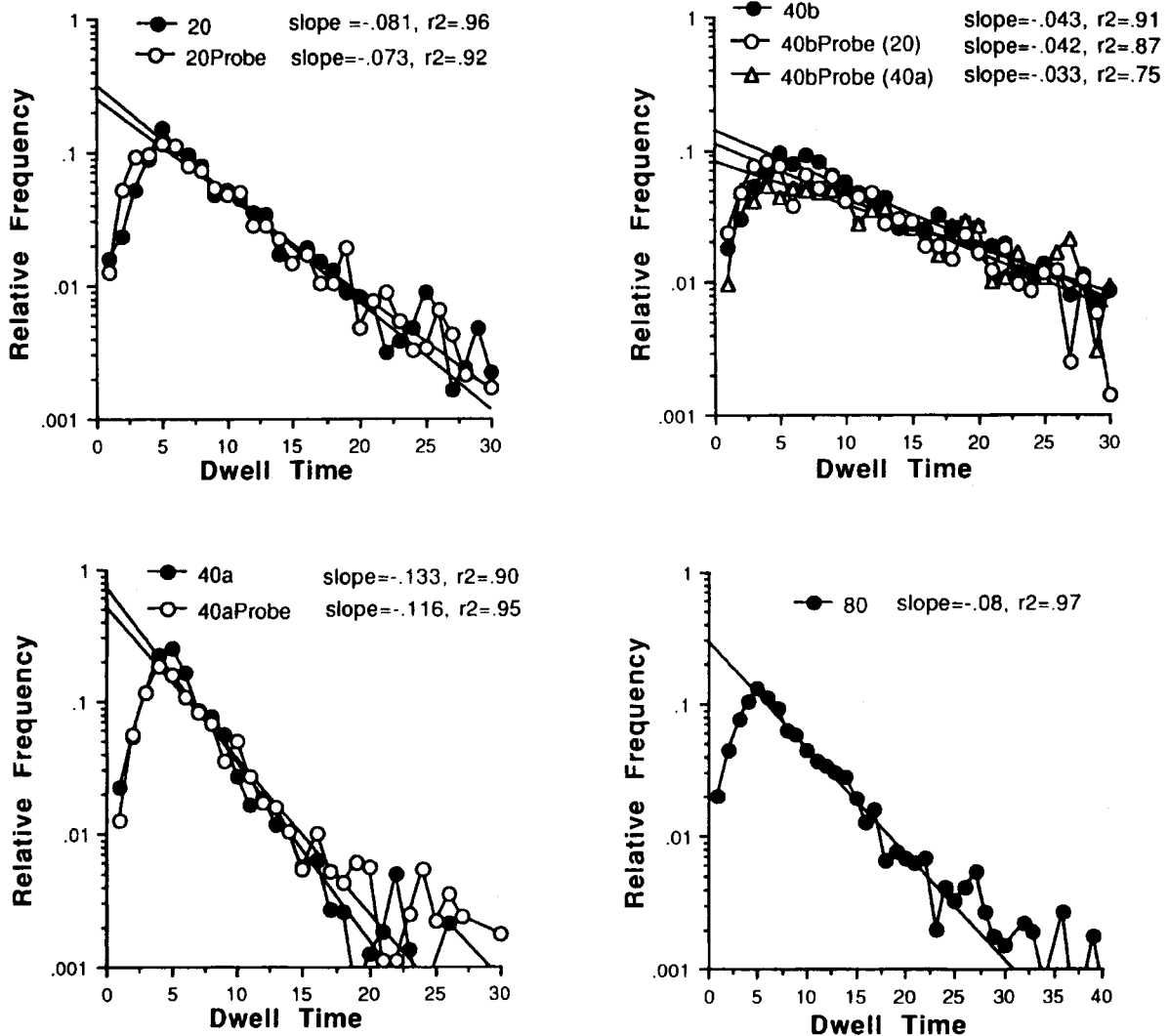


Figure 4. Dwell-time distributions averaged over 6 subjects for each schedule and probe sessions with that schedule. Relative frequency is on a log scale. The 20/40 training pair is shown on the left and the 40/80 pair on the right. Regression fits to relative frequencies between 3 and 23 sec are shown with straight lines. Note the similar slopes for probe and training distributions.

exits from 20, and hence the surprising preference for 40b in this probe test as well.

DISCUSSION

The novel preference for 40b over 20 does not fit with a generalization-of-preference account which might be put forward to accommodate the Belke result. Both of the two probe stimuli were favored 2:1 in the training pairs. Hence, a generalization account would predict, at best, indifference between them. In contrast, the data show that subjects spend about one-third of their time on the richer schedule. That is, better is worse.

These results, and the preceding analysis, implicate a two-component time-allocation system in a stationary Markov chain. One component is a decision process for deciding between two training alternatives. This process, I argue, is well described as sampling from exponential memories and simply choosing the better of the two samples. The second component is the memory sampling rate. Memories are sampled and choices are made at exponentially distributed times, with a mean rate proportional to the overall rate of reinforcement in the training context, that is, proportional to arousal levels. When this rate of reinforcement is high, as in the richer pair in this experiment, sampling occurs faster than when it is low, as in the leaner pair. This results in matching when the choice is between the two training schedules, since they are trained in the same overall reinforcement context. But when the schedules are probed with an alternative trained in a different context, subjects do not adjust their sampling rate. They are slower to decide in the probe signal from the leaner pair than they are in the probe signal from the richer pair. This results in the surprising time-allocation "preference" for the leaner of the two probe schedules when both were preferred in training.

The stochastic properties of this system have interesting implications as well—unappreciated heretofore, perhaps because the hidden transitions were not considered. For example, Buckner et al. (1993) argued that the occurrence of reinforcement on a side lengthened the stay on that side, because their data show that DTs that include a reinforcer average about twice as long as those without reinforcers. But a beautiful result, due to Feller (1966), on two independent Poisson processes shows that this is just what is to be expected if the reinforcement delivery system is independent, moment to moment, from the stochastic switching rates. Conceptually, Feller's result says that DTs that are "lucky enough," so to speak, to "catch" a reinforcement are on average twice the length of other DTs, hence the above finding. Notice that the time between reinforcement and exit from a side then remains at about the overall average DT, as does the time from entrance to reinforcement, as would be expected on a strictly independent basis. Thus, we argue that it is the stochastic properties of these systems that produce the extended DTs which contain reinforcers, not a local effect of reinforcement delivery.

I have argued that the exponential character of DTs is a consequence of exponential memory sampling. However, one might simply assume exponential DTs with the appropriate means. Then the present unusual "preference" for the leaner of two equally preferred schedules also may be predicted by the switching-rate constancies of Myerson and Miezen (1980) and Mark and Gallistel (1994), without recourse to the deeper analysis provided here of how such dwell times are produced. The present account unifies a decision rule that does not itself specify matching as an outcome, with the molar matching result and its exponential DT character, by incorporating memory-driven hidden transitions from one schedule back to itself.

The implications of the memory-sampling-rate analysis are potentially far reaching, especially if the hidden transitions may be experimentally unmasked. It is noteworthy, in this context, that matching is also obtained in discrete trial procedures (e.g., Nevin, 1969) in which there are no hidden transitions and no switching rates, because choice opportunities are experimentally controlled. Williams and Royalty (1989, Experiments 2 and 3) and Williams (1993) also studied discrete-trial preference tests of choice between signals trained in alternative contexts. Their results are in qualitative agreement with control by the scheduled reinforcement probabilities in training. Choice in discrete trials should force decisions based on memory for the trained interreward intervals, consonant with the decision process espoused here and independent of the rate at which subjects would make such choices in free-operant situations.

While arousal apparently controls the rate at which animals sample from their memory to make choices, it is not yet clear what kind of time window for storing interreward intervals into memory is used. Mark and Gallistel (1994) have shown that under some circumstances extremely rapid adjustments to fluctuations in reward rate occur during concurrent choice. In the present context, this suggests that storage into memory may be sensitive to fluctuating payoff rates. It may be shown that the present account of choice does not require storage of more than a few, say three to five, interreinforcement intervals to generate matching. Under these conditions, local time allocation tracks local fluctuations in reward rate. It seems likely that a fluctuating environment may predispose subjects to remember a small number of delays, while stable environments may encourage a larger memory store.

An important next step will be to examine whether the memory-sampling-rate idea applies to probes between signals for exclusive alternatives as opposed to concurrent signals for the schedules themselves. The former, studied in concurrent-chain procedures, may be analogous to patch choice in foraging animals (Shettleworth, 1988). It is an open question as to whether memory sampling obeys the same rules when subjects are choosing between patches, analogous to travel decisions of foragers to approach one or another resource.

REFERENCES

- BAUM, W. M., & RACHLIN, H. C. (1969). Choice as time allocation. *Journal of the Experimental Analysis of Behavior*, **12**, 861-874.
- BELKE, T. W. (1992). Stimulus preference and the transitivity of preference. *Animal Learning & Behavior*, **20**, 401-406.
- BUCKNER, R. L., GREEN, L., & MYERSON, J. (1993). Short-term and long-term effects of reinforcers on choice. *Journal of the Experimental Analysis of Behavior*, **59**, 293-307.
- DAVIES, D. G. S., STADDON, J. E. R., MACHADO, A., & PALMER, R. G. (1993). The process of recurrent choice. *Psychological Review*, **100**, 320-341.
- FELLER, W. (1966). *An introduction to probability theory and its applications* (Vol. 2). New York: Wiley.
- FINDLEY, J. D. (1958). Preference and switching under concurrent scheduling. *Journal of the Experimental Analysis of Behavior*, **1**, 123-144.
- GALLISTEL, C. R. (1990). *The organization of learning*. Cambridge, MA: MIT Press.
- GIBBON, J., CHURCH, R. M., FAIRHURST, S., & KACELNIK, A. (1988). Scalar expectancy theory and choice between delayed rewards. *Psychological Review*, **95**, 102-114.
- GIBBON, J., FARRELL, L., LOCURTO, C. M., DUNCAN, H. J., & TERRACE, H. S. (1980). Partial reinforcement in autoshaping with pigeons. *Animal Learning & Behavior*, **8**, 45-59.
- HERRNSTEIN, R. J. (1970). On the law of effect. *Journal of the Experimental Analysis of Behavior*, **13**, 243-266.
- HERRNSTEIN, R. J., & VAUGHAN, W. (1980). Melioration and behavioral allocation. In J. E. R. Staddon (Ed.), *Limits to action: The allocation of individual behavior* (pp. 143-176). San Diego: Academic Press.
- HEYMAN, G. M. (1979). A Markov model description of changeover probabilities on concurrent variable-interval schedules. *Journal of the Experimental Analysis of Behavior*, **43**, 313-322.
- KILLEEN, P. R., HANSON, S. J., & OSBORNE, S. R. (1978). Arousal: Its genesis and manifestation as response rate. *Psychological Review*, **85**, 571-581.
- MARK, T. A., & GALLISTEL, C. R. (1994). Kinetics of matching. *Journal of Experimental Psychology: Animal Behavior Processes*, **20**, 79-95.
- MCGILL, W. J. (1963). Stochastic latency mechanisms. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 1, pp. 309-360). New York: Wiley.
- MYERSON, J., & HALE, S. (1988). Choice in transition: A comparison of melioration and the kinetic model. *Journal of the Experimental Analysis of Behavior*, **49**, 291-302.
- MYERSON, J., & MIEZEN, F. M. (1980). The kinetics of choice: An operant systems analysis. *Psychological Review*, **87**, 160-174.
- NEVIN, J. A. (1969). Interval reinforcement of choice behavior in discrete trials. *Journal of the Experimental Analysis of Behavior*, **12**, 875-885.
- REAL, P. G. (1983). A time series analysis of changeover performance on concurrent variable-interval schedules. *Animal Learning & Behavior*, **11**, 255-265.
- REAL, P. G., & DREYFUS, L. R. (1985). Levels of aggregation: Relative time allocation in concurrent-schedule performance. *Journal of the Experimental Analysis of Behavior*, **43**, 97-113.
- SHETTLEWORTH, S. J. (1988). Foraging as operant behavior and operant behavior as foraging: What have we learned? In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 22, pp. 1-49). New York: Academic Press.
- SHIMP, C. P. (1969). Optimal behavior in free operant experiments. *Psychological Review*, **76**, 97-112.
- SKINNER, B. F. (1938). *The behavior of organisms*. New York: Appleton Century.
- VAUGHAN, W., JR. (1985). Choice: A local analysis. *Journal of the Experimental Analysis of Behavior*, **43**, 383-406.
- WILLIAMS, B. A. (1988). Reinforcement, choice and response strength. In R. C. Atkinson, R. J. Herrnstein, G. L. Lindzey, & R. D. Luce (Eds.), *Stevens' Handbook of experimental psychology* (2nd ed., pp. 167-244). New York: Wiley.
- WILLIAMS, B. A. (1993). Molar versus local reinforcement probability as determinants of stimulus value. *Journal of the Experimental Analysis of Behavior*, **59**, 163-172.
- WILLIAMS, B. A. (1994). The role of probability in models of choice. *Psychological Review*, **101**, 704-707.
- WILLIAMS, B. A., & ROYALTY, P. (1989). A test of the melioration theory of matching. *Journal of Experimental Psychology: Animal Behavior Processes*, **15**, 99-113.

(Manuscript received August 22, 1994;
revision accepted for publication December 6, 1994.)