Learning and Motivation **33,** 46–62 (2002)
doi:10.1006/lmot.2001.1099, available online at http://www.idealibrary.com on IDEAL®

A Test of Gibbon's Feedforward Model of Matching

C. R. Gallistel and Terence A. Mark

University of California, Los Angeles

Adam King

Department of Computer Science, Fairfield University

and

Peter Latham

Department of Neurobiology, University of California, Los Angeles

Gibbon (1995) elaborated an ingenious model of matching, a feedforward model that is consistent with Heyman's (1982) suggestion that matching behavior does not depend on selection by consequences. Most models (for example, Herrnstein & Vaughan, 1980) have been feedback models, built on the law of effect. Measurements of how rapidly rats adjust to changes in the relative rates of brain stimulation reward on concurrent random interval schedules imply a feedforward process. The adjustments are, however, too fast to be consistent with Gibbon's model. © 2002 Elsevier Science (USA)

John Gibbon pioneered the psychophysical study of interval timing and the application of information-processing models to our understanding of conditioned behavior. Among his many, highly original contributions was a model of matching behavior (Gibbon, 1995), which differed in a fundamental way from previous models. The difference has potentially far reaching implications for our understanding of instrumentally conditioned behavior. Unlike most previous models, Gibbon's model does not assume that the consequences of previous responses feed back to affect the relative strengths of competing behaviors (for a review of models of this type, see Lea & Dow, 1984). Gibbon's model is a purely feedforward model. The experience of different intervals between rewards elicits stay durations inversely proportionate to the ratio of those intervals, without regard to the effect that the animal's behavior has on those intervals.

The law of effect ought to apply with exceptional directness when subjects are given a matching protocol. Thorndike (1911, p. 244) wrote ''The Law of Effect is that: Of several responses made to the same situation, those

0023-9690/02 $35.00

which are accompanied or closely followed by satisfaction to the animal will, other things being equal, be more firmly connected with the situation, so that when it recurs, they will be more likely to recur . . .'' A more contemporary statement of the law comes from Schmajuk (1997, p. 149): ''During operant conditioning, animals learn by trial and error from feedback that evaluates their behavior but does not indicate the correct response.'' In the matching protocol, the subject is offered two response options—most typically, two different manipulanda at two different locations. Responses on the two manipulanda are reinforced on concurrent random interval (RI) schedules. A RI schedule makes the next reward available at exponentially distributed latencies (schedule intervals) following the harvesting (collection) of the previous reward. Once scheduled, a reward remains available until it is harvested by the first subsequent response on the given manipulandum. The parameter of a RI schedule is the expected (average) interval between the harvesting of a reward and the scheduling of the next. Typically, this is shorter for one option than for the other. The shorter this expected interval is, the sooner, on average, responding on that manipulandum will be rewarded. Thus, in a matching experiment, the subject has two response options, and—other things being equal—one of them is rewarded sooner and more frequently than the other.

As Thorndike's formulation predicts, the response rewarded at shorter intervals emerges as the stronger of the two responses in that it occurs more frequently. The question is, does this come about through a process that makes responses more or less firmly connected to situations according as they are more or less likely to yield satisfaction? Or, is this the result of a decision process that translates the experienced temporal distribution of rewards into expected stay durations without regard to the relation between the animal's behavior and reward? On the first view, the animal acts in the world and observes the consequences of its acting in order to choose in the future those actions that yield the greatest satisfaction. On the second view, the animal observes the distribution of rewards in space and time, then chooses its actions without regard to the satisfactions or lack thereof that its previous actions have produced.

MELIORATION: A REPRESENTATIVE LAW-OF-EFFECT MODEL

Herrnstein's melioration model (Herrnstein, 1982; Herrnstein & Prelec, 1991; Herrnstein & Vaughan, 1980) is representative of models that take the law of effect as their point of departure in explaining matching behavior. In this model, the subject is assumed to monitor the average time or number of responses that it *invests* in each option for each reward earned. If the number of responses required to earn a reward from one option is on average fewer than the number required to earn a reward from the other, more responses are allotted to the first option and fewer to the second. Thus, for example, if, a pigeon makes on average 15 pecks to a green key between

one reward and the next and spends on average 20 s pecking at that key between rewards, its investment per reward, when measured in responses, is 15 responses/per reward; measured in time, it is 20 s per reward. The reciprocals of these numbers—amount of reward/response or amount of reward/unit time invested—are what economists call *returns.* The melioration model assumes that when two response options yield different returns, the response that yields the higher return gets stronger and the response that yields the lower return gets weaker.

When rewards are delivered on concurrent random interval schedules, the intervals between rewards are primarily determined by the delays imposed by the schedule rather than by the subject's responding, because subjects shift back and forth between the two options—in our case between two levers on opposing sides of a box—at intervals substantially shorter than the average of the scheduled delays. Under these circumstances, increasing the investment on one side (that is, the average stay on that side and hence the average number of lever presses on that side per unit of session time) and decreasing the investment on the other has little effect on the number of rewards that the subject obtains from the two levers in the course of a session (Heyman, 1982). Put another way, changes in the expected stay durations have little effect on expected *income.* In economics, the income from an investment is the amount of reward that the investment yields per unit of time—not per unit of time or effort invested, but simply per unit of time. Thus, if reward magnitude is assumed to be constant, the income that a subject obtains from pressing a lever is the number of rewards it gets from that lever per minute of session time, regardless of how much or how little time the subject spends pressing that lever, that is, regardless of how many or how few responses it made to obtain those rewards. (Investments measured in responses and investments measured in time spent responding are so closely correlated that they may be treated as interchangeable (see Baum & Rachlin, 1969).)

By contrast, changes in the expected stay durations have a strong effect on returns. The expected (average) return from a response (or from a unit of time invested in a response option) is approximately inversely proportional to the average stay duration. Thus, for example, doubling the average duration of stays on the richer side and halving it on the poorer side approximately halves the return from the richer side and doubles the return from the poorer side (while having very little effect on the incomes from the two sides). Return, which is also called expected value, quantifies the relation between behavior and its consequences, whereas income specifies what the animal has obtained without regard to its investment (how much it did).

The inverse relation between investment and return is the key to the melioration model's explanation of matching behavior. As the investment in the richer option goes up, but the income realized remains almost constant, so the return from that option goes down. Similarly for the poorer option: as

the investment declines, the income remains almost constant, hence the return goes up. The adjustment of relative investments continues until a ratio of investments is reached that equates the returns. This *equilibrium point* is reached when the ratios between the average stay durations at the two locations matches the ratio of the average incomes. In models where behavior is driven by its consequences (see, for example, Montague, Dayan, & Sejnowski, 1996; Schultz, Dayan, & Montague, 1997; Sutton & Barto, 1998), it is the returns that matter. When the subject matches its investment ratio to the income ratio, the expected values of the responses are equal. Inequalities in the returns drive the behavioral process to this equilibrium point.

THE GIBBON MODEL

In the Gibbon model what matters are the incomes not the returns. The subject remembers the intervals between rewards at each foraging location. The interval from one reward to the next divided into the magnitude of the reward gives an income datum for that location. In the typical matching experiment, reward magnitude does not vary, so remembering the income data is equivalent to remembering the interreward intervals, which are proportional to the reciprocals of the incomes. (Reward magnitude is the constant of proportionality.) By visiting the two locations and responding on the two manipulanda, the subject obtains two populations of remembered intervals. In deciding to stay at a given location or leave it, the subject continually draws a pair of samples, one from each of these two populations. After each sampling, it chooses to visit (or to continue visiting) the location associated with the shorter sample.

The odds that a sample from one population of exponentially distributed intervals will be shorter than a sample from another such population are the inverse of the ratio of the expectations (Rachlin, Logue, Gibbon, & Frankel, 1986). Thus, for example, if the average interval between rewards in one population is half as long as the average interval in the other, then the odds are 2:1 that a sample from the first population will be shorter than the sample from the second. In that case, the probability that after any one sampling the subject will decide to leave the richer location to visit the poorer is half the probability that it will decide to leave the poorer location to visit the richer. The expected durations of the stays at the richer locations will be twice the expected durations of the stays at the poorer location, because the expected duration of a stay is the sampling interval times the reciprocal of the probability that a sampling results in the decision to leave a location. If, for example, the subject samples once per second and the probability that a sample will cause it to leave is 1 in 4, then the expected duration of its stays is 4 s.

In the limit, income is sensitive to behavior, because the interval between successive rewards experienced at a location cannot be shorter than the interval between the termination of the last visit and the beginning of the next. However, in feedforward models, the process that generates behavior is not

sensitive to the dependence of income on behavior. Feedforward models are predicated on the implicit assumption that the animal's behavior is unlikely to affect how rewards are distributed in space and time. Insofar as this has been true during the evolution of the mechanisms that determine behavior, it is better to base behavior on the observed distributions of interreward intervals rather than on the observed returns, because returns are inherently noisier than incomes. The variability in returns is the result of the variability in the temporal distribution of rewards and the variability in the subject's sampling of that distribution. Thus, if a subject's behavior generally has no effect on whether a reward is or is not available to it—and if the manner in which it samples the world does not systematically distort what it observes—then it is better to base behavior simply on what has been observed, without regard to whatever effect the subject's behavior may have had on what it observed.

DISTINGUISHING BETWEEN FEEDBACK AND FEEDFORWARD MODELS

Models based on the law of effect are feedback models. The subject discovers the behavior that yields the greatest return by varying its behavior and assessing the resulting variation in returns. When the expected delays of reward change, the time that it takes the process to adjust to the new expectations cannot be shorter than some multiple of the time that it takes for a change in behavior to become manifest in a change in the returns. The subject must first discover that its returns are no longer equal. Then, it must discover by trial and error the ratio of response strengths (investments) that equates returns (expected values) in the new situation. The process of discovering by trial and error a critical point in a space defined by behavioral (output) parameters is called hill climbing in the computer science literature. Models based on the law of effect are hill climbing models. The need to repeatedly observe the effects of repeated changes in one's behavior limits the speed with which the hill can be climbed.

In the Gibbon model, matching behavior is not the result of a hill-climbing process. There is no need to repeatedly observe the effects of repeated changes in the parameters of behavior. The adjustment to a change in the relative rates of reward takes no longer than the time it takes to replace the prechange populations of remembered intervals with remembered intervals that come entirely from the period after the change in the programmed rates of reward. How long that takes depends on how large the populations are from which the subject samples, a question we will return to in a later section. For the moment, suffice it to note that, in principle at least, a feedforward system can adjust to changes more rapidly than a feedback system, because there is no need to determine the effects of intermediate changes in output en route to the final output state. Reflexes (feedforward behavioral mechanisms) respond to changes faster than servomechanisms (mechanisms that employ feedback).

Heyman (1982) formulated the distinction we are after here as the distinction between conditioned and unconditioned behavior, when he suggested that matching was unconditioned behavior. By ''conditioned'' he meant operantly conditioned and by unconditioned he meant ''innate.'' For an unconditioned behavior to occur, it suffices simply for an animal to experience a situation, because the response to that situation is innate. For a conditioned behavior to occur, the animal must not only experience a situation, it must in addition discover by trial and error which of its behaviors pays off more often in that situation. The emergence of a stable conditioned response to a situation necessarily takes longer than the emergence of an (equivalent) unconditioned response to that same situation, because the former requires more extensive experience. A corollary of this is that for an operantly conditioned response to change from one stable value to another in response to a change in situation, the animal must experience the consequences of changes in its own behavior within the new situation. By contrast, an innate response changes to a new stable form as soon as the animal detects the change in situations, before it has the opportunity to experience the consequences of changes in its behavior in the new situation. This is a consequence of the point that Schmajuk emphasized in the above quote: in operant conditioning, the consequences of behavior evaluate the responses that produced those consequences; they do not ''indicate the correct response.''

THE RAT APPROXIMATES AN IDEAL DETECTOR OF CHANGES IN RATES OF REWARD

We have recently reported the results of an experiment that determined how long it takes rats to adjust their stay duration ratios to unpredictable complementary step changes in the programmed rates of brain stimulation reward (Gallistel, Mark, King, & Latham, 2001). One phase of the experiment lasted for 20 daily 2-h-long sessions. In this phase, rats experienced two changes in the relative rates of scheduled rewards per session. The direction and magnitude of each change was unpredictable, although the two rates always changed in such a way as to preserve the overall rate of reward (the sum of the two rates). One change always occurred between the end of the previous session and the beginning of the next. The timing of the second change was not predictable; it occurred at a randomly chosen moment in the middle 80 min of each 120-min session. The intervals in the RI schedules were produced by an electronically realized random rate process, which generated different sequences of intervals every time it was called. The next intervals in a sequence with a given expectation could not be anticipated from knowledge of the preceding intervals, no matter how often the subject had experienced a schedule with that expectation. Moreover, a given interval could occur in schedules with widely differing expectations, so the experience of a single interval could never specify which schedule was in force. However, the experience of a single interval could, in some cases, suffice

to indicate with high probability that a schedule with a substantially lower or higher expectation had supplanted the schedule previously in force. The question was whether subjects were sensitive to these probabilities.

We found that the subjects were remarkably sensitive to the probabilistic implications of the interreward intervals they experienced. They adjusted their expected stay durations to new rates of reward about as rapidly as is in principle possible. That is, our subjects behaved like ideal detectors of changes in rates of reward; no physically realizable device could have adjusted to the new rates much faster than our subjects adjusted. This means that there was no delay between the detection of the change in situation and their adjustment to that change. They made the adjustment before there was time for them to assess the consequences of changes in their behavior within the new situation. Thus, the change they made could not have been a consequence of such an assessment.

To reach this conclusion, we had to develop a suitable representation of the behavior and of the inputs that might drive it—the experienced incomes and the experienced returns. The expected incomes (the inputs) and the expected stay durations (the outputs) in a matching protocol cannot be known exactly. They must be represented by probability density functions. These functions quantify the objective (inescapable) uncertainty about the true expectation of a random rate process that has only been observed for a finite interval. They give for every rate the relative likelihood that it is the expected rate.

The longer we observe the output of a constant random rate process, the more precisely we can estimate its expectation, that is the average number of events it produces per unit of time. (This statement is a version of the law of large numbers; the longer you observe, the bigger the number of events; hence, the more precisely the average is known.) Thus, the probability density function representing the rate of reward on a given side will show a higher and narrower peak as a session progresses, so long as that rate remains constant. This analytically necessary trend is evident in the pre-change peaks of the probability density functions for the reward rates in Figs. 1 and 2 (left). If the subject believes that rate remains constant within a session, then its estimates of the expected rates of reward at any point in the session will be averages over the populations of intervals observed up to that point. In such a case, its probability density functions at any point in the session after the change would be flatter than before the change, because all post-change estimates of the prevailing rate of reward would be based on two different populations of interreward intervals, one generated by the schedule before the change in rate and one by the schedule after the change. The averaging across observed interreward intervals from before and after the change in rates of reward will slow down the emergence of an accurate estimate of the post-change rates, which is not ideal.

An ideal detector of changes takes into account the possibility that there

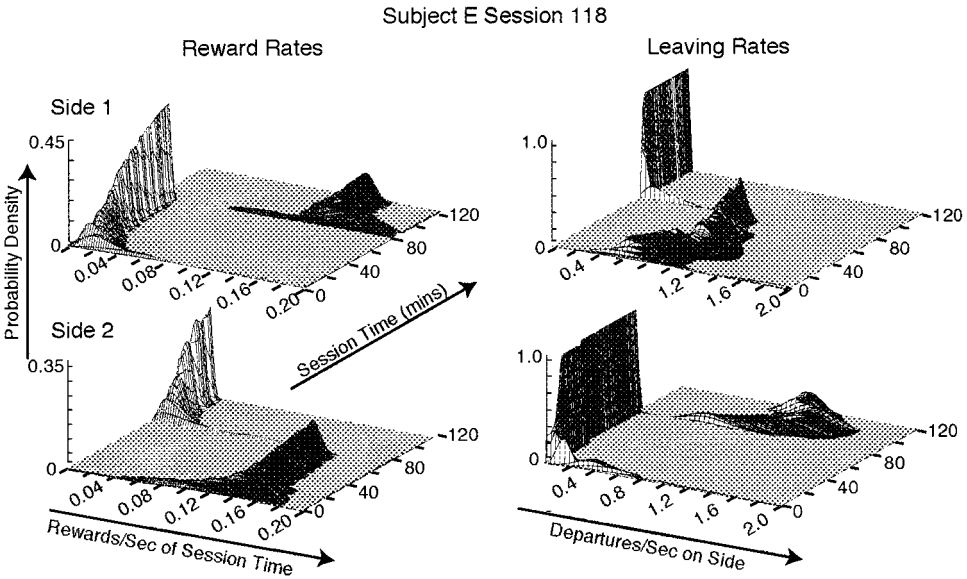

FIG. 1. (Left) Probability density (*y* axis) as a function of rate of reward (*x* axis) at successive session times (*z* axis) as calculated from our model of an ideal detector of changes in rates of reward. (Right) Probability density (*y* axis) as a function of leaving rate (*x* axis) at successive session times (*z* axis). The probability density functions were computed by a Bayesian formula that assumed one and only one change in the rates of reward and leaving rates in the course of each session (see Appendix to Gallistel *et al.*, 2001). Notice that the changes in the probability density functions for the expected stay durations are as abrupt as those for the (known-to-be) step changes in the rates of reward.

has been a change. Instead of averaging over all the intervals since the beginning of a session, it estimates whether the change in rate has occurred, and what the rate now is. In developing a mathematical model of an ideal detector, we used a Bayesian approach to build in this a priori expectation—one change within each session. This allowed us to compare the behavior of our subjects to the behavior of an ideal detector.

In our model of an ideal detector, the probability density function for the reward rate on a given side has two components, one for the case in which the reward rate has not yet changed, and one for the case in which it has. The integrals of the two components sum to 1. As data indicative of a change in rate accumulate, the bulk of the probability shifts from the prechange component to the post-change component. The ratios of the two integrals give the odds that the change has already occurred. The shift is data driven; the model does not know where the change in rate will occur, only that there will be a change. Building the expectation that there will be a single step change into the process that estimates the current rate of reward allows this process to adjust its estimate of the current rates as fast as possible following

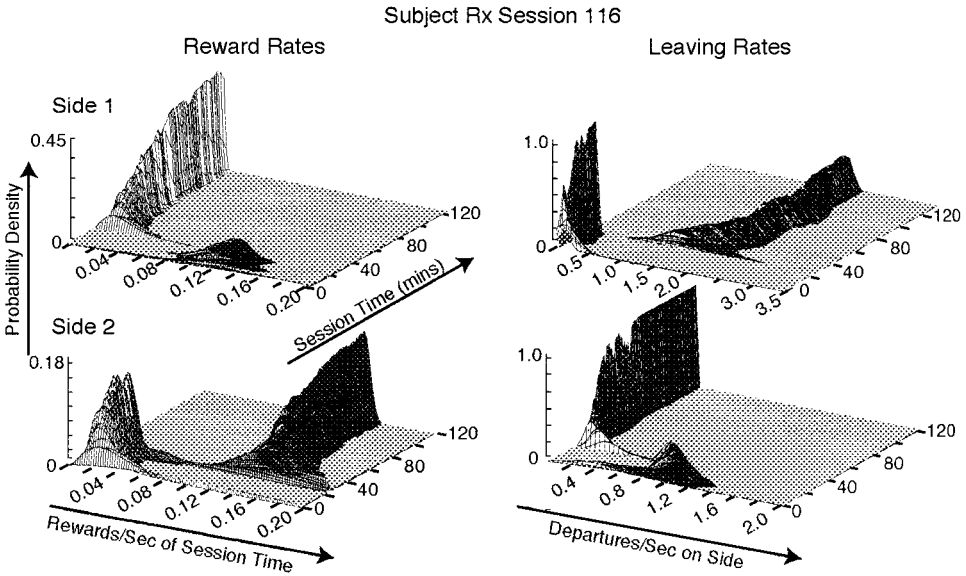

FIG. 2. Same as Fig. 1, except that the probability density functions were computed from the data for a different subject in a different session.

the change. Post-change estimates of the expected intervals are no longer averages over all the interreward intervals observed up to a given point in a session. When the pattern of intervals implies that the change has occurred, the current estimate of the rate of reward (and any behavior based on that estimate) is determined mostly by the intervals observed since the estimated change point. The stronger this implication is, the more completely the current estimate is based only on intervals since the change. (For the mathematical development of this Bayesian model of an ideal detector, see the Appendix of Gallistel *et al.* (2001).)

The left-hand panels in Figs. 1 and 2 show examples of the step-like changes in the estimates of the current rates of reward from our model of an ideal detector. At the beginning of a session, the probability density functions for the rates of reward are broad and low, reflecting the fact that the expected rates of reward cannot be known with any precision from the first few interreward intervals. As the session progresses, these functions become narrow and sharply peaked, reflecting the increasing precision with which the expected rates of reward can be estimated. At approximately the point in each session at which the scheduled rates of reward changed, the probability density landscapes abruptly flatten, reflecting the inescapable uncertainty about the expected rates of reward during a time of change in those rates. Soon after the change point, the ridges representing the likely values of the new rates of reward rise and narrow. The shift from the ridges representing

the prechange estimates to the ridges representing the post-change estimates can be and usually are step-like, because the estimates after the change are unaffected by the interreward intervals before the change.

We can represent the relevant aspect of the rat's behavior in the same way that we represent its objective experience of reward, because the durations of its stays at the two locations where rewards are found are exponentially distributed. This important fact was first reported by Heyman (1982) and later confirmed by Gibbon (1995). It implies that the termination of a visit to a location where rewards are found is determined by a process analogous to the repeated flipping of a coin. In Gibbon's 1995 model, the repeated sampling from the population of interreward intervals is analogous to the repeated flipping of a coin. When the sample interval from the other reward location comes up shorter than the sample interval from the location where the subject is, it leaves to visit the other location.

Because stay durations are exponentially distributed, the subject's behavior at a given location can be represented by an expected rate parameter—a leaving rate. The expected duration of a stay is the reciprocal of the expected leaving rate, just as the expected interval between rewards is the reciprocal of the expected reward rate.

The right-hand panels in Figs. 1 and 2 show the evolution of two different rats' leaving rates over time in representative sessions. The change in the expected stay durations almost coincides with the change in the experienced rates of reward, and it is about equally as abrupt. In other words, step changes in reward rates lead, after a very short latency, to step changes in expected stay durations. The abruptness of the changes sometimes observed is shown in Fig. 3 in another way. Here we plot the cumulative time at one location against the cumulative time at the other. The slope of this plot is the ratio of the expected stay durations. In the example shown in Fig. 3, the expected stay durations changed completely within the span of a single visit cycle.

The abruptness of the changes in expected stay durations implies that the process that determines them is like our model of an ideal detector in that it does not average across the population of interreward intervals before and after a change. Like our model, it must have a mechanism or procedure that detects the changes and, in effect, divides the population of interreward intervals on which post-change behavior is based into a prechange and a post-change population. Only the latter population determines post-change behavior.

We confirmed the impression of short-latency, abrupt changes by computing temporal probability density functions for the changes in rates of reward and the changes in leaving rates. The temporal probability density functions for the change in the rates of reward represent the accuracy with which any system not privy to the computer program could in principle determine the temporal locus of the change in rates of reward. Similarly, the temporal prob-

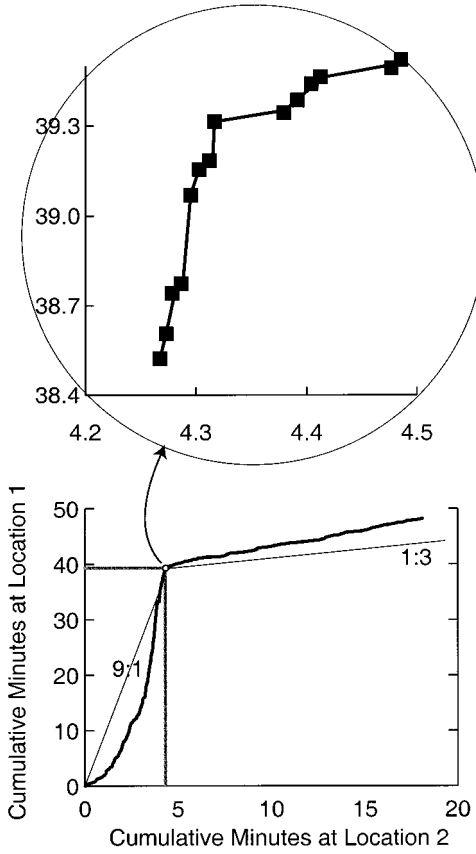

FIG. 3. (Bottom) The cumulative duration of the subject's stays at Location 1 plotted against the cumulative duration of its stays at Location 2 over one session. The gray lines give the coordinates of this plot at the moment when the ratio of the scheduled rates of reward changed, from 9:1 in favor of Location 1 to 1:3 in favor of Location 2. The thin lines labeled 9:1 and 1:3 show what the slope would be if the ratio of the subject's expected stay durations matched the ratio of the scheduled rates of reward. The slope of the plot at any point is the ratio of the expected stay durations at that point. The abrupt change in this ratio more or less coincides with the change in the relative rates of reward. The function at this point (the portion enclosed within the small circle at the intersection of the gray lines) is shown greatly enlarged in the top panel. The square points in this enlarged plot give the function at the completion of successive visit cycles (composed of one visit to each location). The change in the ratio of the expected stay durations was completed within the span of a single visit cycle.

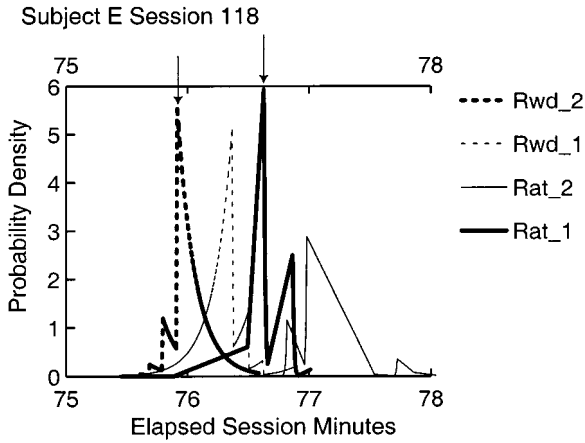

FIG. 4. The temporal probability density functions computed from the data underlying the rate probability density functions in Fig. 1. These functions specify the location in time of the changes in rates of reward and leaving rates seen in Fig. 1. The overlap in these functions is measured by the *t* statistic.

ability density functions for the leaving rates represent the accuracy with which we can know where the change in the rat's behavior occurred. The degree of overlap in these two functions represents how close the rat comes to being an ideal detector of the change in reward rates, how soon it detects the change relative to what is in principle possible. Figure 4 shows probability density functions for the temporal location of the changes portrayed in Fig. 1. In the example given, it comes close, indeed. This is not an unrepresentative example; indeed, it is longer than the typical case. The modal lag over 20 transitions in each of six subjects, as measured by the *t* statistic (see Fig. 4), was 0.25 (see Fig. 8 of Gallistel *et al.* (2001) for the distribution of these measures of overlap).

The adjustment to the change in the rates of reward often occurred before that change had an effect on the returns. Figure 5 gives an example. The top two panels plot the cumulative rewards obtained as a function of the cumulative amounts of time spent at each of the two locations. The slopes of these plots are the returns, the number of rewards per unit of time invested in a location. The axes of the right-hand top panel have both been scaled up by the same factor relative to the axes in the left-hand top panel; the fact that the slopes of the two plots are approximately the same reflects the fact that when a subject matches its time-allocation ratio to the ratio of the rewards obtained per unit of session time, it equates its returns. It will be recalled that it was this fact that was the key to the melioration model: matching equates returns.

The change in the programmed rates of reward occurred at the points indicated by the vertical broken lines extending across all three pairs of panels

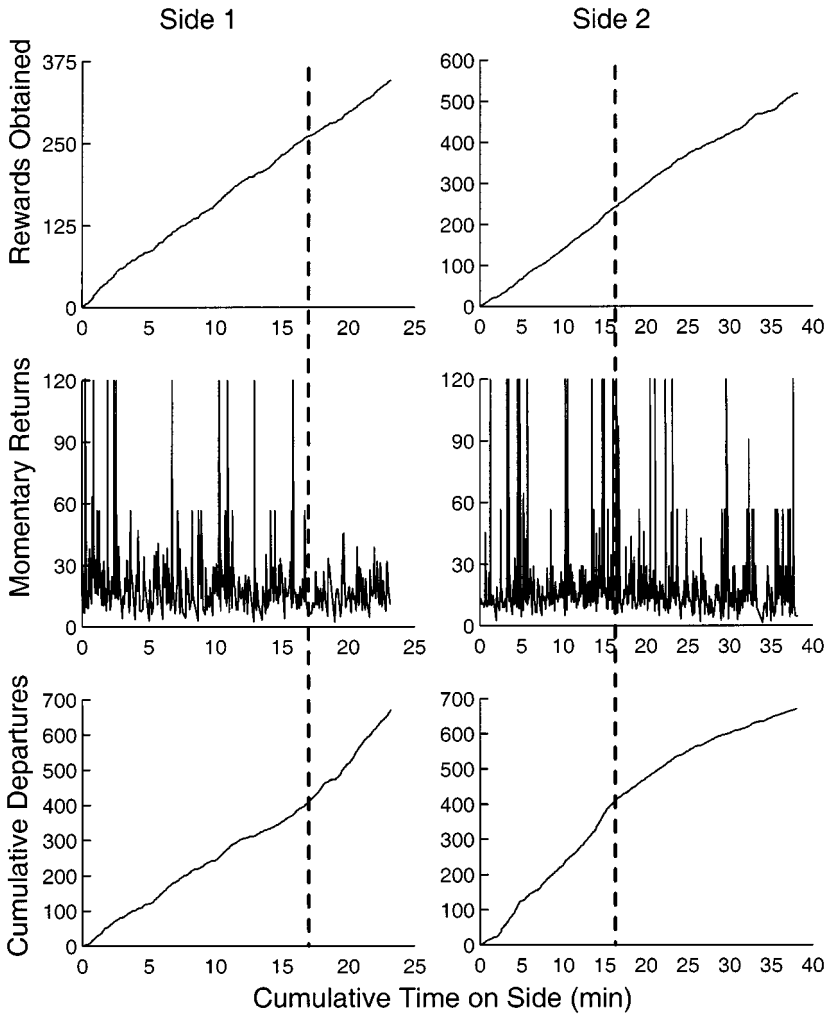

FIG. 5. (Top) The cumulative numbers of rewards on each side plotted against the cumulative time on location up to a given reward. The slope of this plot at a point is the expected return at that point. The vertical dashed lines indicate the moment at which the relative rates of scheduled rewards changed. Note the absence of any perturbation in expected returns at or immediately after this point. (Middle) At the moment when a reward is delivered, the momentary return at that location is the reciprocal of the time the subject has spent at that location since the last delivery. These momentary returns are plotted against the cumulative time on location. Note the extreme variability in these momentary returns. (Bottom) The cumulative number of departures plotted against the cumulative time on location. The slope of this plot at a point is the expected leaving rate at that point. Note the changes in expected stay durations immediately after the change in scheduled rates of reward.

in Fig. 5. The slopes of the plots in the bottom panels are the leaving rates. There are inflection points in these plots at or very shortly after the broken line, indicating that the subject adjusted its leaving rates almost immediately. The problem for the melioration model—and for any hill climbing model using returns—is that this adjustment occurred before the change in reward rates had an effect on the returns. There is no discernible perturbation in the slopes of the plots in the top panels preceding the clear change in the slopes of the plots in the bottom panels. (Nor do statistical tests for the presence of a change in returns yield any evidence of it; see Gallistel *et al.* (2001).) According to the melioration model, the change in rates of reward must first cause an observable difference in the returns, and then the subject must discover by a process of trial and error the departure rates (expected stay durations) that equate its returns under the new conditions. Thus, in this and other hill-climbing or feedback models, there must be an observable perturbation in the returns, which are then restored to equality by trial and error. This was not the case.

The plots of the momentary returns in the middle panels of Fig. 5 suggest how it is possible for the subject to react to a change in rates of reward before the change has affected its returns. When return is calculated on a reward by reward basis, it is a very noisy variable, because, as already noted, its variability is the combined result of the variability in scheduled interreward intervals and the variability in the subject's visits. Random variation in visit durations can mask for a while the effect on returns of a change in the obtained rates of reward.

Consider the case in which a decrease in the rate of reward at one location and an increase at the other happen to coincide with a string of shorter visits to the first location and longer visits to the second (relative to the respective expectations). During the sequence of fortuitously shorter visits to the first location, the intervals between experienced rewards (measured in session time) are longer than their expectation. If the subject's behavior is based on income, there is a clear signal to drive the observed behavioral change. The effects of these longer interreward intervals on the returns experienced is, however, masked by the fortuitous shortness of the visits, which reduces the subject's investment. The effect of the increased rate of reward at the other location is similarly masked; the subject experiences unexpectedly short interreward intervals, but it does not experience an increase in its returns because of its fortuitously lengthier visits. Thus, there is a change signal from the income variable but not from the return variable.

IMPLICATIONS FOR GIBBON'S MODEL

The finding that the rat adjusts to changes in the rates of reward about as fast as it is in principle possible to do implies that Gibbon (1995) was on the right track in abandoning the law of effect and propounding a feedforward model for matching behavior. Heyman (1982) was the first to suggest

that matching behavior was, in his words, ''unconditioned behavior,'' by which he meant that it was not the result of what is usually understood to be the essence of operant conditioning, namely, the shaping of behavior by the fed back effects of reinforcement, that is, selection by consequences. Rather, as Heyman argued, matching behavior is elicited by a certain pattern of reward without regard to the role, if any, that the subject's behavior played in producing them. This is the essence of the feedforward conception. Our results strongly support that conception.

On the other hand, the adjustments we observed occur too rapidly to be explained by the simple and ingenious feedforward model that Gibbon elaborated. The rate at which his model responds to a change in the experienced rates of reward is determined by the sizes of the populations of remembered interreward intervals from which the subject continually samples. The larger these populations are, the longer it takes to replace the prechange intervals with postchange intervals. Thus, to make the model respond rapidly to changes, one has to assume that the populations being sampled are small. One has to assume that the subject only samples from smallish populations of the most recently experienced incomes. However, the sampling populations cannot be too small. The model's behavior becomes unstable when the populations being sampled get too small, as Gibbon clearly recognized (personal communication to CRG).

If, for example, the subject samples from a population consisting of only the last three intervals, then sooner or later the population for the location with the longer expectation will happen to contain three relatively long intervals. In that case, there will be long stretches when the samples for the location with the shorter expectation are always shorter than the samples from the side with the longer expectation. This will lead to long stretches when the subject does not visit the location with the longer expectation. When it finally does visit (after an exceptionally long interval is sampled from the richer side), the interreward interval it experiences will necessarily be a long one. This will lead to ever lengthening intervals between visits and consequently ever longer interreward intervals in memory—a positive feedback process eventuating in memory stores such that the subject never visits the location with the longer expectation. The ever longer interreward intervals that it experiences are, of course, the consequence of its own long absences, but it is the essence of a feedforward model that it takes no account of the impact of the subject's own behavior on what it has observed. Income (by definition) takes no account of investment, even when investment is a crucial determinant of income, as it is in the example just discussed.

In several discussions of this problem, Gibbon and one of us (CRG) were not able to work out how big the populations had to be to make the risk of this kind of instability acceptably low. It was clear, however, that they had to consist of substantially more than three intervals. This means that, in his model, the completion of the subject's adjustment to the new rates of reward

would have to be substantially slower than many of the adjustments we observed. We often observed adjustments—indeed, even overadjustments—that went to completion in the span of one to three visit cycles. (Fig. 3 is an example.)

We conclude that Gibbon (1995) and Heyman (1982) were correct in assuming that matching behavior arises from a purely feedforward process. Matching is what the subject does after it has made certain observations about the distribution of rewards in space and time, without regard to the effect of its own behavior on the observed distributions.

The failure of the law of effect to apply under the simple and directly apposite circumstances created by concurrent random interval schedules of reward for competing operants raises the question, Under what circumstances is behavior determined through selection by its consequences? We know from the autoshaping literature (Bilbrey & Winokur, 1973; Brown & Jenkins, 1968) that the classic operants—key pecking in the pigeon and lever pressing in the rat—can be produced by Pavlovian contingencies, that is from experimental protocols in which the subject's behavior has no effect on reinforcement, so that the process leading to conditioned responding does not depend on such a contingency. We know from the negative auto-maintenance literature that when Pavlovian (feedforward) control is pitted against operant control (selection by consequences), feedforward control prevails (Williams & Williams, 1969). Subjects respond to a stimulus that predicts reward, even when their responding causes the omission of the predicted reward. Thus feedforward contingency prevails over feedback contingency. It is time to ask the radical question, Is the law of effect generally relevant to the understanding of instrumental behavior? If so, to what aspects of that behavior? That Gibbon's work leads us to ask such radical questions is one measure of its originality and importance.

REFERENCES

Baum, W. M., & Rachlin, H. C. (1969). Choice as time allocation. *Journal of the Experimental Analysis of Behavior,* **12,** 861–874.

Bilbrey, J., & Winokur, S. (1973). Controls for and constraints on autoshaping. *Journal of the Experimental Analysis of Behavior,* **20,** 323–332.

Brown, P. L., & Jenkins, H. M. (1968). Autoshaping of the pigeon's key-peck. *Journal of the Experimental Analysis of Behavior,* **11,** 1–8.

Gallistel, C. R., Mark, T. A., King, A., & Latham, P. (2001). The rat approximates an ideal detector of changes in rates of reward: Implications for the law of effect. *Journal of Experimental Psychology: Animal Behavior Processes,* **27,** 354–372.

Gibbon, J. (1995). Dynamics of time matching: Arousal makes better seem worse. *Psychonomic Bulletin and Review,* **2,** 208–215.

Herrnstein, R. J. (1982). Melioration as behavioral dynamism. In M. L. Commons, R. J. Hernnstein, & H. Rachlin (Eds.), *Quantitative analyses of behavior: Matching and maximizing accounts* (vol. 2). Cambridge, MA: Ballinger.

Herrnstein, R. J., & Prelec, D. (1991). Melioration: A theory of distributed choice. *Journal of Economic Perspectives,* **5,** 137–156.

Herrnstein, R. J., & Vaughan, W. J. (1980). Melioration and behavioral allocation. In J. E. R. Staddon (Ed.), *Limits to action: The allocation of individual behavior* (pp. 143–176). New York: Academic.

Heyman, G. M. (1982). Is time allocation unconditioned behavior? In M. Commons, R. Herrnstein, & H. Rachlin (Eds.), *Quantitative Analyses of Behavior: Matching and Maximizing Accounts* (vol. 2, pp. 459–490). Cambridge, MA: Ballinger Press.

Lea, S. E. G., & Dow, S. M. (1984). The integration of reinforcements over time. In J. Gibbon & L. Allan (Eds.), *Timing and time perception* (vol. 423, pp. 269–277). New York: Annals of the New York Academy of Sciences.

Montague, P. R., Dayan, P., & Sejnowski, T. J. (1996). A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *Journal of Neuroscience,* **16,** 1936–1947.

Rachlin, H., Logue, A. W., Gibbon, J., & Frankel, M. (1986). Cognition and behavior in studies of choice. *Psychological Review,* **93,** 33–45.

Schmajuk, N. A. (1997). *Animal learning and cognition: A neural network approach.* Cambridge, UK: Cambridge University Press.

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science,* **275,** 1593–1599.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning.* Cambridge, MA: MIT Press.

Thorndike, E. L. (1911). *Animal intelligence.* New York: MacMillan.

Williams, D. R., & Williams, H. (1969). Automaintenance in the pigeon: Sustained pecking despite contingent non-reinforcement. *Journal of the Experimental Analysis of Behavior,* **12,** 511–520.

Received April 1, 2001