JOURNAL OF THE EXPERIMENTAL ANALYSIS OF BEHAVIOR 2005, **84,** 555–579 NUMBER 3 (NOVEMBER)

DYNAMIC RESPONSE-BY-RESPONSE MODELS OF MATCHING BEHAVIOR IN RHESUS MONKEYS

BRIAN LAU AND PAUL W. GLIMCHER

NEW YORK UNIVERSITY

We studied the choice behavior of 2 monkeys in a discrete-trial task with reinforcement contingencies similar to those Herrnstein (1961) used when he described the matching law. In each session, the monkeys experienced blocks of discrete trials at different relative-reinforcer frequencies or magnitudes with unsignalled transitions between the blocks. Steady-state data following adjustment to each transition were well characterized by the generalized matching law; response ratios undermatched reinforcer frequency ratios but matched reinforcer magnitude ratios. We modelled response-by-response behavior with linear models that used past reinforcers as well as past choices to predict the monkeys' choices on each trial. We found that more recently obtained reinforcers more strongly influenced choice behavior. Perhaps surprisingly, we also found that the monkeys' actions were influenced by the pattern of their own past choices. It was necessary to incorporate both past reinforcers *and* past choices in order to accurately capture steady-state behavior as well as the fluctuations during block transitions and the response-by-response patterns of behavior. Our results suggest that simple reinforcement learning models must account for the effects of past choices to accurately characterize behavior in this task, and that models with these properties provide a conceptual tool for studying how both past reinforcers and past choices are integrated by the neural systems that generate behavior.

Key words: choice, matching law, dynamics, model, eye movement, monkey

Much of our understanding of choice behavior comes from the rich body of research using concurrent schedules of reinforcement like those Herrnstein used to describe the matching law (Herrnstein, 1961). Neuroscientists studying the biological mechanisms of choice behavior are poised to take advantage of the techniques and results that have been developed from studies on matching behavior (reviewed in Davison & McCarthy, 1988; de Villiers, 1977; B. Williams, 1988). However, monkeys are commonly used as subjects in neuroscientific research, but they are rarely used as subjects in matching experiments, and the behavioral research that has been done with monkeys (Anderson, Velkey, & Woolverton, 2002; Iglauer & Woods, 1974) differs in a number of procedural details from the behavioral methods used in neuroscience. In neurophysiological experiments, monkeys are head-restrained to allow for stable electrophysiological recordings, eye movements rather than arm movements are often used as a response measure, individual responses occur in discrete-trials to allow for precise control of timing, and water rather than food is used as a reinforcer to minimize recording artifacts caused by mouth movements. And perhaps most importantly, the reinforcement contingencies differ from the concurrent variable-interval (VI) schedules used to study matching behavior (but see Sugrue, Corrado, & Newsome, 2004). These differences make it difficult to relate monkey choice behavior as described by neuroscientists to the existing literature on the matching law.

One of our goals in the present study was to rigorously characterize the behavior of monkeys using apparatus typical of neurophysiological experiments while they performed a repeated choice task with reinforcement contingencies similar to the concurrent VI VI schedules used to study matching behavior. We used a discrete-trial task where the probability of reinforcement for choosing a particular alternative grows with the number of trials spent not choosing that alternative, similar to the exponential growth obtained

We thank Michael Davison, Nathaniel Daw, Peter Dayan, Mehrdad Jazayeri, and Aldo Rustichini for helpful discussions. Portions of this paper were presented at the 2004 meeting of Society for the Quantitative Analyses of Behavior and at Computational and Systems Neuroscience 2004. Supported by a NDSEG fellowship to Brian Lau and NEI EY010536 to Paul W. Glimcher.

All experimental procedures were approved by the New York University Institutional Animal Care and Use Committee, and performed in compliance with the Public Health Service's Guide for the Care and Use of Animals.

Address correspondence to Brian Lau, Center for Neural Science, New York University, 4 Washington Place, Room 809, New York, New York 10003 (e-mail: brian.lau@nyu.edu).

doi: 10.1901/jeab.2005.110-04

with constant-probability VI schedules (e.g., Houston & McNamara, 1981; Staddon, Hinson, & Kram, 1981). And like concurrent VI VI schedules, this method of arranging reinforcers renders exclusive choice of an alternative suboptimal for maximizing reinforcer rate because at some point the probability of reinforcement of the unchosen alternative exceeds that of the chosen alternative. In fact, for tasks like ours, maximizing reinforcer rate is closely approximated when subjects match the proportion of choices allocated to an alternative to the proportion of reinforcers received from choosing it (Houston & McNamara, 1981; Staddon et al., 1981). This also is true of concurrent VI VI schedules (Staddon & Motheral, 1978), and allows us to compare our results with previous work on the matching law.

Classic descriptions of choice relate behavioral output, typically measured as the long-term average ratio of choices, to average reinforcer input, typically measured as the long-term average ratio of reinforcers. The strict matching law equates these two quantities; the relative number of choices matches the relative reinforcer frequency (Herrnstein, 1961). This description has been extended to include other reinforcer attributes such as magnitude, delay, and quality (reviewed in Davison & McCarthy, 1988; de Villiers, 1977; B. Williams, 1988), and generalized to account for deviations from strict matching (Baum, 1974). The generalized matching law for reinforcer frequency and magnitude assumes that these attributes have independent effects (Baum & Rachlin, 1969), but allows for differential levels of control by each variable;

$$\frac{C_1}{C_2} = c\left(\frac{R_1}{R_2}\right)^a \left(\frac{M_1}{M_2}\right)^b, \qquad (1)$$

where C, R, and M denote the number of responses, number of reinforcers, and reinforcer magnitude for each alternative, respectively. The coefficients a and b relate the ratio of responses to reinforcer frequency and magnitude, and are interpreted as sensitivity to each variable, whereas c is a constant bias towards one alternative not related to reinforcer frequency or magnitude.

The generalized matching law describes an enormous amount of data collected from many species (de Villiers, 1977), and allows one to predict *average* choice behavior under different conditions, such as for arbitrary combinations of reinforcer frequency and magnitude. However, the generalized matching law does not specify how animals produce matching behavior at a response-by-response level. As such, the generalized matching law is less useful for making neurophysiological predictions because it does not constrain how neural activity should vary on a response-by-response basis. For neurobiological studies, a model of performance that specifies the computations performed by animals when allocating behavior would be of tremendous value because interpreting neural data often requires understanding how choices are produced. Ideally, such a model would furnish response-by-response estimates of the underlying decision variables (e.g., reinforcer rate) that predict the animal's choices. Measurements of brain activity then can be correlated with these theoretical variables to determine whether they predict brain activity as well as choice (e.g., O'Doherty et al., 2004; Platt & Glimcher, 1999; Sugrue et al., 2004).

Within psychology, a number of theoretical models have been formulated to explain matching behavior (B. Williams, 1988). Some of these models make predictions at the response-by-response level (e.g., momentary maximizing, Shimp, 1966), and considerable effort has been devoted to dissecting the local structure of choice to select amongst these different response-by-response models. Many of the analytical methods used to understand how local variations in behavior lead to matching can be broadly categorized as relating local variations in behavior to (a) local variations in reinforcer history or (b) local variations in behavioral history itself. Examples of the first class include transfer function formulations used by Palya and colleagues (Palya, Walter, Kessel, & Lucke, 1996, 2002; see McDowell, Bass, & Kessel, 1992, for a theoretical exposition in a behavioral context) to predict response rate using weighted sums of past reinforcers. Related techniques have been applied to predict choice allocation in concurrent schedules on short time scales (response-by-response, Sugrue et al., 2004) as well as longer time scales (within and between sessions, Grace, Bragason, & McLean, 1999; Hunter & Davison, 1985; Mark & Gallistel, 1994). An example of the second class of

techniques is estimating the probability of a particular choice conditional on a specific sequence of past choices, which can reveal the types of strategies animals use under different reinforcement conditions (e.g., Heyman, 1979; Nevin, 1969; Shimp, 1966; Silberberg, Hamilton, Ziriax, & Casey, 1978). Although it is conceptually useful to separate the effects of reinforcers and the potentially intrinsic response-by-response patterning of behavior (e.g., tendencies to generate particular sequences of behavior), behavior generally represents some combination of these two (Davison, 2004; Heyman, 1979; Palya, 1992). This suggests that analyses of reinforcer effects or sequential choice effects in isolation may not have a straightforward interpretation. Consider a situation in which an animal strictly alternates between two choice options. Reinforcement only is available from one alternative, and when received, causes the animal to produce one extra choice to it after which strict alternation is continued. Focusing solely on the relation between choice behavior and past reinforcers might lead to the incorrect interpretation that reinforcement changes choice behavior in an oscillatory manner.

These considerations led us to formulate a response-by-response model that treats the effects due to past reinforcers and choices as separable processes. We used this statistical model to predict individual choices based on weighted combinations of recently obtained reinforcers as well as previous behavior (c.f., Davison & Hunter, 1979), which allowed choice predictions to be influenced both by which alternatives had recently yielded reinforcers, as well as by sequential patterns in choice behavior.

The present results help test existing models of choice behavior. Many of the more plausible—in the sense that animals may actually implement them—theoretical models incorporate some form of reinforcer rate estimation using linear weightings of past reinforcers (e.g., Killeen, 1981; Luce, 1959). Recently, models using reinforcer rate estimation have been used to make predictions about neural activity. Single neurons in a number of cortical areas have activity correlated with reinforcer frequency (Barraclough, Conroy, & Lee, 2004; Platt & Glimcher, 1999; Sugrue et al., 2004) and magnitude (Platt & Glimcher, 1999). These and related studies (Glimcher, 2002; Montague & Berns, 2002; Schultz, 2004) are beginning to lay the groundwork for a detailed understanding of the mechanisms underlying choice behavior, and highlight the importance of behavioral models in forging a bridge between behavior and brain. We have found that local linear weightings of past reinforcers predict the choice behavior of monkeys performing a matching task, with weights similar to those suggested by reinforcer rate estimation models, but we extend this to show that accounting for recent choices also is required for accurate predictions by models of this type. The present results suggest that behavioral models incorporating dependence on both past reinforcers and choices may provide more accurate predictions, which will be essential for correlation with neural activity.

METHOD

Subjects

Two male rhesus monkeys (*Macaca mulatta*) were used as subjects (Monkey B and Monkey H, 10.5 kg and 11.5 kg respectively at the start of the experiments). Monkey H had been used previously in another experiment studying eye movements. Monkey B was naive at the start of the experiments described here. Both monkeys experienced a similar sequence of training (described below), and both were naive to the choice task used in this experiment.

Apparatus

Prior to behavioral training, each animal was implanted with a head restraint prosthesis and a scleral eye coil to allow for the maintenance of stable head position and recording of eye position. Surgical procedures were performed using standard aseptic techniques under isoflurane inhalant anaesthesia (e.g., Platt & Glimcher, 1997). Analgesia and antibiotics were administered during surgery and continued for 3 days postoperatively.

Training sessions were conducted in a dimly lit sound-attenuated room. The monkeys sat in an enclosed plexiglass primate chair (28 cm by 48 cm) that permitted arm and leg movements. The monkeys were head-restrained using the implanted prosthesis. Body movements were monitored from a separate room with a closed-circuit infrared camera. Eye movements were monitored and recorded

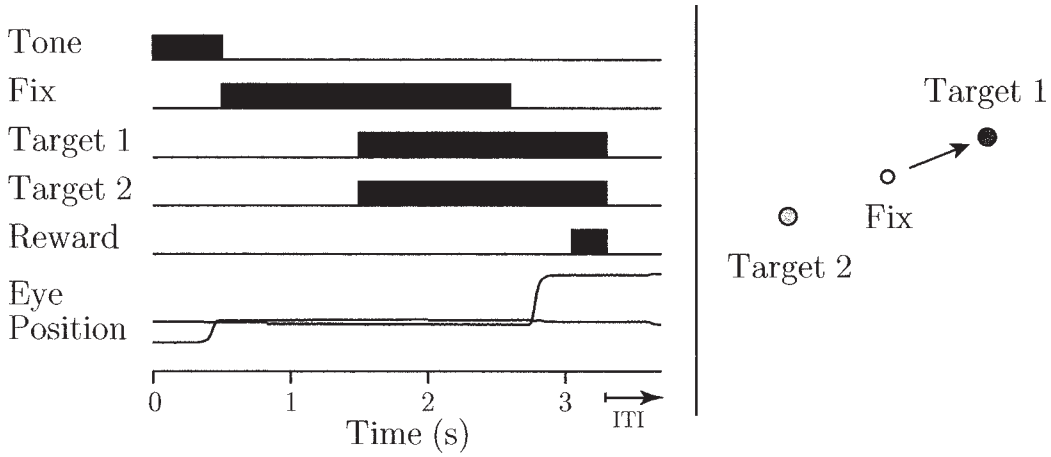

Fig. 1. Timeline and spatial arrangement of the two-alternative choice task.

using the scleral search coil technique (Judge, Richmond, & Chu, 1980) with a sampling rate of 500 Hz.

Visual stimuli were generated using an array of light-emitting diodes (LEDs) positioned 145 cm from the subject's eyes. The array contained 567 LEDs (21×27, 2° spacing), and each LED subtended about 0.25° of visual angle.

Procedure

Training began after a postoperative recovery period of 6 weeks. Water delivered into a monkey's mouth via a sipper tube was used as the reinforcer. The monkeys were trained on weekdays, and obtained the bulk of their water on these days during the 2 to 4 hr training sessions. The monkeys were not trained on the weekends, during which ad libitum water was provided. To ensure that the monkeys were adequately hydrated, we monitored the monkeys' weights daily and the University veterinarian periodically assessed the condition of the animals.

Initial training consisted of habituating the monkeys to being head-restrained in the primate chair. They were brought up from their home cages daily, and sitting quietly in the primate chair was reinforced with water. After 1 to 2 weeks of chair training, the monkeys were trained to shift their point of gaze to visual targets and to hold fixation using water reinforcement. We used manually controlled reinforcement to shape immediate and rapid movements to briefly illuminated LEDs, and incrementally delayed reinforcement to shape prolonged fixation. Once the monkeys could shift gaze rapidly to targets presented anywhere in the LED array and maintain fixation for 1 to 2 s on that LED, we placed reinforcement under computer control. The monkeys then were trained to shift gaze to eccentric visual targets from a central fixation point. When the monkeys were performing these tasks reliably, we began training on a choice task.

The data reported in this paper were collected while the monkeys performed the two-alternative choice task shown in Figure 1. Each trial started with a 500 ms 500 Hz tone, after which the monkey was given 700 ms to align their gaze within 3° of a yellow LED in the center of the visual field. After maintaining fixation for 400 ms, two peripheral LEDs (one red and one green) were illuminated on either side of the centrally located fixation point. These peripheral LEDs were positioned an equal distance from the fixation point; this was occasionally varied by a few degrees from session to session, but the distance was, on average, 15° of visual angle. One second later, the central fixation point disappeared, cueing the monkey to choose one of the peripheral LEDs by shifting gaze to within 4° of its location and fixating for 600 ms. If a reinforcer had been scheduled for the target chosen, it was delivered 200 ms after the eye movement was completed. The trial durations were the same whether or not the animal received reinforcement. Each trial lasted 3.2 to 3.65 s,

and only one choice could be made on each trial. Trials were separated by a 2-s intertrial interval (ITI) beginning at the end of the time a reinforcer would have been delivered. No LEDs were illuminated during the ITI. The range in trial durations comes from the variability in the amount of time taken by the monkeys to select and execute their eye movements. Relative to the duration of the trial, this time was short (mean reaction time 187 ms with a standard deviation of 71 ms).

A trial was aborted if the monkey failed to align its gaze within the required distances from the fixation or choice targets, or if an eye movement was made to one of the choice targets before the fixation point was extinguished. When an abort was detected, all the LEDs were extinguished immediately, no reinforcers were delivered, and the trial was restarted after 3 s. The monkeys rarely aborted trials (7% and 6% of trials for Monkey H and Monkey B, respectively), and these trials were excluded from our analyses.

Reinforcers were arranged by flipping a separate biased coin for each alternative using a computer random-number generator; if ''heads'' came up for a particular alternative, a reinforcer was armed, or scheduled, for that alternative. Reinforcers were scheduled using independent arming probabilities for each alternative, meaning that on any trial both alternatives, neither alternative, or only one alternative might be armed to deliver a reinforcer. If a reinforcer was scheduled for the alternative the monkey did *not* choose, it remained available until that alternative was next chosen (however, no information regarding held reinforcers was revealed to the subjects). For example, if a subject chose the left alternative when both alternatives were armed to deliver reinforcers, he would receive the reinforcer scheduled for the left alternative while the reinforcer for the right alternative would be held until the next time the subject chose it, which could be any number of trials into the future. A changeover delay or changeover response was not used; reinforcers were not delayed or withheld for an extra response when subjects chose to switch from selecting one alternative to the other. This method of arranging reinforcers in a discrete-trial choice task has been referred to as ''dual assignment with hold'' (Staddon et al., 1981).

In one set of sessions, we held the magnitude of reinforcement obtained from choosing either alternative equal while varying relative reinforcer frequency within sessions. We used arming probabilities that summed to about 0.3 (see Table 1 for precise values), and in each single session, the monkeys performed a series of trials under four different arming probability ratios (in Condition 1 for example, the ratios were 6:1, 3:1, 1:3, and 1:6). We switched between these ratios in an unsignalled manner as described below.

In another set of sessions, we held the programmed arming probabilities constant and equal while varying the relative reinforcer magnitudes within sessions. The amount of water delivered was controlled by varying the amount of time a solenoid inline with the water spout was held open. In each session, the monkeys performed a series of trials under four different magnitude ratios (in Condition 12 for example, the ratios were 3:1, 3:2, 2:3, 1:3), where the programmed arming probability for both alternatives was 0.15. We switched between these ratios in an unsignalled manner as described below.

Note that for the arming probabilities used, the subjects did not receive reinforcement on every trial; the monkeys received, on average, a reinforcer every three to four trials. However, aside from whether or not a reinforcer was delivered, the timing and appearance of each trial was identical to the subjects.

In each session, the monkeys performed a series of choice trials consisting of blocks of trials at different relative reinforcer frequencies or different relative reinforcer magnitudes. For example, in a frequency condition, a monkey might perform 124 trials where the right and left alternatives had an arming probability of 0.21 and 0.07, respectively (a 3:1 ratio) then 102 trials at a 1:6 ratio followed by 185 trials at a 3:1 ratio. Transitions between blocks of trials with different ratios were unsignalled, and the monkeys had to learn by experience which alternative had the higher frequency or magnitude of reinforcement. When blocks were switched, the richer alternative always changed spatial location, but its degree of richness was variable; the two possible ratios to switch to were chosen with equal probability. For example, a 3:1 or 6:1 ratio was followed by a 1:6 or a 1:3 ratio with equal probability. There were minor variations

Table 1

Experimental conditions for each monkey. The arming probability and magnitude of reinforcement are listed for each alternative. The magnitude of reinforcement is in units of milliliters.

	Monkey H				Monkey B		
	Arming probabilities	Magnitudes	Number of blocks		Arming probabilities	Magnitudes	Number of blocks
Frequency condition				Frequency condition			
1	0.24/0.04	0.35/0.35	16	5	0.25/0.05	0.4/0.4	5
	0.21/0.07	0.35/0.35	12		0.22/0.08	0.4/0.4	9
	0.07/0.21	0.35/0.35	11		0.08/0.22	0.4/0.4	7
	0.04/0.24	0.35/0.35	19		0.05/0.25	0.4/0.4	5
2	0.24/0.04	0.4/0.4	5	6	0.25/0.05	0.5/0.5	18
	0.21/0.07	0.4/0.4	8		0.22/0.08	0.5/0.5	15
	0.07/0.21	0.4/0.4	8		0.08/0.22	0.5/0.5	11
	0.04/0.24	0.4/0.4	4		0.05/0.25	0.5/0.5	18
3	0.24/0.04	0.45/0.45	7	7	0.29/0.04	0.5/0.5	12
	0.21/0.07	0.45/0.45	10		0.24/0.09	0.5/0.5	15
	0.07/0.21	0.45/0.45	5		0.09/0.24	0.5/0.5	13
	0.04/0.24	0.45/0.45	10		0.04/0.29	0.5/0.5	17
4	0.24/0.04	0.5/0.5	10	8	0.283/0.047	0.55/0.55	9
	0.21/0.07	0.5/0.5	13		0.248/0.082	0.55/0.55	8
	0.07/0.21	0.5/0.5	10		0.082/0.248	0.55/0.55	10
	0.04/0.24	0.5/0.5	12		0.047/0.283	0.55/0.55	3
5	0.25/0.05	0.4/0.4	9	9	0.283/0.047	0.6/0.6	3
	0.22/0.08	0.4/0.4	3		0.248/0.082	0.6/0.6	5
	0.08/0.22	0.4/0.4	7		0.082/0.248	0.6/0.6	5
	0.05/0.25	0.4/0.4	8		0.047/0.283	0.6/0.6	3
6	0.25/0.05	0.5/0.5	21				
	0.22/0.08	0.5/0.5	22				
	0.08/0.22	0.5/0.5	17				
	0.05/0.25	0.5/0.5	18				
Magnitude condition				Magnitude condition			
10	0.15/0.15	0.6/0.2	12	10	0.15/0.15	0.6/0.2	7
	0.15/0.15	0.5/0.3	10		0.15/0.15	0.5/0.3	9
	0.15/0.15	0.3/0.5	12		0.15/0.15	0.3/0.5	10
	0.15/0.15	0.2/0.6	10		0.15/0.15	0.2/0.6	2
11	0.15/0.15	0.6/0.2	42	11	0.15/0.15	0.6/0.2	20
	0.15/0.15	0.48/0.32	38		0.15/0.15	0.48/0.32	19
	0.15/0.15	0.32/0.48	36		0.15/0.15	0.32/0.48	21
	0.15/0.15	0.2/0.6	33		0.15/0.15	0.2/0.6	17
12	0.15/0.15	0.75/0.25	29	12	0.15/0.15	0.75/0.25	18
	0.15/0.15	0.6/0.4	27		0.15/0.15	0.6/0.4	21
	0.15/0.15	0.4/0.6	27		0.15/0.15	0.4/0.6	21
	0.15/0.15	0.25/0.75	20		0.15/0.15	0.25/0.75	20

in the ratios (Table 1), but we used the same method for block transitions in all frequency and magnitude conditions as described next.

We randomized the number of trials in each block to discourage any control over behavior by the number of trials completed in a block. The number of trials in a block for all conditions was 100 trials plus a random number of trials drawn from a geometric distribution with a mean of 30 trials. The mean number of trials per block observed was 121 trials for Monkey H and 125 trials for Monkey B, after excluding aborted trials.

The data analyzed were 67,827 completed trials from Monkey H and 47,160 completed trials from Monkey B. All the data reported were collected after at least 3 months of training on the choice task.

RESULTS

We present our results in two sections. First we describe the steady-state behavior at the

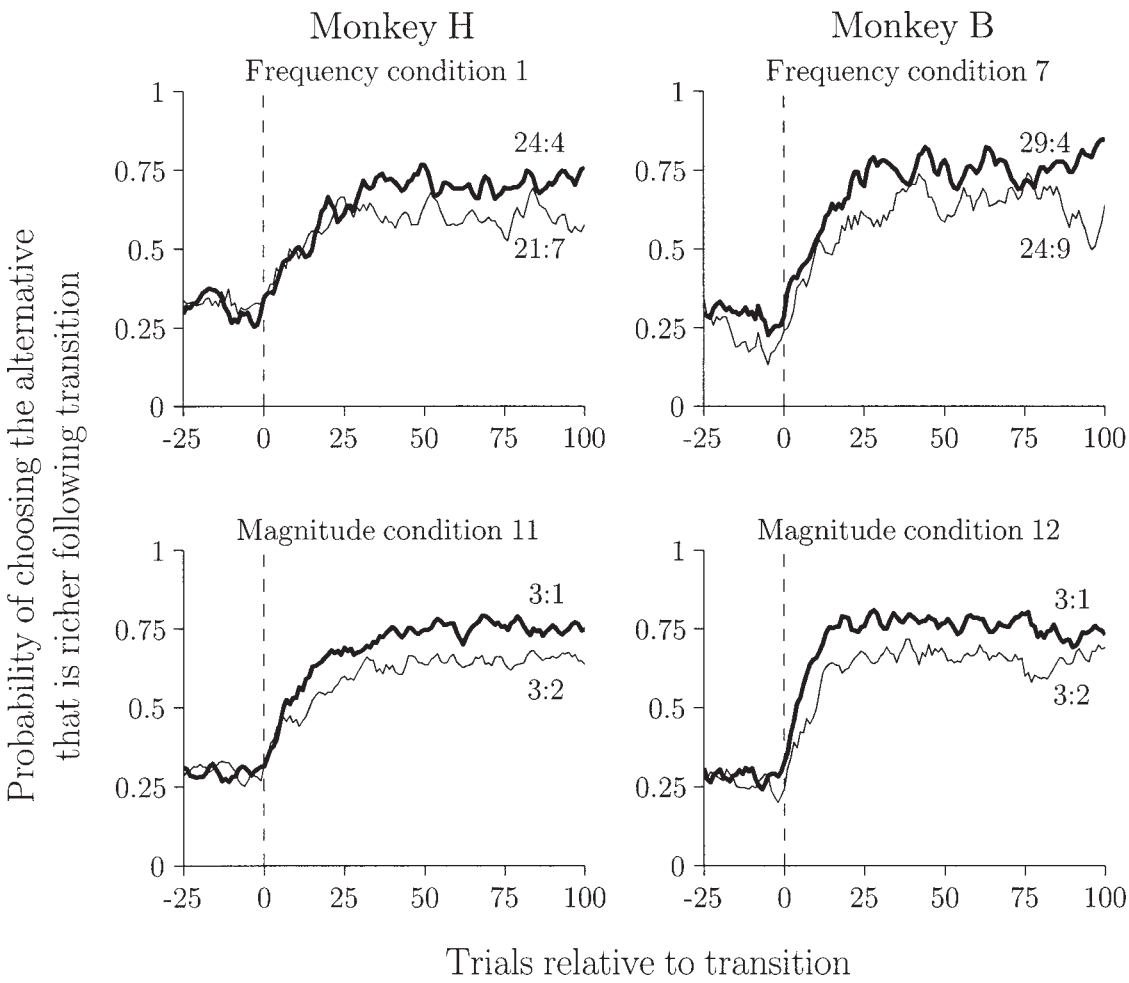

Fig. 2. Example choice data aligned on the trial (dashed line) that a transition between ratios occurred. The upper panels are data from conditions (Table 1) where reinforcer frequency was manipulated, and the lower panels are data from conditions where reinforcer magnitudes were manipulated. The data were compiled with respect to the alternative that was richer following the transition, and averaged separately for the two possible programmed posttransition ratios (pretransition ratios were on average ~1:4 for the top panels, and ~1:2 for the bottom panels). The data were smoothed using a five-point moving average.

end of each block in terms of the generalized matching law. This establishes the behavioral endpoint that we seek to explain using response-by-response models in the subsequent section.

Steady-State Analysis

Figure 2 shows examples of responses to block transitions for each monkey. The curves show the probability of choosing the alternative that was richer *following* an unsignalled block transition, aligned on the trial at which the transition occurred. These probabilities were computed by averaging choices for trials preceding and following a block transition. For example, the thick line in the upper left panel is an average of all transitions that ended with an arming probability ratio of 24:4. Likewise, the thin lines in the lower panels are averages of all transitions that ended with a magnitude ratio of 3:2. The curves in Figure 2 are similar prior to the transition because both are averages of the two possible pretransition ratios (see Methods). There was a rapid acquisition period following the transition, and the curves diverged for the two possible posttransition ratios. We quantified the speed of acquisition using the number of trials it

took the monkeys to reach a 50% rate of choosing the richer alternative following a block transition. This was 11.4 trials for Monkey H and 10.42 trials for Monkey B averaged over all the conditions where reinforcer frequency was manipulated. There was no difference between the acquisition times for the two possible posttransition ratios in each condition (~3:1 and ~6:1). The acquisition time was 11.56 trials for Monkey H and 10.86 trials for Monkey B averaged over all the conditions where reinforcer magnitude was manipulated. However, for the reinforcer magnitude conditions, there was a difference between these times that depended on the posttransition reinforcer magnitude ratio; transitions to a 3:1 ratio were faster (9.84 trials for Monkey H and 8.55 trials for Monkey B) than transitions to a 3:2 ratio (13.27 trials for Monkey H and 13.17 trials for Monkey B).

Figure 2 shows that there was a period following acquisition where the animals' behavior fluctuated about a steady-state. We characterized the differences in steady-state preference by analyzing the last 65 trials of each block. We chose this number based on plots like Figure 2, which show that behavior had, on average, reached a steady-state by this point in each block.

Figure 3 shows log ratio responses from example reinforcer frequency and reinforcer magnitude conditions. We fit the data for each condition with the logarithmic form of the generalized matching law,

$$\log\left(\frac{C_1}{C_2}\right) = a\log\left(\frac{R_1}{R_2}\right) + b\log\left(\frac{M_1}{M_2}\right) + \log c, \quad (2)$$

using least-squares regression. Equation 2 was used without the magnitude term for the conditions where we manipulated reinforcer frequency, whereas for the magnitude conditions, we included the term for reinforcer frequency because the reinforcement schedules did not enforce precise equality between the obtained reinforcer frequencies from the two alternatives. The estimated parameters from these fits are listed in Table 2. The coefficients and quality of fits were similar for the 2 monkeys, consistent with the examples shown in Figure 3. Both animals showed greater sensitivity to reinforcer magnitude (b) than reinforcer frequency (a).

Figure 4 shows log response ratios as a function of log obtained reinforcer frequency and log reinforcer magnitude ratios for all our data. The plane through the data represents a least-squares fit of Equation 2, excluding data in which no reinforcers were obtained from one of the alternatives (six of 551 blocks for Monkey H, 10 of 376 blocks for Monkey B). The generalized matching law provided good fits to the data, accounting for 90% of the variance in the data for both monkeys.

We assessed the degree to which our data deviated from the generalized matching law by checking for a dependence of the observed behavior on nonlinear transformations of relative reinforcer frequency and relative reinforcer magnitude not captured by Equation 2. We did this by adding polynomial covariates to Equation 2; for the reinforcer frequency conditions, we fit additional coefficients for log reinforcer frequency ratio raised to the second and third powers, whereas for the reinforcer magnitude conditions, we fit additional coefficients for log reinforcer frequency and log reinforcer magnitude ratios raised to the second and third powers. Including these covariates accounted for quadratic and cubic deviations in the data, but only increased the average percentage of variance explained by 1.8% (range, 0.31 to 3.81%) and 1.5% (range, 0.13 to 2.63%) for the reinforcer frequency and reinforcer magnitude conditions, respectively. This indicates that the generalized matching law is a reasonable description of our data, even though it typically is used to describe data after many sessions, rather than trials, of training under the same reinforcer conditions.

Although the generalized matching law fits the steady-state behavior that we measured, it only describes average choice. It is silent regarding the sequences of choices underlying these averages. At one extreme, we might hypothesize that the response-by-response process that achieves steady-state behavior operates is purely deterministic. At the other extreme, we might hypothesize that behavior is allocated according to a purely probabilistic strategy. We examined how probabilistic the monkeys were by studying how runlengths, the number of consecutive choices to a particular alternative, vary with the degree of preference, as measured by the relative responses from each block. If the monkeys chose in a purely

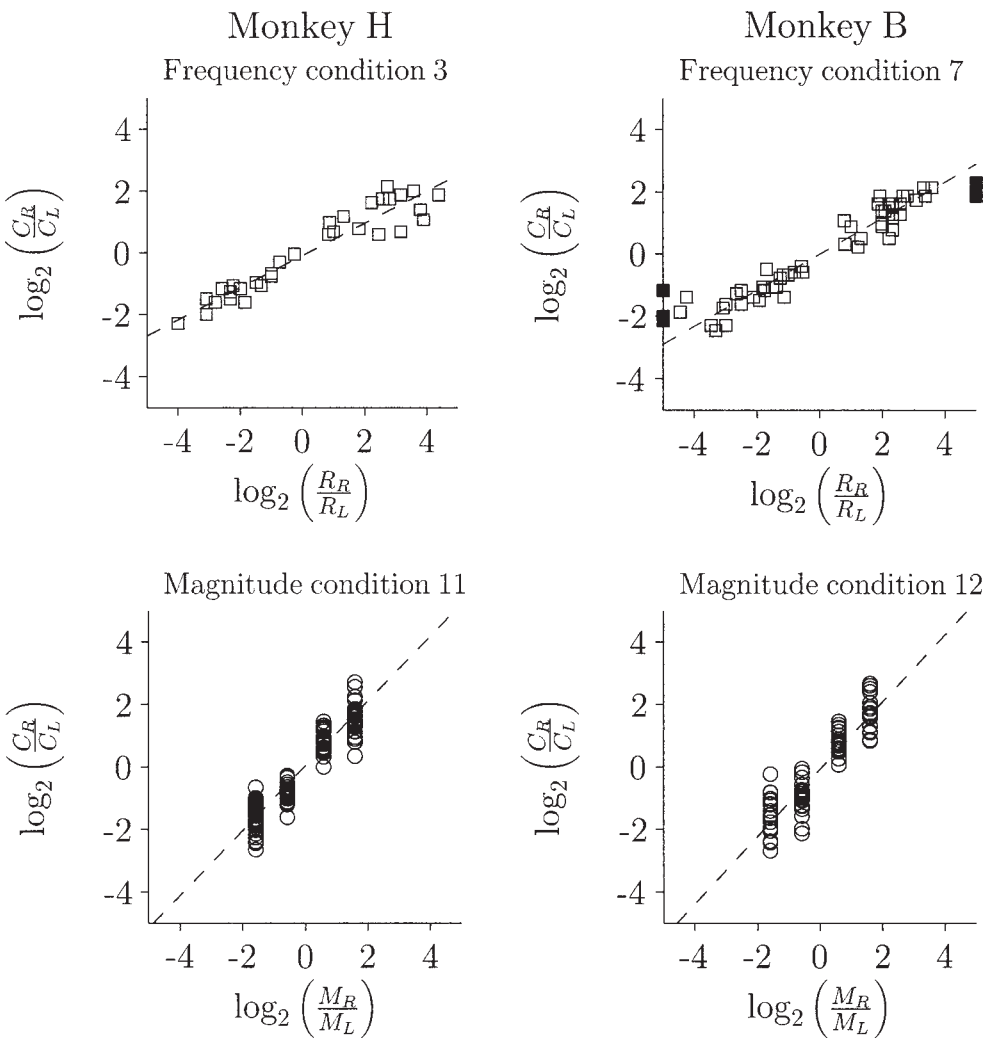

Fig. 3. Log choice ratios (right over left) from individual conditions (see Table 1 for ratios used in each condition) as a function of obtained log reinforcer frequency ratios (upper panels) or log reinforcer magnitude ratios (lower panels). Each point was obtained by averaging the number of choices and reinforcers from the last 65 trials of a block. The lines represent least-squares fits of the generalized matching law, the coefficients of which are listed in Table 2. The solid symbols plotted at the extremes of the abscissa (for Monkey B) represent blocks where no reinforcers were obtained from one of the alternatives. For the magnitude conditions, the effect due to reinforcer frequency has been subtracted from the log choice ratios.

probabilistic manner by flipping a weighted coin on each trial, with the weight dictated by the degree of preference (e.g., a 2:1 preference gives a two-thirds probability of choosing the preferred alternative on each trial), the runlengths would be distributed geometrically with a mean equal to the reciprocal of the coin weight. This is a useful baseline model because the assumption of independence between trials that defines this model means that the degree of weighting of the coin fully specifies the runlength distributions (e.g., Houston & Sumida, 1987).

Figure 5 shows runlength as a function of preference, with these variables measured from the last 65 trials in each block. Runlengths were measured for both alternatives; for example, the sequence of choices RLLLLLR from a block where L was preferred would contribute a runlength of five to the runlengths for the rich, or preferred, alternative and a runlength of one to the runlengths

Table 2

Fits of the generalized matching law (Equation 2) to the steady-state data from the end of each block. The condition labels in the first column correspond to those in Table 1.

	Monkey H				Monkey B			
	log c (*SE*)	a (*SE*)	b (*SE*)	R^2	log c (*SE*)	a (*SE*)	b (*SE*)	R^2
Frequency condition								
1	0.06 (0.06)	0.51 (0.03)	—	0.88				
2	0.10 (0.09)	0.50 (0.04)	—	0.89				
3	−0.08 (0.08)	0.52 (0.03)	—	0.91				
4	0.16 (0.07)	0.44 (0.03)	—	0.87				
5	0.00 (0.07)	0.57 (0.03)	—	0.93	−0.01 (0.09)	0.51 (0.04)	—	0.88
6	0.05 (0.05)	0.54 (0.02)	—	0.89	−0.10 (0.07)	0.56 (0.03)	—	0.86
7					0.00 (0.05)	0.58 (0.02)	—	0.93
8					−0.01 (0.08)	0.49 (0.03)	—	0.89
9					−0.10 (0.08)	0.56 (0.04)	—	0.95
pooled	0.05 (0.03)	0.51 (0.01)	—	0.88	−0.04 (0.03)	0.54 (0.01)	—	0.89
Magnitude condition								
10	0.08 (0.05)	0.31 (0.06)	1.02 (0.04)	0.96	−0.15 (0.11)	0.43 (0.16)	1.31 (0.11)	0.92
11	0.05 (0.03)	0.34 (0.05)	1.04 (0.03)	0.92	−0.18 (0.05)	0.49 (0.07)	1.34 (0.05)	0.95
12	0.08 (0.04)	0.31 (0.06)	0.94 (0.04)	0.90	−0.05 (0.06)	0.56 (0.08)	1.08 (0.06)	0.88
pooled	0.06 (0.02)	0.32 (0.03)	1.00 (0.02)	0.92	−0.11 (0.03)	0.53 (0.05)	1.22 (0.04)	0.91

for the lean, or nonpreferred, alternative. Because neither animal exhibited a significant spatial bias, runlengths were compiled for the preferred and nonpreferred alternatives irrespective of spatial location. The points have been jittered slightly along the abscissa to reveal data that otherwise would have stacked on top of each other. Almost as soon as one of the alternatives was preferred, the runlengths for the nonpreferred alternative became very short, typically one trial (mean across all preferences was 1.05 trials with a standard deviation of 0.09).

Figure 5 also shows the predicted average runlengths under the probabilistic strategy described above. The actual runlength data systematically deviate from and hence reject this model, implying that average preference as specified by the generalized matching law is insufficient to specify the average runlength

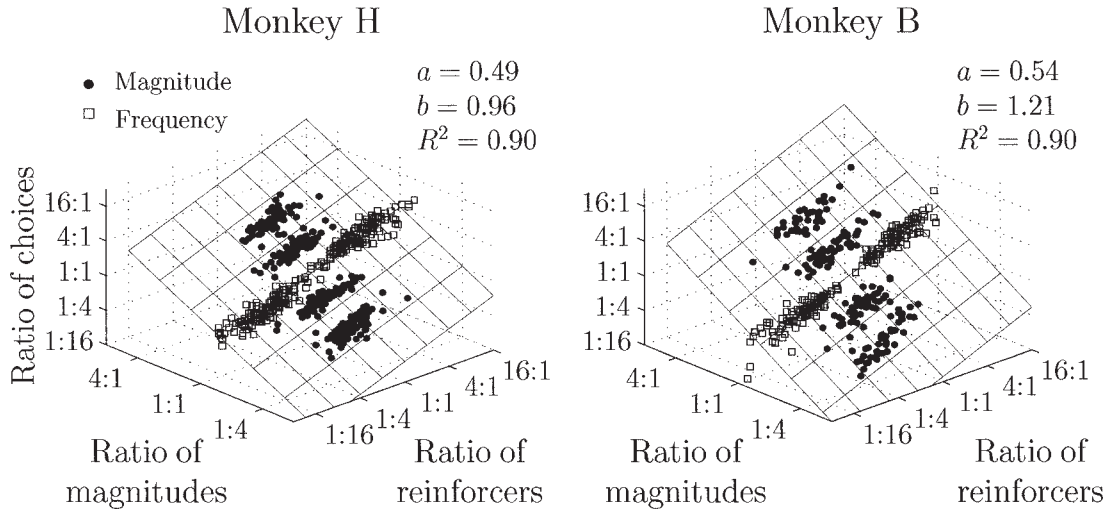

Fig. 4. Log choice ratios (right over left) as a function of obtained log reinforcer frequency ratios and log reinforcer magnitude ratios. Each point was obtained by averaging choices and reinforcers from the last 65 trials of a block. All such data from both the frequency (open squares) and magnitude (filled circles) experiments are plotted. The planes through the data are fits of the generalized matching law to the entire data set for each monkey.

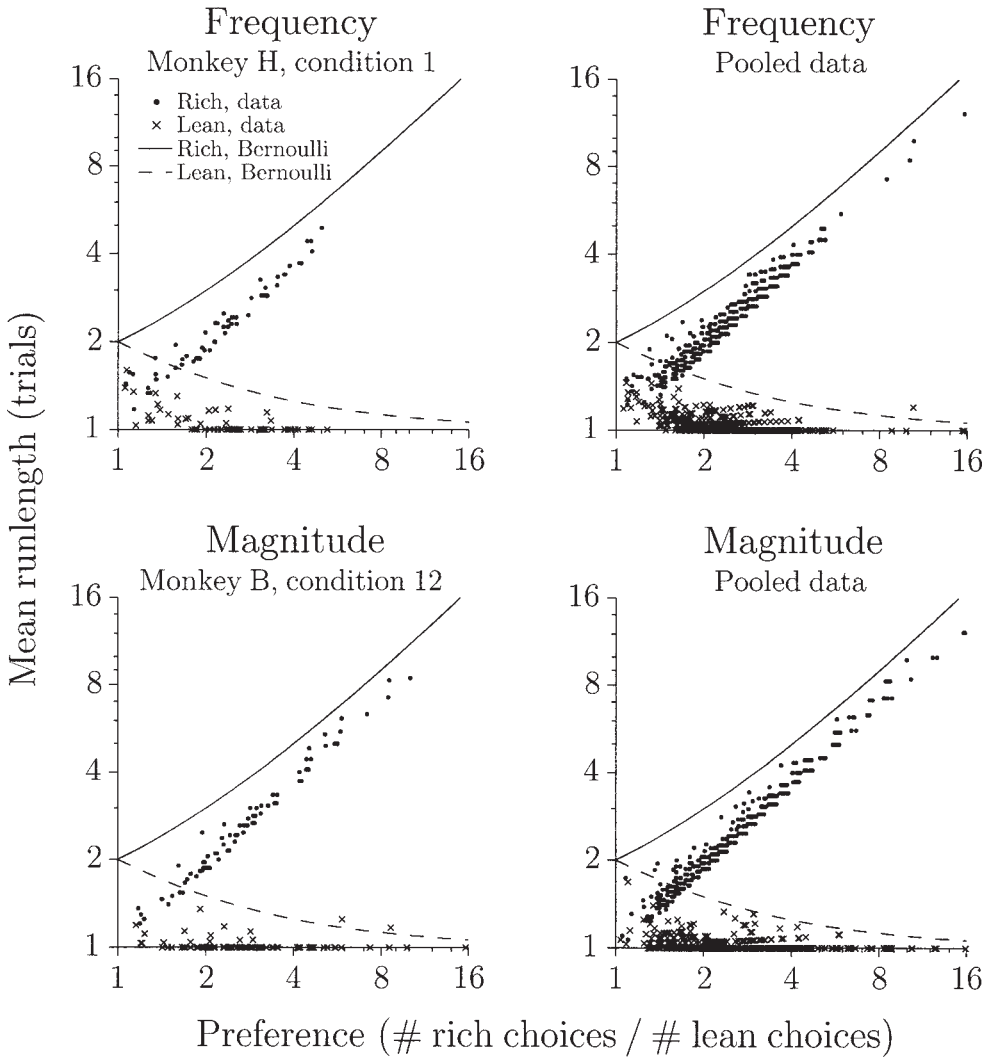

Fig. 5. Mean runlength as a function of preference. Each point represents the mean runlength in the last 65 trials of each block, plotted separately for choices of the rich (circles) and lean alternatives (×s). The data are plotted on log-log coordinates, and the points are jittered slightly along the abscissa to reveal points that otherwise would have stacked on top of each other. The lines represent mean runlengths predicted by a Bernoulli process that allocates choice independently from trial-to-trial, as in a series of independent weighted coin flips. The left panels show examples from single conditions for each monkey. The right panels show runlengths for all the data combined; for clarity, the data points for the conditions used in the left panels are not included in the pooled data.

patterns. Instead, they show that the monkeys tend to stay at the richer alternative, switching briefly to sample the leaner alternative. Baum, Schwendiman, and Bell (1999) observed this pattern of behavior in pigeons under concurrent VI VI schedules, terming it ''fix-and-sample.'' The runlength analysis highlights the fact that the monkeys' choices on individual trials were not independent during steady-state behavior. These response-by-response dependencies could have arisen from the effects of past reinforcers and/or past choices, but how this occurs cannot be revealed through sequential statistics like runlengths because they do not take into account when reinforcers were delivered. In the next section we characterize these response-by-response dynamics by constructing statistical models that predict choice on each trial based on the history of reinforcers and choices.

Dynamic Analysis

We used response-by-response models to predict choice on each trial using the past history of reinforcers and choices. Here we introduce some necessary notation. Let $c_{R,i}$ and $c_{L,i}$ represent the choice to left and right alternatives on the ith trial. These variables are binary; a value of 1 indicates a choice of a particular alternative. In order to fit a model to the data, we seek to estimate the probability, $p_{R,i}$, of choosing the right alternative (i.e., the probability $c_{R,i} = 1$). We assume that past reinforcers and choices are linearly combined to determine choice on each trial, which allows us to write the logistic regression,

$$\log\left(\frac{p_{R,i}}{p_{L,i}}\right) = \sum_{j=1} \alpha_{R,j} r_{R,i-j} + \sum_{j=1} \alpha_{L,j} r_{L,i-j} + \sum_{j=1} \beta_{R,j} c_{R,i-j} + \sum_{j=1} \beta_{L,j} c_{L,i-j} + \gamma, \quad (3)$$

where r is the magnitude of reinforcement in milliliters of water received for choosing a particular alternative on the jth past trial, and zero otherwise. Like the generalized matching law, Equation 3 implies that the probability and the magnitude of reinforcers have independent effects (i.e., they combine multiplicatively). The α and β coefficients measure the influence of past reinforcers and choices, and the intercept term γ captures preference not accounted for by past reinforcers or choices, similar to the bias term in the generalized matching law. The coefficients represent changes in the natural logarithm of the odds of choosing the right alternative (or equivalently left since $p_{L,i} = 1 - p_{R,i}$). The model is linear in the log odds, but nonlinear in the probability of choice, which can be recovered by exponentiating both sides of Equation 3 and solving for $p_{R,i}$.

We simplify the model by assuming that the effects of past reinforcers and choices are symmetric; a past reinforcer or a choice to the right alternative changes the odds of choosing right by the same amount that a past reinforcer or a choice to the left alternative changes the odds of choosing left. Note that this assumption is required for the past choices (but not past reinforcers) in order to be able to fit the model because there are only two choice alternatives, $c_{L,i} = 1 - c_{R,i}$. Then, the model reduces to

$$\log\left(\frac{p_{R,i}}{p_{L,i}}\right) = \sum_{j=1} \alpha_j \left(r_{R,i-j} - r_{L,i-j}\right) + \sum_{j=1} \beta_j \left(c_{R,i-j} - c_{L,i-j}\right) + \gamma. \quad (4)$$

Here a unit reinforcer obtained j trials in the past increases the log odds of choosing an alternative by α_j if the reinforcer was received for choosing that alternative, otherwise it decreases the log odds by α_j. This applies similarly to the effects of past choices, where a significant β_j means that the current choice depends on a choice made j trials ago (Cox, 1970).

Note that excluding the terms associated with the β parameters (forcing all $\beta_j = 0$) yields a model that only depends on the history of obtained reinforcers, producing results similar to those obtained by transfer function estimation (c.f., Palya et al., 1996, 2002; Sugrue et al., 2004). Alternatively, excluding the terms associated with the α parameters (forcing all $\alpha_j = 0$) yields a regression that only depends on the history of choices taken. Including both α and β allows us to assess the independent effects reinforcer and choice history have on current choice.

The intercept term γ shifts preference towards one of the alternatives irrespective of reinforcement (it captures a bias not due to either reinforcer frequency or reinforcer magnitude). It also is possible that the monkeys were biased towards the rich alternative, independent of its spatial location. This is distinct from the effect captured by γ, because a bias towards the rich alternative will switch repeatedly within a session (as the location of the richer alternative switched), so long as the animal can actually identify the richer alternative (Davison & Baum, 2003). We tested for this effect by adding a dummy variable that was +1 on trials when the right alternative was rich and −1 when the left alternative was rich. This allows for a bias towards or away from the rich alternative depending on the sign of the fitted coefficient for the dummy variable. We assigned the identity of the rich alternative according to the first reinforcer the animal obtained in a block, and this remained fixed until the first reinforcer obtained in the following block, at which point the sign of

the dummy variable switched. A rich bias is separable from the effects of past reinforcers because the first sum in Equation 4 only contributes when reinforcers are delivered (recall that reinforcers are not delivered on every trial), whereas a bias towards the rich alternative can influence behavior even when no reinforcers are delivered.

The complete model therefore had four components: a dependence on past reinforcers and choices, as well as two kinds of bias, one for a constant bias towards an alternative and another for a bias towards the rich alternative. We fit this model to the data using the method of maximum likelihood (e.g., Fahrmeir & Tutz, 2001; McCullagh & Nelder, 1989). The model was fit twice for each monkey, once for all of the data pooled across reinforcer frequency conditions (32,126 trials for Monkey H and 24,337 trials for Monkey B), and again for all of the data pooled across reinforcer magnitude conditions (35,701 trials for Monkey H and 22,823 trials for Monkey B). The intersession spacing was dealt with by padding the design matrix with zeros so that reinforcers and choices from different sessions were not used to predict responses from other sessions.

The results obtained from fitting the model are plotted in Figure 6. The coefficients represent changes in the log odds of choosing an alternative due to a unit reinforcer (upper panels) or a previous choice (middle panels). The coefficients are plotted for models fit using a history of 15 trials. We address how long a history is required by the most parsimonious model in a later section; extending the history does not affect the results presented here. The change due to a reinforcer j trials in the past (upper panels) can be determined by reading off the coefficient value from the ordinate and multiplying it by the magnitude of reinforcement in milliliters. For example, the effect of a typical 0.5 ml reinforcer four trials in the past for Monkey H in the probability conditions is $0.83 \times 0.50 = 0.42$. This is an increase in log odds, so exponentiating gives an increased odds of 1.52 for choosing the reinforced alternative on the current trial. The pattern of results for reinforcer history is similar for both animals; the most recent reinforcers strongly increased the odds of choosing the reinforced alternative again, and this influence decayed as reinforcers receded in time. For both monkeys, the coefficients for reinforcer history are similar in shape for both the probability and magnitude conditions, although the coefficients are larger in the probability conditions. Relative to Monkey H, the coefficients for Monkey B are initially larger, but also decay more quickly, indicating that recent reinforcers would shift preference more strongly, but that their effects are more local in time.

The middle panels in Figure 6 indicate that the effects of recent choices are somewhat more complicated. These coefficients should be interpreted in the same way as the coefficients for past reinforcers, except that there is no need to scale the coefficients because the choice variables are binary. The negative coefficient for the last trial means that choosing a particular alternative decreased the odds of choosing that alternative again on the next trial, independent of any effects due to reinforcer history. Alone, this coefficient would produce a tendency to alternate. Positive coefficients for the trials further in the past than the last trial mean that choosing a particular alternative some time in the past increases the likelihood of choosing it again, independent of the effects due to past reinforcers. Alone, these coefficients would produce a tendency to persist on an alternative.

Taken together the model coefficients indicate that recent reinforcers and choices act to produce persistence on the rich alternative with occasional sampling of the lean alternative. This pattern comes about because larger or more frequent reinforcers increase the likelihood of repeating choices to the alternative that yields them. Momentary decreases in rich reinforcers eventually cause a switch to the lean alternative, followed by a tendency to switch immediately back to the rich alternative due to longer-term choice history effects.

The bias terms are plotted in the bottom panels of Figure 6. There was little or no spatial bias towards either alternative, which was consistent with our steady-state analysis using the generalized matching law. There was, however, a bias towards the rich alternative that was larger for the magnitude conditions.

We now return to the issue of determining the length of reinforcer and choice histories supported by the data. Equation 4 does not specify a unique model; each combination of

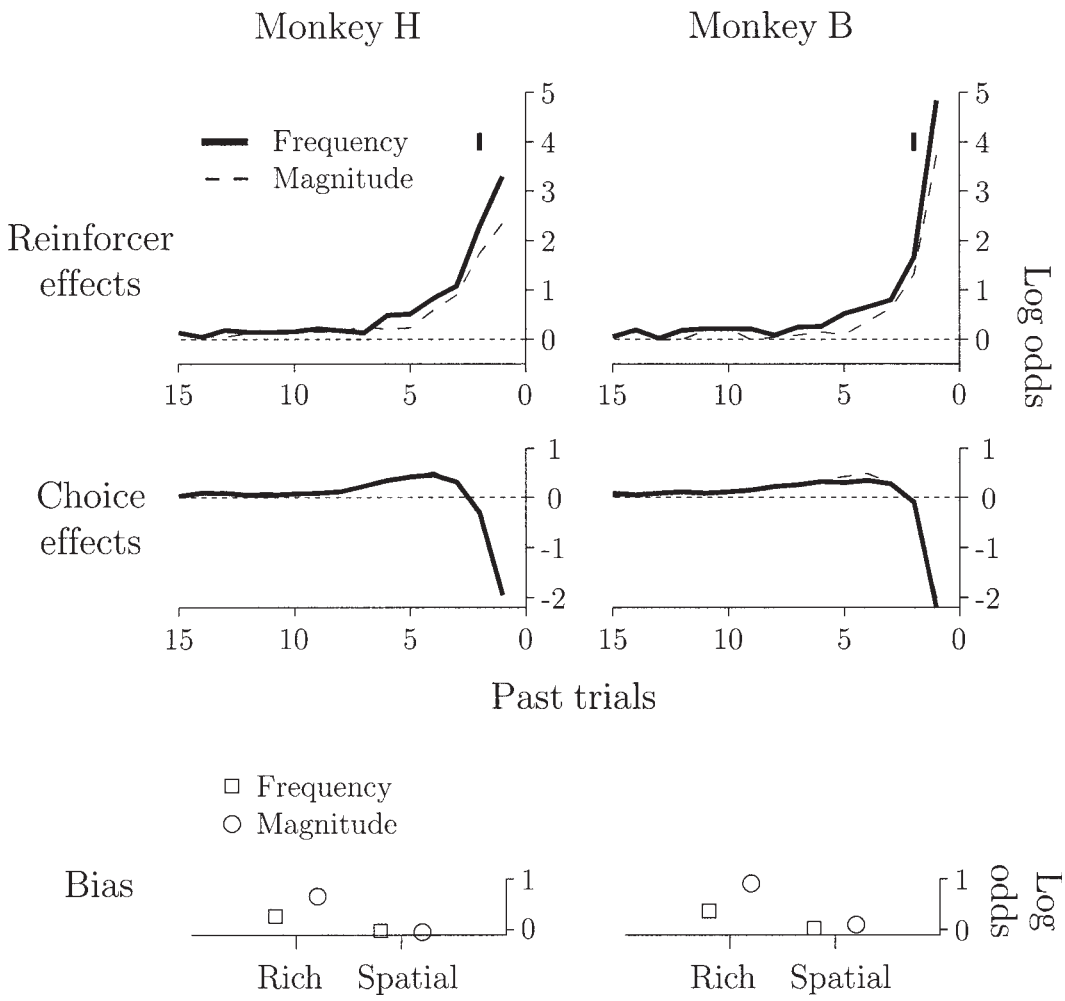

Fig. 6. Coefficients for the fitted dynamic linear model as a function of the number of trials in the past relative to the current trial. The coefficients for past reinforcers, past choices, and biases are plotted in the upper, middle, and lower panels, respectively. The vertical ticks in the upper panels represent the largest 95% confidence intervals for the past reinforcers. Confidence intervals are not plotted for the other coefficients, as they were smaller than the symbols.

lengths of reinforcer and choice histories constitutes a candidate model. In principle, one could include nearly all past reinforcers and choices in the model; however, this lacks parsimony, and the resulting parameters can be highly variable because estimation is limited by the amount of data available. Alternatively, including too few parameters leads to a poor fit. We used Akaike's Information Criterion (AIC; Akaike, 1974) to select objectively a best model given the data. The AIC estimates the information lost by approximating the true process underlying the data by a particular model (see Burnham & Anderson, 1998). For each candidate model, the AIC is computed as

$$AIC = -2\ln(\mathcal{L}) + 2k, \tag{5}$$

where $\mathcal{L}$ is the maximized log-likelihood of the model fit and k is the number of parameters. Equation 5 enforces parsimony by balancing the quality of each fit with the increase in model complexity brought by adding more parameters; good fits obtained by adding more parameters decrease the first term but also increase the second term. This makes intuitive sense; at some point adding more parameters

Table 3

Model comparison using Akaike's Information Criterion. The three rows for each condition correspond to three separate models, with each row representing a more complex model. The third row for each condition is the best-fitting model according to the *AIC* metric, and the Δ*AIC* column gives the difference between each model and the best model's *AIC* value. The numbers in parentheses for the best model correspond to the number of past reinforcers and choices, respectively. The total number of parameters for the models including past reinforcers and choices includes the two bias terms.

	Monkey H			Monkey B		
	Number of parameters	*AIC*	Δ*AIC*	Number of parameters	*AIC*	Δ*AIC*
Frequency conditions						
Bias	1	39451	12502	1	32062	11973
+ Reinforcers	17	35901	8952	17	28297	8208
+ Choices	43 (13, 28)	26949	0	42 (14, 26)	20089	0
Magnitude conditions						
Bias	1	46162	14829	1	30184	10757
+ Reinforcers	17	41340	10007	17	25429	6002
+ Choices	38 (16, 20)	31333	0	28 (11, 15)	19427	0

is less informative because the extra parameters are fitting random variations in the data. Selecting the model with the smallest AIC value identifies the model that is the best approximation within the model set to the true underlying process in an information-theoretic sense.

We computed AIC values for a family of models in which we varied the lengths of the reinforcer and choice histories. We varied the lengths up to 50 trials each, and selected the best model according to the minimal AIC value across the 2,500 candidate models. The results are listed in Table 3. Note that only relative differences between AIC values are meaningful; the absolute values associated with any one model are not. The best models are very local, integrating relatively few past reinforcers (average 14 trials) and choices (average 22 trials), which confirms the visual impression provided by Figure 6. Part of the reason that the effects of past choice extends further into the past than the effects of past reinforcers is due to the fact that there are many more choices than there are reinforced choices because subjects were not reinforced on every response. This provides more data with which to fit the choice history coefficients. This is why the confidence intervals for the choice history coefficients are smaller than those for the reinforcer history coefficients in Figure 6.

For comparison, results for two simpler models also are shown in Table 3; one includes only a constant bias, which is similar to a model that simply guesses with a probability of 0.5 on each trial, whereas the other adds the effects of past reinforcers (without past choices). The differences in AIC values in Table 3 represent the degree of evidence in favor of the best-fitting model, and the simpler models were selected to give a sense of the contribution of each model component. The larger the difference the less plausible a model is compared to the best model; values greater than 10 on this scale provide strong support for the model with the smallest AIC value (Burnham & Anderson, 1998). Even though incorporating the effects of past choices into the model increases its complexity, its improved ability to fit the data makes up for the cost of adding these extra parameters.

The AIC measure indicates that the best-fitting model requires relatively few parameters. However, this measure can only determine which model is the best of the set of models considered. Even a set of poorly fitting models will have a ''best'' model, so we checked the goodness-of-fit by asking how well the best model (a) predicts average choice behavior including transitions between reinforcer ratios; and (b) how well it captures fine response-by-response structure.

Figure 7 shows an example of the choice probabilities measured in a single session along with the corresponding probabilities predicted by the best-fitting model. To convey the fit of the model to average choice behavior, Figure 8 shows choice probabilities

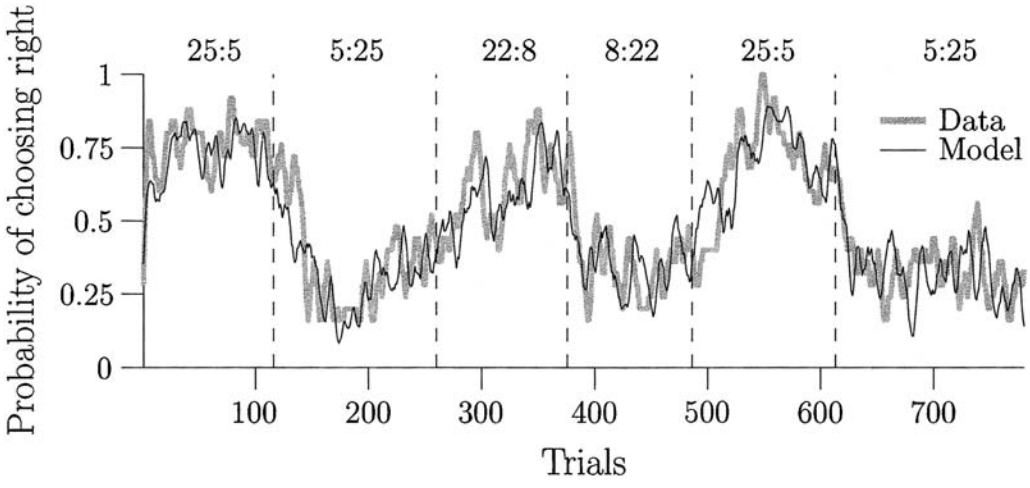

Fig. 7. Predicted and observed choice data for a single session of data from Monkey H (Condition 6). The dashed vertical lines mark the unsignalled block transitions. The data were smoothed with a nine-point moving average.

in a format like Figure 2. Also plotted are predictions from the best-fitting model, which were averaged just like the actual choice data. The model predictions captured the steady-state average as well as the dynamics of the transitions between blocks. There is a misprediction, however, that is obvious only in the extreme transitions for the magnitude conditions (lower right panel in Figure 8). For these transitions, the monkeys were quicker to respond to new reinforcer conditions than the model predicted.

Although Figure 8 reveals that the model accurately captures average behavior in our task, this comparison cannot determine whether the model adequately characterizes fine structure at the response-by-response level because it is an average over many transitions. The runlength analysis above (Figure 5) showed clear structure at the response-by-response level, and we tested whether the model sufficiently characterized the fine structure in behavior by analyzing model residuals, the response-by-response differences between the predicted choice probabilities and the observed choice data. These residuals represent variation in the data not explained by the model, and are useful because inadequate models will yield structured residuals (e.g., Box & Jenkins, 1976).

We used autocorrelation functions to assess whether the residuals showed any response-by-response structure. These functions represent the average correlation between residuals separated by a certain number of trials. A random process that is independent from trial-to-trial, like a fair coin, has zero autocorrelation for any trial separation (a fair coin flipped now does not depend on the outcome of any previous flip). Significantly nonzero autocorrelations in the residuals, therefore, would reveal systematic failures of the model. We estimated autocorrelations by shifting the residuals one trial and computing the correlation between the shifted residuals and the original, repeating this for each of 25 shifts. Figure 9 shows residual autocorrelations for the best-fitting model. The horizontal grey lines represent approximate 95% confidence intervals for autocorrelations expected from an independent random process. For both monkeys, the autocorrelations stayed mostly within these confidence intervals, indicating that the residuals were largely unstructured. For comparison, the thinnest lines in the bottom panels of Figure 9 are the residual autocorrelations from a model that only incorporates reinforcer history (the choice history regressors were excluded). This model is inadequate, and clearly indicates the need to incorporate the effects of past choices to account for the response-by-response structure in our task. Indeed, the structure of the autocorrelation functions for this model reveals the underlying tendency of the monkeys to alternate and persist as shown by the full model (middle panels of Figure 6).

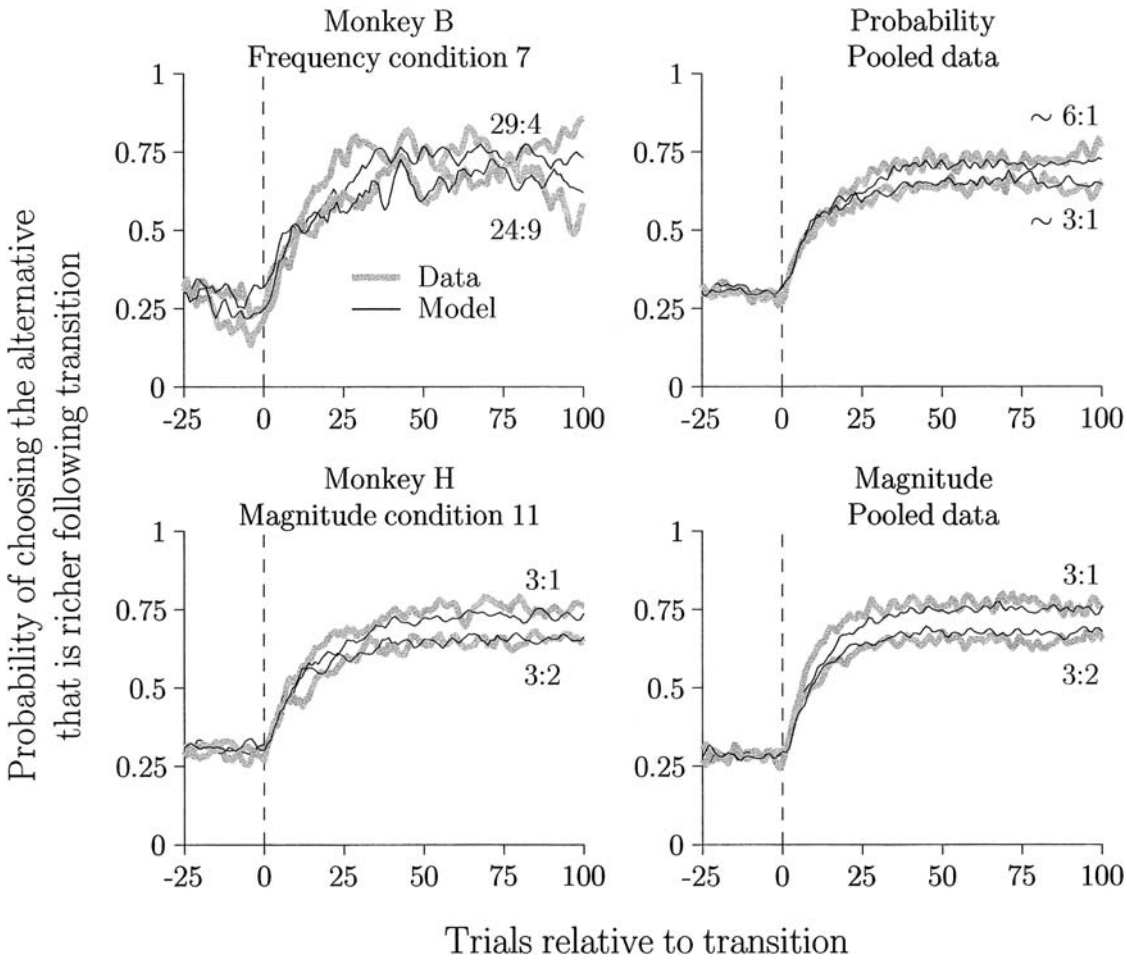

Fig. 8. Predicted and observed choice data aligned on the trial that a transition between ratios occurred. The upper panels are data from conditions where reinforcer frequency was manipulated, and the lower panels are data from conditions where reinforcer magnitudes were manipulated. The left panels show examples from single conditions for each monkey; the right panels show averages across the entire data set. These data were compiled with respect to the two possible posttransition ratios for each condition (see Figure 2 for details). The data were smoothed using a five-point moving average.

Taken together, the analyses suggest that a local *linear* model that includes an effect for past choices accounted for behavior in our task.

DISCUSSION

We studied the choice behavior of monkeys in a task where reinforcement contingencies were varied within a session. We showed that even though the monkeys' choice behavior was variable at a response-by-response level, behavior averaged over tens of trials was a lawful function of reinforcer frequency and magnitude. Monkeys adjusted their choice behavior in response to abrupt reversals in the prevailing reinforcer conditions, reaching an average steady-state of responding after obtaining relatively few reinforcers under the new schedule conditions. Steady-state behavior at the end of each block was well accounted for by the generalized matching law (Baum, 1974). When we varied the relative frequencies of reinforcement, we found that monkeys preferred the rich alternative less than would be expected if they strictly matched the ratio of their choices to the ratio of reinforcers obtained, a finding known as undermatching.

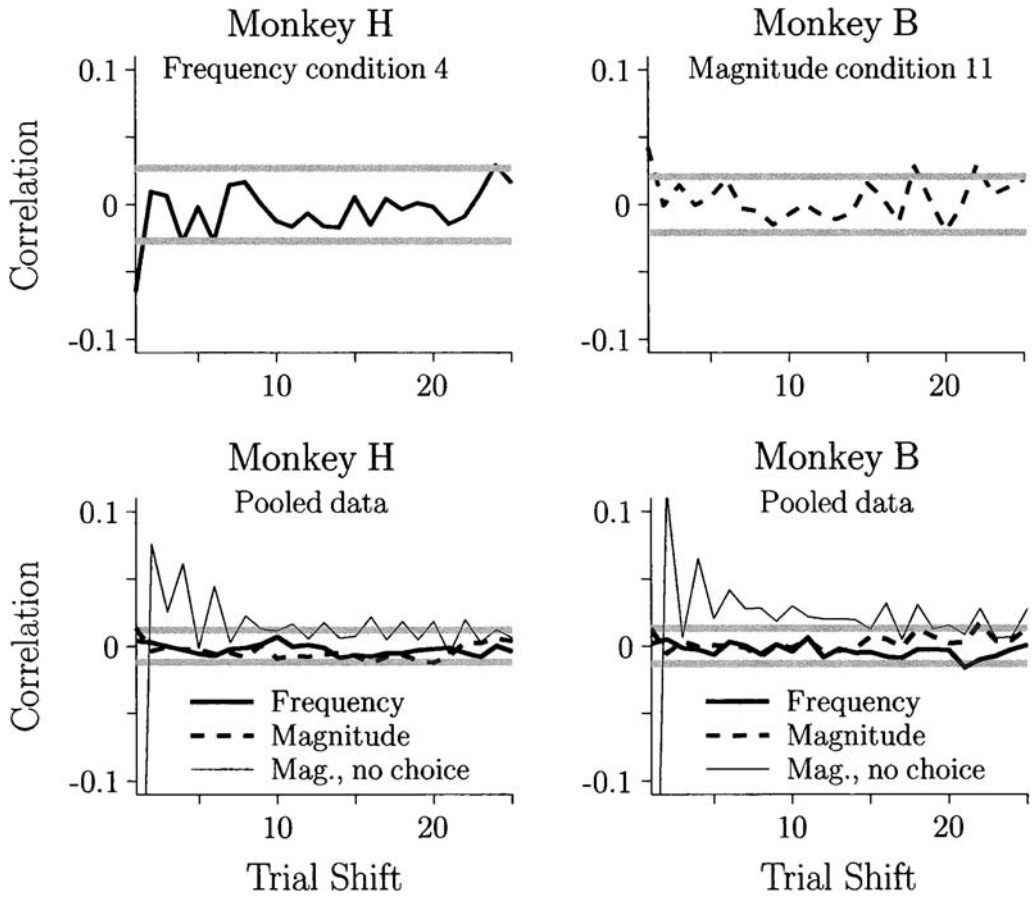

Fig. 9. Autocorrelation functions of model residuals. The thick black line and the dashed line are the autocorrelation functions of the residuals from the best-fitting models (see Table 3) for the probability and magnitude conditions, respectively. The top panels show examples from single conditions for each monkey; the bottom panels show averages across the entire data set. The thin black line in the lower panels is the autocorrelation function of the residuals from a model that does *not* account for the effects of past choices (only shown for magnitude conditions). The gray horizontal bands give approximate 95% confidence intervals for an independent stochastic process.

However, in the magnitude conditions we found that both monkeys matched, or slightly overmatched, their choice allocation to the ratio of obtained reinforcers.

These results may appear at odds with findings in other species, which typically show that animals slightly undermatch for rate of reinforcement and strongly undermatch for magnitude of reinforcement (e.g., Baum, 1979; Davison & McCarthy, 1988). However, like us, Anderson et al. (2002) reported strong undermatching in monkeys performing under standard concurrent VI VI schedules for food reinforcers. This was true despite the fact that their study, unlike ours, used a changeover delay. In addition, we may have observed stronger undermatching because we reversed the reinforcement contingencies many times during a single session. Indeed, strong undermatching is commonly observed under conditions like those we used, such as frequent (Davison & Baum, 2000) or continuous (Palya & Allan, 2003) schedule changes. Regarding the high sensitivity to magnitude, the published data are varied, with observations of near-strict matching (Brownstein, 1971; Keller & Gollub, 1977; Neuringer, 1967) as well as undermatching (Schneider, 1973; Todorov, 1973) being reported.

A principal feature of the choice task we used was the unpredictable and frequent changes in reinforcer frequency and magni-

tude ratios within sessions. Periods of steady-state behavior were preceded by transitions during which the monkeys adjusted their behavior to new reinforcer conditions. Choice behavior began to shift towards the richer alternative after relatively few reinforcers had been delivered in the new reinforcer conditions, despite the fact that block changes were unsignalled and unpredictable. Similarly rapid behavioral adjustments have been observed in response to changes in the frequency (Davison & Baum, 2000; Dreyfus, 1991; Gallistel, Mark, King, & Latham, 2001; Mark & Gallistel, 1994; Sugrue et al., 2004) and magnitude of reinforcement (Davison & Baum, 2003) under concurrent VI VI schedules, as well as under other types of reinforcement schedules (Bailey & Mazur, 1990; Dorris & Glimcher, 2004; Grace et al., 1999; Mazur, 1992).

We found that choice behavior during both steady-state periods at the end of each block as well as in the rapid shifts during transitions between blocks were well characterized by a response-by-response model that linearly combined past reinforcers and past choices to predict current choice. We found a decaying effect of past reinforcers with more recently obtained reinforcers more strongly influencing choice behavior. Our analysis differs from other methods that have been used to examine local reinforcer effects in that the effects of past reinforcers were separated from the effects of past choices (c.f., Davison & Baum, 2000, 2003; Dorris & Glimcher, 2004; Mark & Gallistel, 1994; McCoy, Crowley, Haghighian, Dean, & Platt, 2003; Sugrue et al., 2004). Analyses that do not separate these effects can be more difficult to interpret. Behavior that is independent of reinforcement, such as a tendency to alternate between two actions, may appear to be caused by reinforcement if sequential patterns of choice are not analyzed as well. For example, estimating the likelihood of repeating a choice to an alternative by averaging responses or response ratios for a number of trials following each reinforcer to that alternative often produces oscillatory conditional probabilities (e.g., Bailey & Mazur, 1990; Davison & Baum, 2000; Mazur, 1992), which these authors recognize is due to sequential structure in choice behavior. These oscillations can obscure the structure of the effects due to reinforcers. In our analysis, when the effects of past choices were accounted for, we found that the effects of past reinforcers were highly local, becoming essentially ineffective by the time they have moved 5 to 10 trials into the past (Figure 6). In contrast, excluding the effects of past choices can result in biased estimates of the effects of past reinforcers, which results from this reduced model's attempt to account for the effects of past choices using only responses that yielded reinforcers. We have found that even if alternation is accounted for by incorporating the effects of the last choice, the estimates for the effects of past reinforcers can be biased upwards. In our data, failure to account for the effects of enough past choices resulted in reinforcer effects that remained significant up to 30 trials into the past. Separating the effects of past reinforcers and choices may be of neurobiological importance if the mechanisms of reinforcement learning are distinct from the mechanisms of structuring sequential choice behavior (e.g., Daw, Niv, & Dayan, 2005).

The linear model we used is only one method to separate the effects of past reinforcers from past choices. Another method is to estimate conditional probabilities with respect to time rather than responses in free-operant tasks, which effectively averages out choice structure unrelated to reinforcement (e.g., Davison & Baum, 2002). Alternatively, one could treat specific sequences of choices as units of analysis (Sugrue et al., 2004). Both of these methods also reveal decaying effect of reinforcers, but they are not focused on characterizing the effects of past choices. One advantage of including covariates for past choices in the manner we did is that it allows one to assess quantitatively the *relative* effects of past reinforcers and choices.

By accounting for the effects of past choices we showed that our monkeys have a short-term tendency to alternate as well as a longer-term tendency to persist on the more frequently chosen alternative. Combined with the effects of past reinforcers, this produced a fix-and-sample pattern of choice; there was a tendency to persist on the rich alternative with brief visits to the lean alternative. This sequential pattern of choices was first explored under concurrent VI VI schedules using detailed analyses of runlengths (Baum et al., 1999) and interresponse intervals (Davison, 2004). Moreover, Baum et al. showed that the

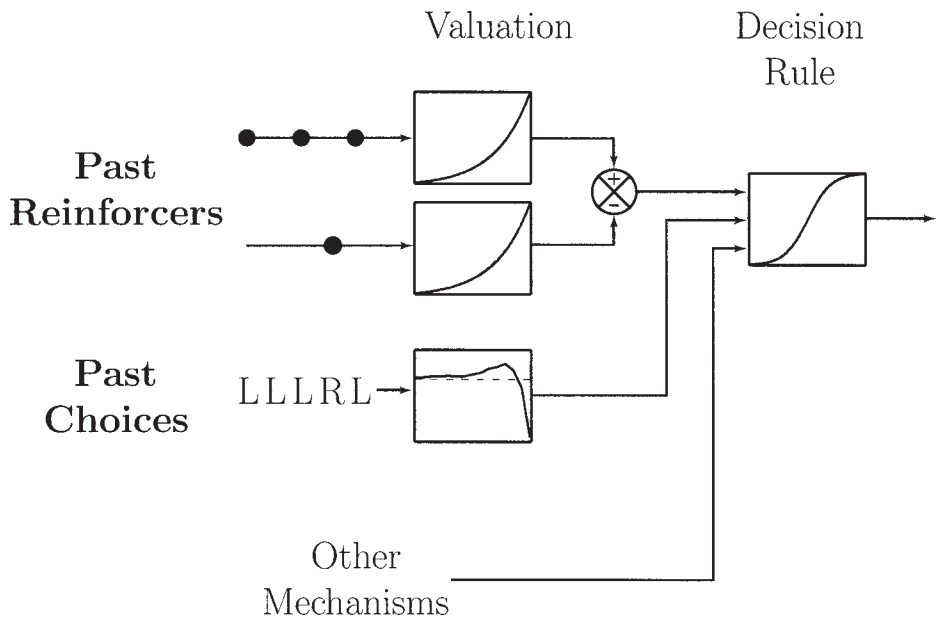

Fig. 10. Two-stage description of choice. The first stage involves valuation of each alternative based on past reinforcer and choices, and possibly other factors like satiety. The second stage generates the choice by comparing the value of each alternative and selecting one using a decision rule.

sequential structure varied depending on presence or absence of a changeover delay; pigeons adjust their patterns of choice behavior to respect the fluctuations in local probability of reinforcement. Similar adjustments have been observed when animals have been trained to exhibit random sequences of behavior (Lee, Conroy, McGreevy, & Barraclough, 2004; Neuringer, 2002). These results show that animals can adjust the degree to which memories of past reinforcers and choices influence behavior. Incorporating choice as a covariate in dynamic response-by-response models allows one to isolate these changes from those due to past reinforcers (Lee et al., 2004; Mookherjee & Sopher, 1994).

Theoretical Models

The statistical model we presented is principally descriptive. However, we can recast it in a different way for comparison with substantive theoretical models. Doing so helps relate our work to neurophysiological work by making clear which variables might be explicitly represented by the nervous system. Figure 10 shows our statistical model separated into two stages. In the first stage, the subjective value of each alternative is estimated from past experience, and in the second stage a decision is formed by comparing these estimates. It is common to split the decision process in this manner (e.g., Davis, Staddon, Machado, & Palmer, 1993; Houston, Kacelnik, & McNamara, 1982; Luce, 1959). As we have sketched it, the first stage includes the effects of past reinforcers, which we think of as providing estimates of the rate of reinforcement. To see this more clearly, we can write the logistic regression in terms of the probability of choice rather than the log odds,

$$p = \frac{1}{1+\exp(\sum \alpha(r_R - r_L))}, \qquad (6)$$

where we ignore past choices momentarily and suppress trial subscripts for clarity. The terms $\Sigma\ \alpha(r_R - r_L)$ encapsulate our assumptions about the internal representations animals form about recently obtained reinforcers. Because the model is linear in the parameters, the coefficients weighting past reinforcers can be rescaled so they sum to 1. This equivalent formulation has a simple interpretation; past reinforcers are weighted to provide local estimates of the difference in average reinforcer rate from each alternative. This estimate has

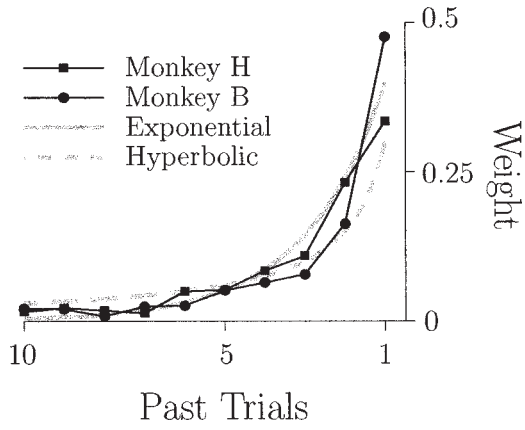

Fig. 11. Comparison of coefficients for past reinforcers with exponential and hyperbolic weightings predicted by theoretical choice models. Each of the four curves has been normalized to unit area. The time constant of the exponential is three trials, and the hyperbolic function is given by the reciprocal of the trial number.

units of reinforcer magnitude per trial, and is passed to the second stage to form a choice probability.

Many theoretical models of foraging and repeated choice behavior assume that the rate of prey encounter or reinforcement is a fundamental variable that animals use to achieve the computational goals of maximizing fitness or reinforcement (Stephens & Krebs, 1986). Rate estimation is usually implemented by a weighted average with exponential decay (Killeen, 1981, 1994), which captures the intuitive notion that because natural environments are nonstationary, recent events should weigh more heavily in memory than long-past events (Cowie, 1977). Alternative weightings have been proposed. For example, Devenport and Devenport (1994) propose a temporal weighting rule that weights past reinforcers hyperbolically (with respect to units of time). In Figure 11, we replot the coefficients for past reinforcers along with examples of exponential and hyperbolic sequences of weights. The variability between subjects precludes determining which particular form of weighting is more appropriate for our data, but the general shape of the reinforcer coefficients supports the idea that animals represent reinforcement rate using algorithms that approximate these linear weightings.

How should we think about the effects of past choices? The decision stage of many choice models uses rules including simply selecting the most valuable alternative (Herrnstein & Vaughan, 1980), matching choice proportions to the local rates of reinforcement (Kacelnik, Krebs, & Ens, 1987), and the logistic function we have used, which is commonly referred to as the softmax (Sutton & Barto, 1998) or logit (Camerer & Ho, 1998) decision rule. None of the models using these decision rules explicitly allows for sequential structure in choice behavior beyond that induced by structure in the obtained reinforcers. This is true for models using an exponentially weighted average because all knowledge about past reinforcers is represented by a single number; this formulation thus possesses an independence-of-path (Bush & Mosteller, 1955; Luce, 1959).

Models incorporating choice dependence are motivated by the nature of the choice task we used. Because the probability of reinforcement grows geometrically with time or responses spent away from the lean alternative in tasks like the one we used, it eventually pays to choose that alternative, but because its probability is reset to its initial low value once chosen it is never worth staying more than one trial. Houston and McNamara (1981) prove that reinforcer rate is maximized by repeating a deterministic sequence; an optimal subject would choose the rich alternative a fixed number of times (dictated by the particular schedules) followed by a single choice of the lean alternative (see also Staddon et al., 1981). Thus, behaving optimally in tasks like ours requires counting the number of choices since the last changeover. Empirical tests of the hypothesis that animals use trial counting to maximize reinforcement rate are equivocal, with evidence for (Hinson & Staddon, 1983; Shimp, 1966; Silberberg et al., 1978) and against (Herrnstein, data published in de Villiers, 1977; Heyman, 1979; Nevin, 1969, 1979) the hypothesis. This leaves the question of optimality unresolved, but the evidence for sequential structure in choice behavior is interesting because models that only estimate reinforcement rates are not designed to generate these patterns. Machado (1993) remedied this by incorporating short-term memory for past responses into his behavioral models. Appropriate use of recent memory for past responses can generate tendencies to shift away from, or to persist on, recently chosen

alternatives (negative recency and positive recency), and is necessary for producing patterned sequences of responses. This captures the idea that one can only remember so much information, but it is distinct from what one does with that memory, which presumably depends on reinforcement contingencies and other factors (Glimcher, 2005; Neuringer, 2002).

Neurophysiology

One of the most interesting uses of the behavioral models described above is that they may provide insight into the underlying neurobiological mechanism of choice, thus allowing tests of the models at the mechanistic level. As a starting point, a literal interpretation of models with the form in Figure 10 makes two predictions. First, certain brain structures may implement a decision rule, functioning to compare the subjective value of each available alternative to generate choice. Second, other brain structures may implement valuation mechanisms, functioning to encode the various components that define value in a particular environment. For an animal foraging for food, evaluating which patch to forage at is a complex combination of estimating the reinforcement rate at each patch, the likelihood that an alternative patch has recently become richer, the risk of predation, and so on. These variables must be represented in the nervous system if they are to affect choice, and neuroscientists are beginning to test these predictions by combining behavioral experiments with techniques such as single-unit electrophysiology and functional magnetic resonance imaging (fMRI).

A number of laboratories have focused on the role of the parietal cortex in the decision process. Single neurons in this cortical region have been shown to encode the accumulation of sensory evidence indicating which alternative will yield a reinforcer (Roitman & Shadlen, 2002; Shadlen & Newsome, 2001; Z. M. Williams, Elfar, Eskandar, Toth, & Assad, 2003) as well as the probability and magnitude of reinforcement (Musallam, Corneil, Greger, Scherberger, & Andersen, 2004; Platt & Glimcher, 1999). Recently, Dorris and Glimcher (2004) and Sugrue et al. (2004) recorded from parietal neurons while monkeys were engaged in choice tasks with variable reinforcement probabilities. They found that the activity of neurons in parietal cortex was correlated with estimates of value derived by fitting models using reinforcement rate to predict the monkeys' behavior (similar to Figure 10, but without taking into account past choices). These results indicate that parietal cortex receives information about both sensory and reinforcing aspects of the environment.

The covariation of neural activity in the parietal cortex with behaviorally relevant variables likely reflects computations carried out in other brain areas. For example, the dopaminergic neurons in the substantia nigra pars compacta (SNC) and the ventral tegmental area (VTA) appear to be important for estimating reinforcement rate. Experiments with monkeys and humans have shown that these neurons encode a *prediction error*, the difference between the reinforcer expected and the reinforcer obtained within a trial (O'Doherty et al., 2004; Schultz, 1998). Bayer and Glimcher (2005) linked SNC activity to reinforcement-rate estimation using the exponentially weighted average described above. They reasoned that if the value of a choice alternative is computed using an exponentially weighted average,

$$v_i = v_{i-1} + \alpha \underbrace{(r_i - v_{i-1})}_{\text{prediction error}},$$

and SNC neurons encoded the prediction errors for these computations, then their responses should be proportional to the difference between the most recent reinforcer (r_i) and an exponentially weighted average of all past reinforcers (v_{i-1}). They showed that the number of action potentials discharged by individual SNC neurons is proportional to the magnitude of the difference between the reinforcer expected (v_{i-1}) and the reinforcer actually obtained (r_i), thereby lending strong support to the hypothesis that dopamine neurons underlie the computation of prediction error. These neurons send dense projections to the prefrontal cortex and the striatum, and current evidence suggests that these areas may use a prediction error to estimate reinforcer rate (Barraclough et al., 2004; Haruno et al., 2004; O'Doherty et al., 2004).

Research is rapidly proceeding in a number of brain areas, but one relatively unexplored aspect is how animals integrate reinforcement history with other aspects of their ongoing behavior. We have shown that monkeys engaged in a choice task appear to integrate both past reinforcers and past choices. Incorporating memory for both of these elements was necessary to account for the structured patterns of choice behavior we observed. This is important for guiding interpretation of the physiological data; characterizing the effects of past reinforcers and choices allows a finer separation of these effects at the neural level. Barraclough et al. (2004) present evidence from a repeated-choice task that some neurons in prefrontal cortex encode the identity of the last response, independent of whether the last response was reinforced or not. This raises the intriguing possibility that frontal cortical areas are important for evaluating what choice to make next in light of what choices were made in the past. This idea is consistent with evidence that frontal cortex and the basal ganglia—which are densely interconnected—are important for the expression and learning of motor sequences (Hikosaka et al., 1999; Tanji, 2001; Tanji & Hoshi, 2001). Physiological recordings from these brain areas within the context of a choice task may reveal their role in the generation of temporally extended patterns of choice behavior.

REFERENCES

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transaction on Automatic Control, 19*, 716–723.

Anderson, K. G., Velkey, A. J., & Woolverton, W. L. (2002). The generalized matching law as a predictor of choice between cocaine and food in rhesus monkeys. *Psychopharmacology, 163*, 319–326.

Bailey, J., & Mazur, J. (1990). Choice behavior in transition: Development of preference for the higher probability of reinforcement. *Journal of the Experimental Analysis of Behavior, 53*, 409–422.

Barraclough, D. J., Conroy, M. L., & Lee, D. (2004). Prefrontal cortex and decision making in a mixed-strategy game. *Nature Neuroscience, 7*, 404–410.

Baum, W. (1974). On two types of deviation from the matching law: Bias and undermatching. *Journal of the Experimental Analysis of Behavior, 22*, 231–242.

Baum, W. (1979). Matching, undermatching, and overmatching in studies of choice. *Journal of the Experimental Analysis of Behavior, 32*, 269–281.

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

Baum, W., Schwendiman, J., & Bell, K. (1999). Choice, contingency discrimination, and foraging theory. *Journal of the Experimental Analysis of Behavior, 71*, 355–373.

Bayer, H., & Glimcher, P. (2005). Midbrain dopamine neurons encode a quantitative reward prediction error signal. *Neuron, 47*, 129–141.

Box, G. E. P., & Jenkins, G. M. (1976). *Time series analysis: Forecasting and control*. (Rev. ed.). San Francisco: Holden-Day.

Brownstein, A. (1971). Concurrent schedules of response-independent reinforcement: Duration of a reinforcing stimulus. *Journal of the Experimental Analysis of Behavior, 15*, 211–214.

Burnham, K. P., & Anderson, D. R. (1998). *Model selection and inference: A practical information-theoretic approach*. New York: Springer.

Bush, R. R., & Mosteller, F. (1955). *Stochastic models for learning*. New York: Wiley.

Camerer, C., & Ho, T. (1998). Experience-weighted attraction learning in coordination games: Probability rules, heterogeneity, and time-variation. *Games and Economic Behavior, 42*, 305–326.

Cowie, R. (1977). Optimal foraging in great tits (parus major). *Nature, 268*, 137–139.

Cox, D. R. (1970). *The analysis of binary data*. London: Methuen.

Davis, D. G., Staddon, J. E., Machado, A., & Palmer, R. G. (1993). The process of recurrent choice. *Psychological Review, 100*, 320–341.

Davison, M. (2004). Interresponse times and the structure of choice. *Behavioral Processes, 66*, 173–187.

Davison, M., & Baum, W. M. (2000). Choice in a variable environment: Every reinforcer counts. *Journal of the Experimental Analysis of Behavior, 74*, 1–24.

Davison, M., & Baum, W. M. (2002). Choice in a variable environment: Effects of blackout duration and extinction between components. *Journal of the Experimental Analysis of Behavior, 77*, 65–89.

Davison, M., & Baum, W. M. (2003). Every reinforcer counts: Reinforcer magnitude and local preference. *Journal of the Experimental Analysis of Behavior, 80*, 95–129.

Davison, M., & Hunter, I. (1979). Concurrent schedules: Undermatching and control by previous experimental conditions. *Journal of the Experimental Analysis of Behavior, 32*, 233–244.

Davison, M., & McCarthy, D. (1988). *The matching law: A research review*. Hillsdale, NJ: Erlbaum.

Daw, N., Niv, Y., & Dayan, P. (2005). Actions, policies, values and the basal ganglia. In E. Bezard (Ed.), *Recent breakthroughs in basal ganglia research* (pp. XX–XX). New York: Nova Science Publishers.

de Villiers, P. (1977). Choice in concurrent schedules and a quantitative formulation of the law of effect. In W. Honig & J. Staddon (Eds.), *Handbook of operant behavior* (pp. 233–287). Englewood Cliffs, NJ: Prentice-Hall.

Devenport, L., & Devenport, J. (1994). Time-dependent averaging of foraging information in least chipmunks and golden-mantled ground squirrels. *Animal Behavior, 47*, 787–802.

Dorris, M., & Glimcher, P. W. (2004). Activity in posterior parietal cortex is correlated with the relative subjective desirability of action. *Neuron, 44*, 365–378.

Dreyfus, L. (1991). Local shifts in relative reinforcement rate and time allocation on concurrent schedules. *Journal of Experimental Psychology: Animal Behavior Processes, 17,* 486–502.

Fahrmeir, L., & Tutz, G. (2001). *Multivariate statistical modelling based on generalized linear models* (2nd ed.). New York: Springer.

Gallistel, C. R., Mark, T. A., King, A. P., & Latham, P. E. (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.

Glimcher, P. W. (2002). Decisions, decisions, decisions: Choosing a biological science of choice. *Neuron, 36,* 323–332.

Glimcher, P. W. (2005). Indeterminacy in brain and behavior. *Annual Review of Psychology, 56,* 25–56.

Grace, R. C., Bragason, O., & McLean, A. P. (1999). Rapid acquisition of preference in concurrent chains. *Journal of the Experimental Analysis of Behavior, 80,* 235–252.

Haruno, M., Kuroda, T., Doya, K., Toyama, K., Kimura, M., Samejima, K., et al. (2004). A neural correlate of reward-based behavioral learning in caudate nucleus: A functional magnetic resonance imaging study of a stochastic decision task. *Journal of Neuroscience, 24,* 1660–1665.

Herrnstein, R. J. (1961). Relative and absolute strength of response as a function of frequency of reinforcement. *Journal of the Experimental Analysis of Behavior, 4,* 267–272.

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

Heyman, G. (1979). Markov model description of changeover probabilities on concurrent variable-interval schedules. *Journal of the Experimental Analysis of Behavior, 31,* 41–51.

Hikosaka, O., Nakahara, H., Rand, M. K., Sakai, K., Lu, X., Nakamura, K., et al. (1999). Parallel neural networks for learning sequential procedures. *Trends in Neurosciences, 22,* 464–471.

Hinson, J. M., & Staddon, J. E. (1983). Hill-climbing by pigeons. *Journal of the Experimental Analysis of Behavior, 39,* 25–47.

Houston, A., Kacelnik, A., & McNamara, J. (1982). Some learning rules for acquiring information. In D. McFarland (Ed.), *Functional ontogeny* (pp. 140–191). London: Pitman Books.

Houston, A., & McNamara, J. (1981). How to maximize reward rate on two variable-interval paradigms. *Journal of the Experimental Analysis of Behavior, 35,* 367–396.

Houston, A., & Sumida, B. (1987). Learning rules, matching and frequency dependence. *Journal of Theoretical Biology, 126,* 289–308.

Hunter, I., & Davison, M. (1985). Determination of a behavioral transfer function: White-noise analysis of session-to-session response-ratio dynamics on concurrent VI VI schedules. *Journal of the Experimental Analysis of Behavior, 43,* 43–59.

Iglauer, C., & Woods, J. (1974). Concurrent performances: Reinforcement by different doses of intravenous cocaine in rhesus monkeys. *Journal of the Experimental Analysis of Behavior, 22,* 179–196.

Judge, S. J., Richmond, B. J., & Chu, F. C. (1980). Implantation of magnetic search coils for measurement of eye position: An improved method. *Vision Research, 20,* 535–538.

Kacelnik, A., Krebs, J., & Ens, B. (1987). Foraging in a changing environment: An experiment with starlings (sturnus vulgaris). In M. Commons, A. Kacelnik, & S. Shettleworth (Eds.), *Quantitative analyses of behaviour, Vol. 6: Foraging* (pp. 63–87). Mahwah, NJ: Erlbaum.

Keller, J. V., & Gollub, L. R. (1977). Duration and rate of reinforcement as determinants of concurrent responding. *Journal of the Experimental Analysis of Behavior, 28,* 145–153.

Killeen, P. R. (1981). Averaging theory. In C. Bradshaw, E. Szabadi, & C. Lowe (Eds.), *Quantification of steady state operant behavior* (pp. 21–34). North Holland, Amsterdam: Elsevier.

Killeen, P. R. (1994). Mathematical principles of reinforcement. *Behavioral and Brain Sciences, 17,* 105–172.

Lee, D., Conroy, M., McGreevy, B., & Barraclough, D. (2004). Reinforcement learning and decision making in monkeys during a competitive game. *Cognitive Brain Research, 22,* 45–58.

Luce, R. D. (1959). *Individual choice behavior; a theoretical analysis.* New York: Wiley.

Machado, A. (1993). Learning variable and stereotypical sequences of responses: Some data and a new model. *Behavioural Processes, 30,* 103–129.

Mark, T. A., & Gallistel, C. R. (1994). Kinetics of matching. *Journal of Experimental Psychology: Animal Behavior Processes, 20,* 79–95.

Mazur, J. E. (1992). Choice behavior in transition: Development of preference with ratio and interval schedules. *Journal of Experimental Psychology: Animal Behavior Processes, 18,* 364–378.

McCoy, A. N., Crowley, J. C., Haghighian, G., Dean, H. L., & Platt, M. L. (2003). Saccade reward signals in posterior cingulate cortex. *Neuron, 40,* 1031–1040.

McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models* (2nd ed.). London; New York: Chapman and Hall.

McDowell, J. J., Bass, R., & Kessel, R. (1992). Applying linear systems analysis to dynamic behavior. *Journal of the Experimental Analysis of Behavior, 57,* 377–391.

Montague, P. R., & Berns, G. S. (2002). Neural economics and the biological substrates of valuation. *Neuron, 36,* 265–284.

Mookherjee, D., & Sopher, B. (1994). Learning behavior in an experimental matching pennies game. *Games and Economic Behavior, 7,* 62–91.

Musallam, S., Corneil, B. D., Greger, B., Scherberger, H., & Andersen, R. A. (2004, July 9). Cognitive control signals for neural prosthetics. *Science, 305,* 258–262.

Neuringer, A. (1967). Effects of reinforcement magnitude on choice and rate of responding. *Journal of the Experimental Analysis of Behavior, 10,* 417–424.

Neuringer, A. (2002). Operant variability: Evidence, functions, and theory. *Psychonomic Bulletin and Review, 9,* 672–705.

Nevin, J. (1969). Interval reinforcement of choice behavior in discrete trials. *Journal of the Experimental Analysis of Behavior, 12,* 875–885.

Nevin, J. (1979). Overall matching versus momentary maximizing: Nevin (1969) revisited. *Journal of Experimental Psychology: Animal Behavior Processes, 5*, 300–306.

O'Doherty, J., Dayan, P., Schultz, J., Deichmann, R., Friston, K., & Dolan, R. J. (2004, April 16). Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science, 304*, 452–454.

Palya, W. (1992). Dynamics in the fine structure of schedule-controlled behavior. *Journal of the Experimental Analysis of Behavior, 57*, 267–287.

Palya, W., & Allan, R. (2003). Dynamical concurrent schedules. *Journal of the Experimental Analysis of Behavior, 79*, 1–20.

Palya, W., Walter, D., Kessel, R., & Lucke, R. (1996). Investigating behavioral dynamics with a fixed-time extinction schedule and linear analysis. *Journal of the Experimental Analysis of Behavior, 66*, 391–409.

Palya, W., Walter, D., Kessel, R., & Lucke, R. (2002). Linear modeling of steady-state behavioral dynamics. *Journal of the Experimental Analysis of Behavior, 77*, 3–27.

Platt, M. L., & Glimcher, P. W. (1997). Responses of intraparietal neurons to saccadic targets and visual distractors. *Journal of Neurophysiology, 78*, 1574–1589.

Platt, M. L., & Glimcher, P. W. (1999). Neural correlates of decision variables in parietal cortex. *Nature, 400*, 233–238.

Roitman, J. D., & Shadlen, M. N. (2002). Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. *Journal of Neuroscience, 22*, 9475–9489.

Schneider, J. (1973). Reinforcer effectiveness as a function of reinforcer rate and magnitude: A comparison of concurrent performances. *Journal of the Experimental Analysis of Behavior, 20*, 461–471.

Schultz, W. (1998). Predictive reward signal of dopamine neurons. *Journal of Neurophysiology, 80*, 1–27.

Schultz, W. (2004). Neural coding of basic reward terms of animal learning theory, game theory, microeconomics and behavioural ecology. *Current Opinion in Neurobiology, 14*, 139–147.

Shadlen, M. N., & Newsome, W. T. (2001). Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. *Journal of Neurophysiology, 86*, 1916–1936.

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

Silberberg, A., Hamilton, B., Ziriax, J. M., & Casey, J. (1978). The structure of choice. *Journal of Experimental Psychology: Animal Behavior Processes, 4*, 368–398.

Staddon, J. E., Hinson, J. M., & Kram, R. (1981). Optimal choice. *Journal of the Experimental Analysis of Behavior, 35*, 397–412.

Staddon, J. E., & Motheral, S. (1978). On matching and maximizing in operant choice experiments. *Psychological Review, 85*, 436–444.

Stephens, D. W., & Krebs, J. R. (1986). *Foraging theory*. Princeton, NJ: Princeton University Press.

Sugrue, L. P., Corrado, G. S., & Newsome, W. T. (2004, June 18). Matching behavior and the representation of value in the parietal cortex. *Science, 304*, 1782–1787.

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

Tanji, J. (2001). Sequential organization of multiple movements: Involvement of cortical motor areas. *Annual Review of Neuroscience, 24*, 631–651.

Tanji, J., & Hoshi, E. (2001). Behavioral planning in the prefrontal cortex. *Current Opinion in Neurobiology, 11*, 164–170.

Todorov, J. (1973). Interaction of frequency and magnitude of reinforcement on concurrent performances. *Journal of the Experimental Analysis of Behavior, 19*, 451–458.

Williams, B. (1988). Reinforcement, choice, and response strength. In R. C. Atkinson, R. J. Herrnstein, G. Lindzey, & R. Luce (Eds.), *Stevens's handbook of experimental psychology*. (2nd ed., pp. 167–244). New York: Wiley.

Williams, Z. M., Elfar, J. C., Eskandar, E. N., Toth, L. J., & Assad, J. A. (2003). Parietal activity and the perceived direction of ambiguous apparent motion. *Nature Neuroscience, 6*, 616–623.

Received September 29, 2004
Final acceptance July 17, 2005