ADAPTIVE CRITERION SETTING IN PERCEPTUAL DECISION MAKING

Maik C. Stüttgen, Ali Yildiz, and Onur Güntürkün

UNIVERSITY OF BOCHUM, GERMANY

Pigeons responded in a perceptual categorization task with six different stimuli (shades of gray), three of which were to be classified as ''light'' or ''dark'', respectively. Reinforcement probability for correct responses was varied from 0.2 to 0.6 across blocks of sessions and was unequal for correct light and dark responses. Introduction of a new reinforcement contingency resulted in a biphasic process of adjustment: First, choices were strongly biased towards the favored alternative, which was followed by a shift of preference back towards unbiased choice allocation. The data are well described by a signal detection model in which adjustment to a change in reinforcement contingency is modeled as the change of a criterion along a decision axis with fixed stimulus distributions. Moreover, the model shows that pigeons, after an initial overadjustment, distribute their responses almost optimally, although the overall benefit from doing so is extremely small. The strong and swift effect of minute changes in overall reinforcement probability precludes a choice strategy directly maximizing expected value, contrary to the assumption of signal detection theory. Instead, the rapid adjustments observed can be explained by a model in which reinforcement probabilities for each action, contingent on perceived stimulus intensity, determine choice allocation.

Key words: optimal choice, signal detection theory, psychophysics, expected value, yes-no task, generalized matching law, key peck, pigeon

―――――――――――――――

Optimal choice in natural environments requires the integration of several sources of information, such as sensory evidence (e.g., distinguishing different food types, or appraising potential mating partners) and knowledge about reinforcer availability (e.g. food repletion at different patches; Pyke, Pulliam, & Charnov, 1977). Signal detection theory (SDT; Green & Swets, 1966) specifies how sensory evidence and knowledge about reinforcer availability and magnitude should be integrated in order to optimize choice allocation relative to a specific decision goal, such as maximizing expected value (e.g., the number of reinforcers attained).

A standard laboratory procedure employing SDT analysis is a psychophysical yes–no task. The subject is repeatedly presented with one of two stimuli (S_1 and S_2), occurring in random succession. The subject's task is to identify each stimulus by making response R_1 if S_1 was presented or R_2 if S_2 was presented. Figure 1a shows the four possible outcomes of this procedure, only two of which are correct with probabilities $p(R_1|S_1)$ and $p(R_2|S_2)$. In most animal psychophysics experiments, correct responses are reinforced while incorrect responses are either not reinforced or punished. Furthermore, the magnitude or frequency of reinforcement is usually identical for both types of correct responses (henceforth referred to as a *balanced* or *symmetrical* payoff matrix as opposed to situations where frequency or magnitude of reinforcement differ between the types of correct responses, henceforth referred to as *unbalanced* or *asymmetrical*). The magnitudes (*values*) of positive reinforcement for correct responses are denoted $V_{R1|S1}$ and $V_{R2|S2}$, the magnitudes of punishment (*costs*) are denoted $C_{R2|S1}$ and $C_{R1|S2}$. While most research employing SDT employs balanced payoff matrices, equal reinforcement for all types of correct responses is the exception rather than the rule in natural environments; taking a foraging animal as an example, some locations may be more likely to provide food, or mating partners, or both, than others. SDT takes this into account by allowing the computation of the optimal allocation of responses as a function of both

a

The source of the variability of the representation is not further specified, but could be conceptualized as drifts of attention or random fluctuations in the sensory transduction process. The distribution of the internal representation of the stimulus is usually assumed to be normal (see Figure 1b for illustration). These distributions are therefore likelihood functions, each denoting the likelihood that a certain value x on the decision axis arose from presentation of its corresponding stimulus. If the task is to discriminate S_1 from another stimulus S_2, the subject is assumed to compare the heights of the two likelihood functions at the location of perceived intensity x on a given trial—the likelihood ratio LR:

$$LR = \frac{l(S_2|x)}{l(S_1|x)} \qquad (1)$$

The decision rule is to respond R_2 when the LR exceeds a threshold β, and to respond R_1 otherwise. In Figure 1b, the two stimulus distributions overlap. For example, an internal value of 4 can arise both from S_1 or S_2 presentation, even though the likelihood (height of the bell-shaped curves at $x = 4$) of S_1 is much higher than the likelihood of S_2. At this point, the observer is bound to make errors, since some values on the decision axis are ambiguous as to the stimulus that gave rise to them.
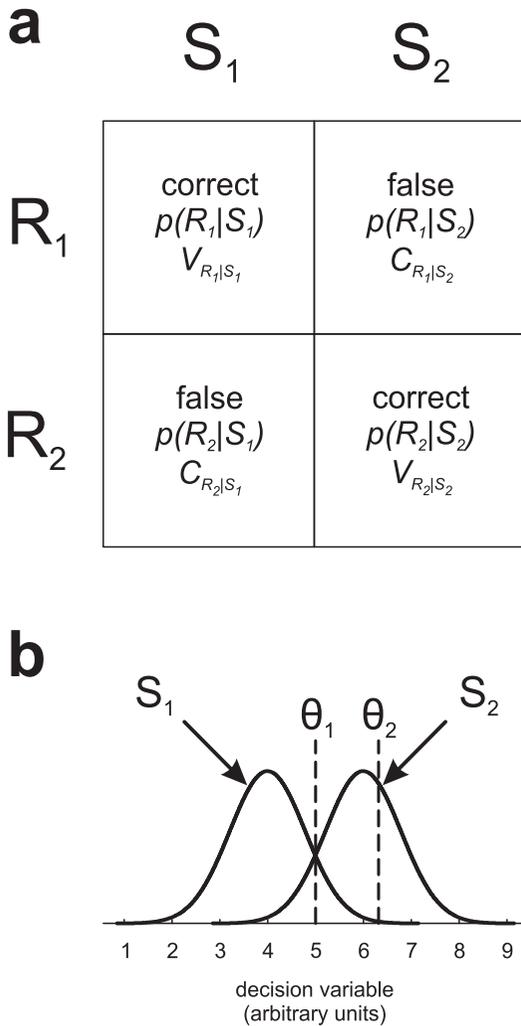
b

Fig. 1. Illustration of signal detection theoretical concepts. (a) Payoff matrix denoting the outcomes of two possible actions, R_1 and R_2, in two possible conditions, presence of stimulus S_1 and presence of stimulus S_2. (b) Presentations of S_1 and S_2 are hypothesized to yield values on an internal decision variable. The observer is assumed to decide which of the two stimuli is present on the basis of an internal decision criterion θ, of which two examples are shown.

SDT proposes that the observer decides on the response to any stimulus on the basis of whether the LR on a given trial exceeds the threshold or not. Decision threshold $θ_1$ in Figure 1b is located right in the middle between the two distributions. At this point where $x = 5$, the likelihoods of S_1 and S_2 are equal, so their ratio is 1. That way, the number of correct S_1 identifications and the number of correct S_2 identifications are equal; the same holds true for the number of incorrect responses to either S_1 or S_2. A decision threshold of β = 1 maximizes overall accuracy (the total fraction of correct responses) without bias for either S_1 or S_2. Hence, a threshold of β = 1 is called neutral or unbiased ($θ_1$ in Figure 1b).

In most signal detection tasks, this neutral decision threshold is also optimal in the sense that it maximizes payoff; that is, receiving the maximum value (from correct responses) while paying the minimum cost (from incor-

sensory uncertainty (asking which stimulus has been presented) and unequal rates of reinforcement (asking which response, if correct, is the more profitable).

SDT assumes that, every time a fixed physical stimulus is presented to an observer, its energy is transformed into a variable internal representation (Boneau & Cole, 1967), the *decision variable*, which can be thought of as perceived stimulus intensity.

rect responses). However, this equality holds only under some specific conditions, such as when the two types of correct responses are equally reinforced, and the two types of errors are equally punished, and both types of signal are presented equally often. The optimal decision threshold β_{opt} is given by

$$\beta_{opt} = \frac{(V_{R_1|S_1} + C_{R_2|S_1}) \times p(S_1)}{(V_{R_2|S_2} + C_{R_1|S_2}) \times p(S_2)} \qquad (2)$$

where $p(S_1)$ and $p(S_2)$ are stimulus presentation probabilities of S_1 and S_2, respectively, and add to 1.

Instead of β, the location of an observer's threshold can be expressed in units of the decision variable. The criterion measure c is related to β by the following equation:

$$\ln \beta = d' c \qquad (3)$$

where d' is the difference between the means of the two distributions, divided by their common standard deviation (MacMillan & Creelman, 2005). A neutral criterion at $\beta = 1$ thus translates to $c = 0$. Accordingly, c can be viewed as the distance of the observed decision criterion from a neutral criterion (in Figure 1b, the threshold at θ_2 corresponds to a c of about 1.3).

With Equation 2, SDT provides a benchmark to evaluate performance of a human or animal subject to optimal performance: the ideal observer. The ideal observer is a hypothetical entity with full knowledge of the stimulus distributions and the values and costs of each possible outcome who places the decision criterion as to maximize a certain decision goal (for our present purposes, this goal is to maximize the total number of attained reinforcers, i.e., expected value).

So far we have considered a balanced payoff matrix: In that case, the numerator and denominator of Equation 2 are equal (assuming equal stimulus presentation probabilities), thus β equals 1 and c equals 0. Now consider a case in which correct S_1 responses yield considerably more reinforcement than correct S_2 responses, that is, $V_{R1|S1} \gg V_{R2|S2}$, with costs identical for both kinds of incorrect responses. In this case, it is desirable to increase the number of correct S_1 responses by moving the decision criterion (e.g., to position θ_2 in Figure 1b). However, this invariably yields a

smaller number of correct S_2 responses, and thereby even less reinforcement for correct S_2 responses in absolute terms. Using decision criterion θ_2 yields almost no errors when S_1 is presented, but more than 50% errors when S_2 is presented. Equation 2 allows us to determine the likelihood ratio, and thus the location of the decision criterion that is statistically optimal to maximize payoff.

Signal detection theory makes some strong assumptions. For example, SDT is limited to the case of the "signal specified exactly", meaning that the nature of the signal, the exact time point of signal occurrence, and the payoff matrix are known to the subject (Stüttgen & Schwarz, 2008; Swets, 1961). While these constraints can be (and frequently are) realized in the laboratory, natural environments are inherently more uncertain. The exact time point, nature, and chance of occurrence of biologically relevant signals such as the sight or sound of a predator are usually not known, and neither is the payoff matrix. Moreover, the payoff matrix of a foraging animal is not stationary over time: some food patches may be unexpectedly pilfered, potential mating partners may have changed territories. Accordingly, animals need to be sensitive to changes in payoff matrices in order to readily adapt their behavior for the maximization of reinforcement. We therefore wondered how animals, when confronted with a signal discrimination problem, adapt to changes in payoff matrices. Although many previous psychophysical studies have manipulated payoff matrices across blocks of experimental sessions (e.g., Alsop & Porritt, 2006; Davison & McCarthy, 1980; Harnett, McCarthy, & Davison, 1984; McCarthy & Davison, 1979; McCarthy & Davison, 1980; McCarthy & Davison, 1984; Nevin, Olson, Mandell, & Yarensky, 1975), usually only steady-state data are reported, that is, performance after the animal has fully adapted its behavior to the changed reinforcement contingencies. Although many studies find that animals behave optimally defined in the sense above (i.e., adopt the decision criterion which maximizes reinforcement; see for example Feng, Holmes, Rorie, & Newsome, 2009; Lea, 1979; Pyke et al., 1977), little is known about the adaptation process after changes in reinforcement contingency, at least for conditional discrimination tasks. However, such data are of interest because reinforcers

affect behavior on a timescale that is not captured by analyses focusing on steady-state behavior, and studies employing concurrent variable-interval (VI) schedules point to the usefulness of analyzing behavioral adaptations to changed reinforcement contingencies on a smaller timescale (e.g., Dreyfus, 1991; Gallistel, Mark, King, & Latham, 2001; Gallistel et al., 2007; Mark & Gallistel, 1994; Mazur, 1995; see Baum, 2010 for review).

We subjected pigeons to a perceptual categorization task with six different stimuli (shades of gray), in which three of the stimuli had to be classified as ''dark'' and ''light'', respectively. We obtained psychometric functions for each session, and varied contingency of reinforcement across blocks of experimental sessions. The payoff matrix was manipulated such that the probabilities of reinforcement for correct dark and light responses were asymmetrical, varying between .2 and .6. We observed the resultant biases in choice behavior and fitted a SDT-based model to the data. We found that the data of all subjects were well described by a model in which an observer shifts an internal decision criterion to maximize payoff, with stimulus distributions unchanged across the entire experiment. After several sessions of exposure to the novel contingencies, performance closely approached optimality within this framework.

METHOD

Subjects

Six pigeons (*Columba livia*), obtained from local breeders and raised in the institute's aviary, served as subjects. Animals were housed individually in wire-mesh cages inside a colony room with a 12-hr dark–light cycle (lights off at 8 p.m.). Water was available at all times, food was restricted to the period of daily testing on workdays, with additional free food available on weekends. During the experiment, the pigeons were maintained at 80–85% of their free-feeding weight. Four of the pigeons were experimentally naive, and 2 others had several months experience on a simple choice task. All subjects were kept and treated according to the German guidelines for the care and use of animals in neuroscience, and the research was approved by a national committee of the State of North Rhine-Westphalia, Germany.

Apparatus

Testing was conducted in an operant chamber. All hardware was controlled by custom-written Matlab code (The Mathworks, Natick, MA; the code is published in Rose, Otto, & Dittrich, 2008). The operant chamber measured 34 cm by 34 cm by 50 cm. On the back wall of the chamber were three translucent response keys (4 cm by 3 cm, bottom height from the floor 19, 20, and 19 cm, required force for activation approximately 25 grams) which could be transilluminated by a flat-screen monitor (ACER AL 1511) mounted against the back wall of the experimental chamber. Each effective key peck produced a feedback click. Food (grain) was provided by a food hopper located below the center key. The chamber was housed in a sound-attenuating shell. White noise was provided at all times to mask extraneous sounds. Sample stimuli were six shades of gray (grayscale values 110, 140, 170, 190, 220, 250, corresponding to illuminances of 22, 35, 49, 59, 76, and 98 lux, respectively). In the following, stimuli with grayscale values of 110, 140, and 170 will be referred to as S_1 or dark, and stimuli with grayscale values of 190, 220, and 250 will be referred to as S_2 or light. Because the precise illuminance values do not matter in this study, we plot behavioral results as a function of gray value rather than illuminance.

Procedure

Figure 2 illustrates the paradigm. At the beginning of each trial, the center key was transilluminated green (initial stimulus), and an alerting sound (1 KHz) was played for 1 s. If the pigeon did not peck within 5 s after initial stimulus onset, the trial was terminated and counted as an ''initialization omission''. Omitted trials were not repeated. Following a single peck on the center key, the sample stimulus was presented for 1 s. Immediately after sample stimulus offset, the center key was again transilluminated green, and another peck was required to turn off the center key and, at the same time, turn on the lateral choice keys. The latter requirement was introduced to make sure that pigeons keep their head in front of the sample key for the whole second of sample presentation in order to prevent the pigeons from moving to the choice keys before the sample stimulus termi-
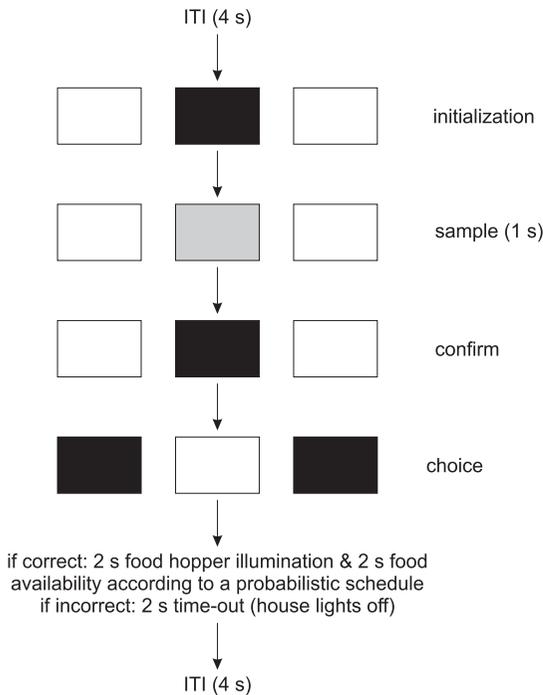
Fig. 2. Schematic of the behavioral paradigm. Sequence of events runs from top to bottom, boxes represent three pecking keys arranged next to each other. After an intertrial interval (ITI) of 4 s, the center key is illuminated green. After a single peck, the center key displays one of six possible sample stimuli (shades of gray) for 1 s. Then, the center key turns green again. After a single peck, the center key is turned off, and the side keys are illuminated orange. The subject has to indicate its decision by pecking either choice key once. If correct, a food hopper is activated for 2 s according to a probabilistic schedule (see Method). If incorrect, all lights are switched off for 2 s (time-out).

nated. The two side (choice) keys were transilluminated orange. If the animal classified a sample stimulus correctly as either light (left choice key) or dark (right choice key), the food hopper was illuminated for 2 s, and, according to a probabilistic schedule, provided 2 s of food access. In case of an incorrect response, all houselights were turned off for 2 s. Stimuli were presented in pseudorandom sequence: A set containing each stimulus type twice was shuffled and presented. This procedure was conducted 25 times, resulting in 300 trials per session. Only trials containing pecks on either choice key entered the analysis. Sessions were conducted daily, usually 5 days per week, and lasted about 45 min each (only one session per day).

Reinforcement probability was the main independent variable in this study. Initially, correct light and dark responses were reinforced with equal probability (gradually decreasing from 1 to .5) to assess baseline performance. After performance had stabilized, the pigeons were exposed to asymmetrical reinforcement probabilities (.6 vs. .3, with half of the animals first being exposed to a S_1/dark-favoring reinforcement schedule of .6 vs. .3, the other half to a S_2/light-favoring reinforcement schedule of .3 vs. .6). Thereafter, biases were switched to .6 versus .2, .2 versus .6, .3 versus .6, and finally again .5 versus .5. Each novel contingency always favored the previously less favorable response. Each asymmetrical reinforcement schedule was maintained for an average of 14.7 consecutive sessions (median: 14.5, minimum: 10, maximum: 22). Table 1 provides descriptive statistics on each animal's experimental schedule. All analyses were done in MATLAB 7.8.0.

RESULTS

The results will be presented in four steps. First, we demonstrate the effect of varying reinforcement contingencies by analyzing steady-state behavior (i.e., averaged over the last five sessions of each condition). Second, we show that responses exhibit a biphasic adjustment process following changes in reinforcement contingency by focusing on individual sessions early in a condition. Third, we develop a SDT-based model that provides an estimate of optimal choice allocation. This estimate serves as a benchmark against which to evaluate each bird's performance. Last, we analyze the data in the framework of the generalized matching law.

Steady-State Behavior

Introduction of asymmetrical reinforcement probabilities strongly affected the pigeons' choice allocation. Figure 3 shows psychometric functions (proportion of left choices per stimulus) for individual birds, averaged across the last five sessions of each of six conditions. With few exceptions, reinforcement schedules favoring S_2 (solid squares and triangles) resulted in higher proportions of left choices, and reinforcement schedules favoring S_1 (open squares and triangles) resulted in lower proportions of left choices. For some birds, the

Table 1

Descriptive statistics for individual pigeons.

	pigeon										
	720		810		919		920		935		947	
no. of sessions in analysis	79		79		80		80		76		87	
Completed trials per session (mean, range)	295 (221–300)		282 (154–300)		296 (235–300)		294 (212–300)		293 (203–300)		282 (181–300)	
% correct per session (mean, range)	85 (58–94)		71 (49–91)		84 (63–93)		85 (67–94)		78 (55–91)		79 (55–90)	
sessions eliminated (>50% omissions)	0		1		0		0		0		4	
order of contingencies [S_1, S_2] and number of test sessions for each contingency	.5l.5	6	.5l.5	6	.5l.5	7	.5l.5	7	.5l.5	7	.5l.5	10
	.3l.6	12	.3l.6	11	.3l.6	12	.6l.3	12	.6l.3	16	.6l.3	17
	.6l.2	18	.6l.2	12	.6l.2	17	.2l.6	12	.2l.6	13	.2l.6	19
	.2l.6	16	.2l.6	13	.2l.6	15	.6l.2	15	.6l.2	18	.6l.2	18
	.6l.3	14	.6l.3	22	.6l.3	15	.3l.6	13	.3l.6	10	.3l.6	13
	.5l.5	13	.5l.5	15	.5l.5	14	.5l.5	21	.5l.5	12	.5l.5	11
mean (median) goodness-of-fit (r^2) for psychometric function[1]	0.997 (0.999)		0.972 (0.989)		0.993 (0.996)		0.995 (0.998)		0.989 (0.995)		0.989 (0.991)	
correlation of thresholds and slopes of psychometric functions[1]	0.394		0.499		0.404		0.529		0.653		0.259	
deviation from optimal criterion	.5l.5	−0.027	.5l.5	−0.017	.5l.5	−0.077	.5l.5	−0.205	.5l.5	−0.458	.5l.5	0.127
	.3l.6	−0.22	.3l.6	−0.45	.3l.6	0.021	.3l.6	−0.123	.3l.6	−0.271	.3l.6	0.064
	.6l.3	−0.198	.6l.3	0.179	.6l.3	−0.061	.6l.3	0.091	.6l.3	−0.272	.6l.3	0.012
	.2l.6	0.19	.2l.6	−1.512	.2l.6	0.106	.2l.6	−0.104	.2l.6	−0.307	.2l.6	0.054
	.6l.2	−0.219	.6l.2	1.485	.6l.2	−0.19	.6l.2	−0.115	.6l.2		.6l.2	−0.22
generalized matching equations and r^2 across all sessions	0.488x − 0.093 r^2=0.866		0.943x + 0.055 r^2=0.808		0.47x + 0.001 r^2=0.874		0.485x − 0.053 r^2=0.879		0.603x + 0.204 r^2=0.932		0.632x + 0.043 r^2=0.888	
generalized matching equations and r^2 for the first 5 sessions of a condition	0.613x − 0.110 r^2=0.956		0.946x − 0.227 r^2=0.921		0.581x − 0.034 r^2=0.944		0.604x − 0.028 r^2=0.943		0.621x − 0.213 r^2=0.964		0.791x + 0.080 r^2=0.963	
generalized matching equations and r^2 for the last five sessions of a condition	0.334x − 0.046 r^2=0.814		1.004x + 0.072 r^2=0.977		0.358x + 0.011 r^2=0.871		0.354x − 0.061 r^2=0.943		0.460x − 0.135 r^2=0.957		0.384x − 0.018 r^2=0.882	
maximum (minimum) beta	12.18 (0.06)		6.56 (0.07)		16.21 (0.06)		17.89 (0.09)		7.39 (0.06)		14.88 (0.05)	

[1] Computations included only values from functions with fitted thresholds between 100 and 260, thus excluding cases where exclusive choice was observed. This applies to 29 out of 79 sessions from pigeon 810, 4 out of 76 sessions from pigeon 935, and 6 out of 88 sessions from pigeon 947.
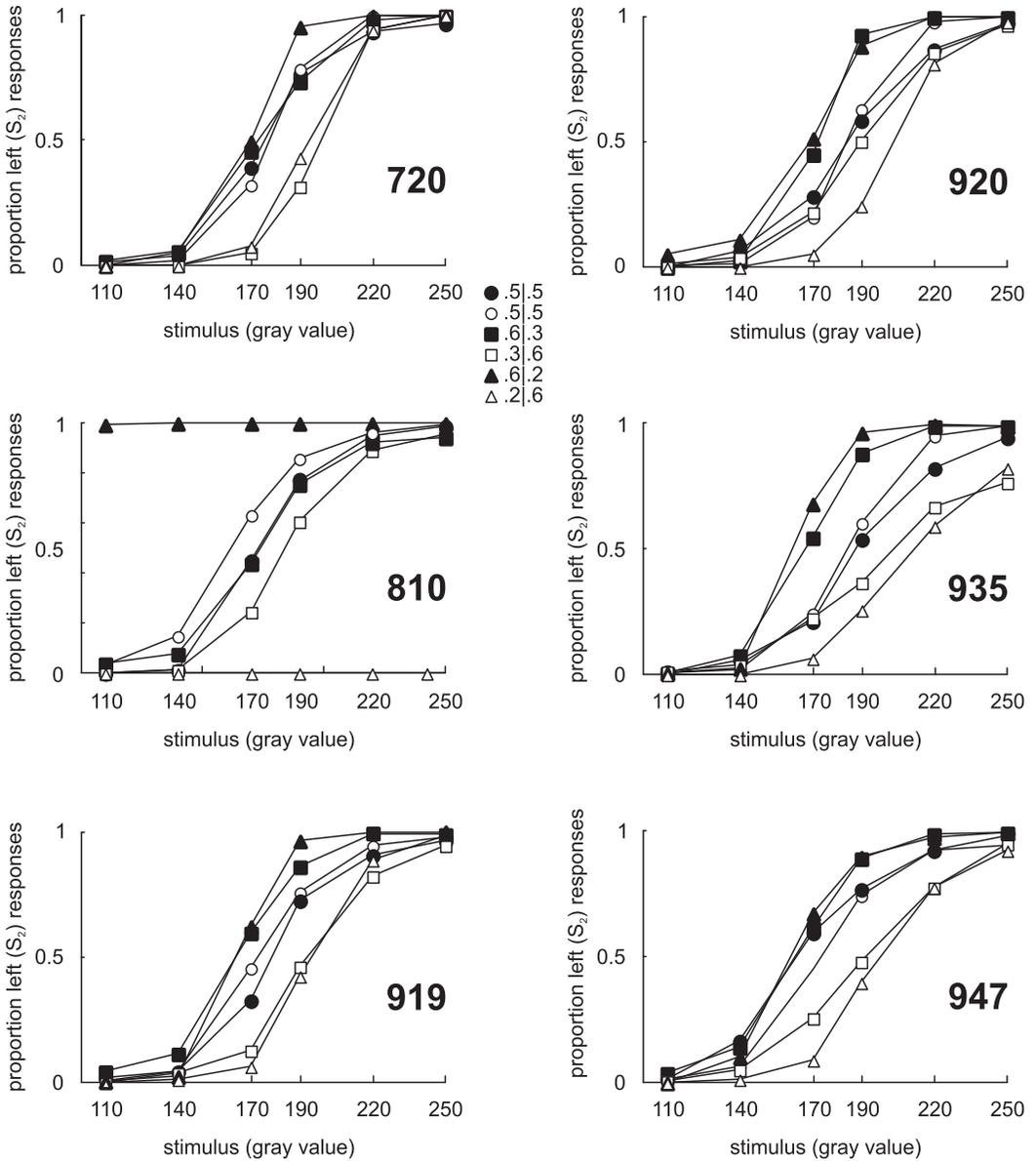
Fig. 3. Mean proportion of left choice responses for the last five sessions of each contingency for individual birds. For the .5|.5 condition, filled circles represent first block, open circles represent last block of experiment.

more extreme reinforcement contingencies (.6 vs. .2, triangles) resulted in more extreme shifts of response proportions than the less extreme contingencies (.6 vs. .3, squares; see, for example, Bird 935). Bird 810 is an extreme case: Here, the more extreme contingencies resulted in exclusive choice of the response with higher reinforcement probability. This general pattern is also visible in the group data, shown in Figure 4. The effect of reinforcement contingency on choice allocation is greatest for stimuli 170 and 190, which were closest to the category boundary (180; compare the variability for stimuli 170 and 190 across conditions to that of 110 and 250), as described previously by Davison and McCarthy (1989).
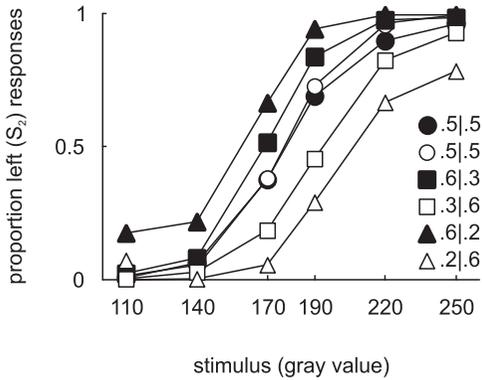
Fig. 4. Mean proportion of left choice responses for the last five sessions of each contingency, averaged over all birds. Conventions as in Figure 3.

Response Changes Following Variations in Reinforcement Contingency

To visualize the dynamics of choice, the psychometric function of each session was fitted with a cumulative Gaussian distribution, with the mean of the function representing threshold and its standard deviation representing slope (that way, larger standard deviation implies shallower slope). Goodness of fit was excellent for nearly all sessions from all birds, with the exception of Bird 810, whose near-exclusive preference in 29 out of 79 sessions prohibited a reasonable fit (see Table 1 for more details).

Figure 5 summarizes the dynamics of choice as changes in psychometric thresholds and slopes across all sessions. The general pattern is that, after the introduction of a new asymmetrical contingency of reinforcement, thresholds rapidly shifted away from the favored stimulus category, implying a larger number of correct responses for that category and, correspondingly, a smaller number of correct responses for the other. Subsequently, thresholds gradually reapproached the category boundary (gray horizontal line). This biphasic pattern of adaptation was particularly pronounced in Birds 919, 920, and 947, but showed up as well in at least two conditions in Birds 720 and 935, with Bird 810 again being the exception (but see condition .6|.3).

Changes in the contingency of reinforcement affected both threshold and slope of the psychometric functions. In fact, thresholds and slopes were positively correlated across all sessions for all animals (see Table 1), with correlations ranging from .26 (Bird 947) to .65 (Bird 935), and an average correlation of .46. This implies a decrease in sensitivity (slope) as threshold increases towards brighter values—in terms of detection theory, d' between neighboring pairs of stimuli decreases with increasing luminance. An interdependence of threshold and slope across different levels of bias induction was previously demonstrated by Davison and McCarthy (1989) in a color discrimination task. They explained the increase in slope with increasing threshold by enhanced sensitivity to wavelength differences when wavelengths become longer (in terms of detection theory, this implies that d' for a given wavelength difference increases with wavelength).

The biphasic adaptation pattern is also visible in group data. Figure 6 plots choice allocation for the first 10 sessions after the introduction of a new reinforcement contingency, separately for each condition. Clearly, changes in choice allocation are more pronounced for the more extreme reinforcement contingencies.

Comparison to Ideal Observer Performance

Shifting the decision criterion in a signal detection or discrimination task from a neutral location can be beneficial when the payoff matrix is asymmetrical (see Introduction). Exact placement of the optimal decision criterion depends on the ratio of reinforcement for the two alternatives and can be derived from a SDT-based model fitted to individual birds' data.

The model fits choice probabilities as arising from six Gaussian distributions on a single decision axis with a variable decision criterion. Decision axis, stimulus location, and criterion location are scaled in units of standard deviations (z-scores). The relative locations of these distributions are modeled separately for each animal, and the decision criterion could assume a different value in each session. Accordingly, differences in the fraction of left responses can only arise from session-wise variations in the decision criterion. The model is explained in more detail in the Appendix.

Figure 7 plots the results of the modeling exercise. For each bird, the left panel shows the relative locations of the six stimulus distributions on the decision axis, superimposed on histograms which depict the fre-
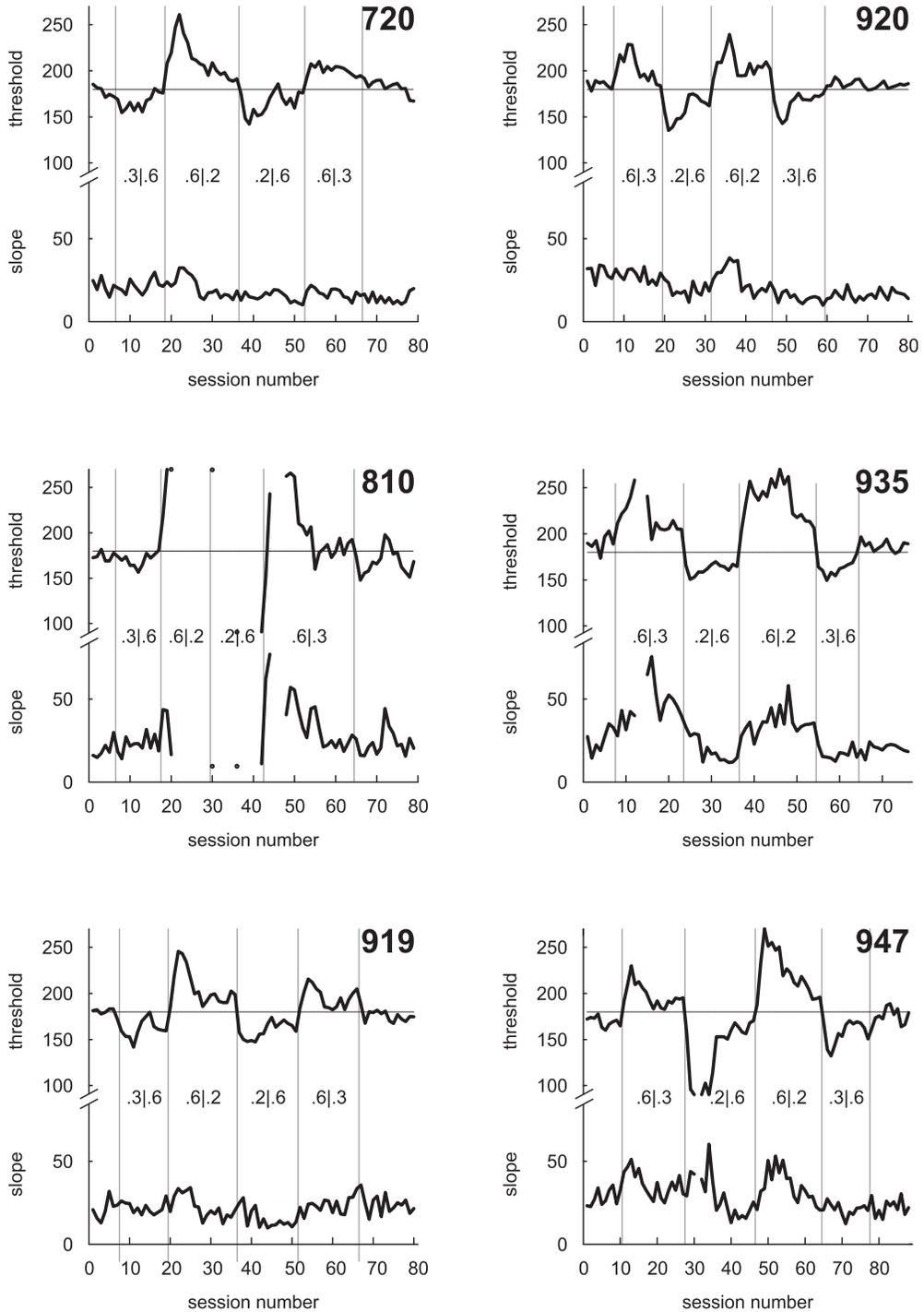
Fig. 5. Changes in threshold and slope across experimental sessions for individual birds. Lines are broken for birds 810, 935, and 947; data for these sessions could not be fitted reasonably well ($r^2 < .65$), and the corresponding data points have been omitted. Vertical gray lines denote changes in reinforcement contingency. Pairs of numbers in the plot indicate reinforcement probabilities for correct responses within one block (S_1 and S_2). The first and last blocks provided equal probabilities of reinforcement (.5) for both stimulus categories. Thin horizontal dotted lines denote unbiased responding.
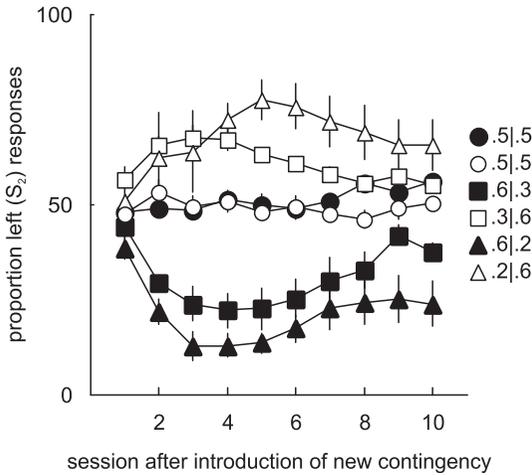
Fig. 6. Changes in choice probability across experimental sessions, averaged across all subjects. Error bars represent the standard error of the mean (SEM). Conventions as in Figure 4.

quency distribution of criterion values for each bird's data. In general, criterion values were unimodally distributed, with most values close to 0 (the unbiased criterion, at which response probabilities are equal). Again, Bird 810 marks an exception: The histogram shows three modes, one at 0, the other two at the extremes of the distribution. It is important to note that the values of the decision criterion and the z-scores are bounded by correcting response ratios of 0 and 1. Such a correction is inevitable with a finite number of stimulus presentations (see Appendix).

Nonetheless, this boundary did not significantly degrade the model's fit to the data. The right panels of Figure 7 show the correlation between the empirically obtained choice probabilities on the abscissa and choice probabilities reconstructed from the model on the ordinate, both expressed as z-scores. The model explained a large portion of the variance in the data: even for Bird 810, r^2 amounted to .84, and the maximum r^2 was .94 for Bird 920.

Assuming the validity of the model, we may now compare the criterion values fitted to the birds' response ratios to those of an ideal observer with the same sensitivity as the pigeons (working with the same internal distributions on the decision axis), but the optimal decision criterion value for each contingency of reinforcement. This was done as follows: The decision criterion was varied

from −5 to +5 in steps of 0.1, and the fraction of correct responses for each stimulus was calculated and multiplied with the reinforcement probability for the respective category of that stimulus for each criterion value. These six products, probability of a correct response × probability of reinforcement for that response, were averaged across stimuli, yielding the expected number of reinforcers per trial (i.e., expected value) for each criterion value. This procedure was repeated for every contingency of reinforcement. We will refer to the dependence of expected value on decision criterion placement as the *objective reward function* (ORF; see Maddox, 2002; sometimes these are also called *molar feedback functions*, Baum, 1981). To distinguish between the decision criterion fitted to each bird's data and the optimal decision criterion, we will refer to these two variables as the *empirical* and *optimal* decision criterion, respectively.

Figure 8 illustrates the relation between the empirical and the optimal decision criteria and the ORFs for individual birds' data. Each panel shows the trajectories of the empirical (bold line) and optimal (thin line) decision criteria over sessions. The grayscale background represents the ORF for each condition (see Figure 9 for another depiction of the ORFs).

The trajectory of the empirical decision criterion values is highly reminiscent of the trajectory of the psychometric thresholds (cf. Figure 5). Comparison of the empirical and the optimal criterion values shows that, within each of the four blocks featuring asymmetrical reinforcement contingencies, the empirical criterion initially overshoots the optimal value, and then gradually reapproaches it over the course of the next few sessions. Relating the trajectory of the empirical decision criterion to the ORF shows that the initial overshoot descends the shallower downward slope of the function. "Shallow" means that moving the decision criterion away from the optimum into this direction by one unit entails a smaller decrease in expected value than moving the criterion in the other direction by the same amount. Accordingly, a change in decision criterion differently affects the amount of reinforcers that will be obtained, depending on the direction of the criterion shift and the initial position of the criterion.

Figure 9 plots the ORF for the different reinforcement contingencies for each animal,

along with the decision criteria averaged over the last five sessions of each condition. Two factors can be appreciated. First, the ORF is highly asymmetrical for conditions with asymmetrical reinforcement contingencies, with plateaus on one side of the neutral criterion. This plateau is more pronounced for conditions with reinforcement probabilities of .6 versus .2. Second, averaged steady-state criterion values are close to the peaks of the ORFs, even for the latter conditions. An exception is again Bird 810, whose values approached optimality in two conditions with probabilities of .6 versus .3, but failed to do so at the more extreme conditions due to the bird's exclusive preference for one choice option. Mean deviations from optimal performance amounted to −0.09, −0.14, −0.04, −0.26, and +0.07 for conditions with reinforcement probabilities of .5|.5, .3|.6, .6|.3, .2|.6, and .6|.2, respectively. Values for individual birds are given in Table 1.

Despite the variability in criterion values (Figure 8), the actual variability in reinforcement density (average number of reinforcers per trial) across sessions was small. Figure 10 plots the ratio of the expected number of reinforcers per trial with the birds' reconstructed criterion values by the expected number of reinforcers per trial obtained by the ideal observer as conceptualized in the SDT model (black line). Even during the initial overshoot phases after conditions were changed, the loss of reinforcers rarely exceeded 5–6%, again with the exception of Bird 810, which showed exclusive preference for the favorable side in condition .6|.2, and failed to adapt to the reversed contingency .2|.6.

It is important to realize that the small losses in reinforcers compared to an ideal observer are not a result of successful and rapid adaptation to the optimal decision criterion. Consider, for example, Bird 810: In Condition .6|.2, this bird maintained exclusive preference for the favorable option for four sessions after contingency reversal (sessions 30–33), thereby losing all reinforcers which could be obtained for correct responses to the other options. Still, this bird attained >90% of reinforcers compared to optimal performance. This is a direct consequence of the flatness of the objective reward function for that contingency as shown in Figure 9.

Furthermore, it can be shown that the birds would have been better off not to adapt to novel contingencies at all: Comparison of the number of reinforcers obtained by an unbiased observer (having the same sensory capacity but maintaining a constant, neutral decision criterion) to those of the ideal observer, shows that the loss of reinforcers per trial never exceeded 3% (Figure 10, gray solid line).

Matching

Figure 11 shows the ratio of pecks (P_{left}/P_{right}) and the ratio of reinforcers (Rf_{left}/Rf_{right}) as a function of session number for each bird. This depiction departs from the familiar visualization of matching, where reinforcer ratios are shown on the abscissa and response ratios on the ordinate, and this helps to visualize the degree of matching for each individual session. It can be seen that there were large fluctuations in both response ratios and reinforcer ratios. Furthermore, response ratios consistently undermatched reinforcer ratios, indicating that the animals worked considerably more for reinforcers from one side compared to the other side. However, there was no apparent trend for the curves to converge (thus showing matching) after introduction of a new contingency: the birds did not equate returns (i.e. reinforcer probabilities per emitted response; Revusky, 1963).

Data from all sessions were fitted with the equation for the generalized matching law (Baum, 1974):

$$\log\left(\frac{P_{left}}{P_{right}}\right) = a \log\left(\frac{Rf_{left}}{Rf_{right}}\right) + \log k \qquad (4)$$

where P_{left} and P_{right} are the relative frequencies of left and right choices, respectively, and Rf_{left} and Rf_{right} are the relative frequencies of reinforcers obtained from left and right responses, respectively. All birds exhibited some degree of undermatching (range: 0.49 to 0.94; see Table 1 for complete data). Fitting Equation 4 separately to data from the first five sessions and the last five sessions of all contingencies revealed that undermatching was more pronounced in the last five sessions (mean slope: 0.48) compared to the first five sessions (mean slope: 0.69).

DISCUSSION

Several studies have related performance in signal detection tasks to the matching law (for example, Alsop & Porritt, 2006; Davison &
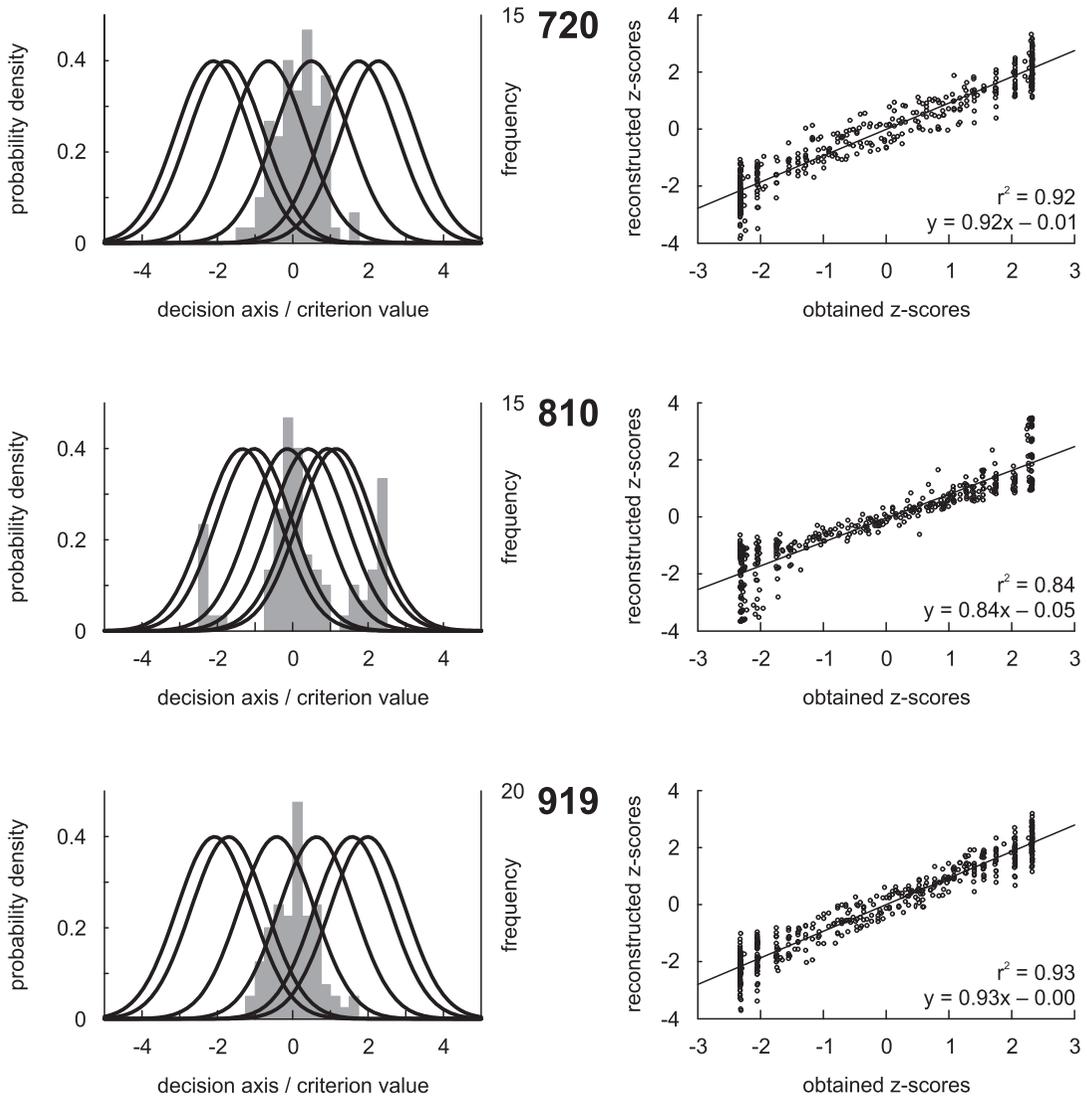
Fig. 7.   Signal-detection-theory-based model applied to the data of each individual pigeon. Left panels show relative locations of six hypothetical stimulus distributions along an internal decision axis. The order of stimulus distributions on the decision axis is perfectly correlated with the order of gray values (left to right, dark to bright). Gray histogram shows distribution of decision criterion values across all sessions as estimated by the model. Right panels show scatterplots of empirical against theoretical fractions of left key pecks across all stimuli and experimental sessions, along with best fitting regression lines, regression equations, and goodness of fit (r^2).

Tustin, 1978; Davison & McCarthy, 1987, 1989; McCarthy & Davison, 1980). An analysis of our data in terms of matching revealed that the animals consistently undermatched; that is, response ratios were always less extreme than reinforcement ratios. Undermatching was more pronounced in the last five sessions of a condition than in the first five sessions (see Figure 11 and Table 1). Slopes for the last five

sessions are similar to those found in studies by McCarthy and Davison, in which animals were exposed to contingencies for more sessions than in the present study (McCarthy & Davison, 1979, 1980).

In conventional matching studies, employing concurrent VI VI schedules, reinforcement ratio does not depend on response ratio in any simple way. If animals emit many more
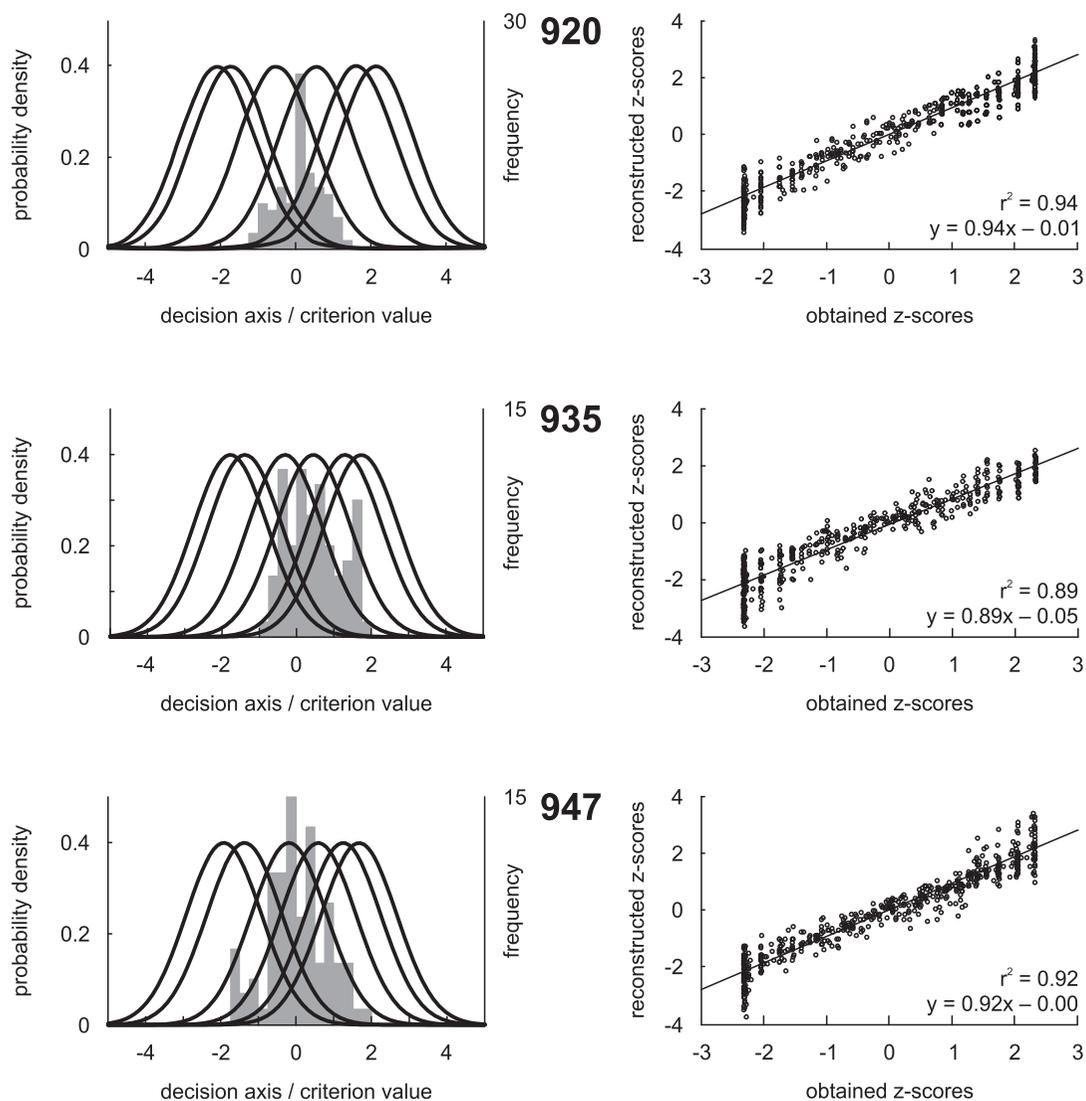
Fig. 7. Continued.

responses than they obtain reinforcers, and switch keys at a high rate, response ratio can vary over a considerable range without having a substantial effect over obtained reinforcement ratio, because almost all scheduled reinforcers are obtained (Herrnstein, 1970; Stubbs & Pliskoff, 1969). In the categorization task with probabilistic reinforcement for correct responses employed in this study, the situation is entirely different: The allocation of responses to different choice options has a direct effect on reinforcement ratio (see Figure 11). In fact, perfect matching (with a slope of 1 in Equation 5) is not possible in our task with asymmetrical reinforcement contingencies, with the exception of exclusive preference for one option. It is difficult to see how the matching framework can account for the present findings—the biphasic adjustment pattern of criterion overshoot and eventual approach to the optimal value—and provide more than a purely descriptive account of choice dynamics. In the following, we will discuss our findings in the light of signal detection theory and extend a decision theoretic model outlined by Boneau and Cole (1967) for a Go–NoGo task to our paradigm.
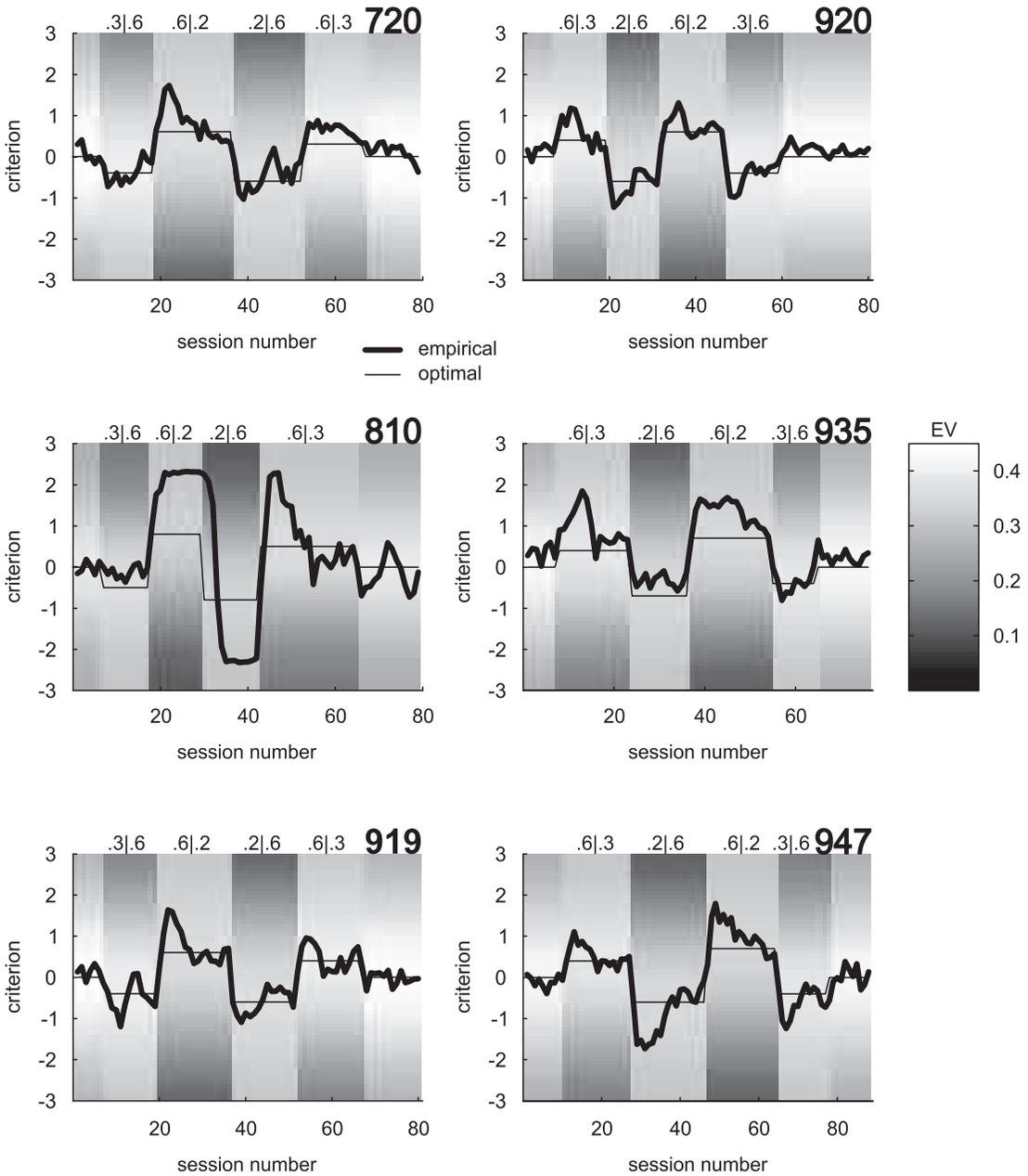
Fig. 8.   Modeled criterion dynamics in relation to reinforcement contingencies and criterion-dependent outcomes for individual birds. Bold lines depict changes in decision criterion experimental sessions, thin solid lines depict optimal placement of decision criteria. Grayscale background represents the objective reward function (expected reinforcers per trial, see colorbar) for each block of sessions (pairs of reinforcement probabilities) and each possible criterion.

The results of our analyses depicted in Figures 9 and 10 suggest that the pigeons, after a period of adjustment, distributed their choices quasioptimally even though this brought about only a small number of additional reinforcers. The surprisingly small changes in payoff that result from comparatively large criterion shifts is due to the flatness of the objective reward functions (Figure 9), which Green and Swets (1966) took to explain that observers in a psychophysical task with asymmetrical payoff matrices stay closer to the
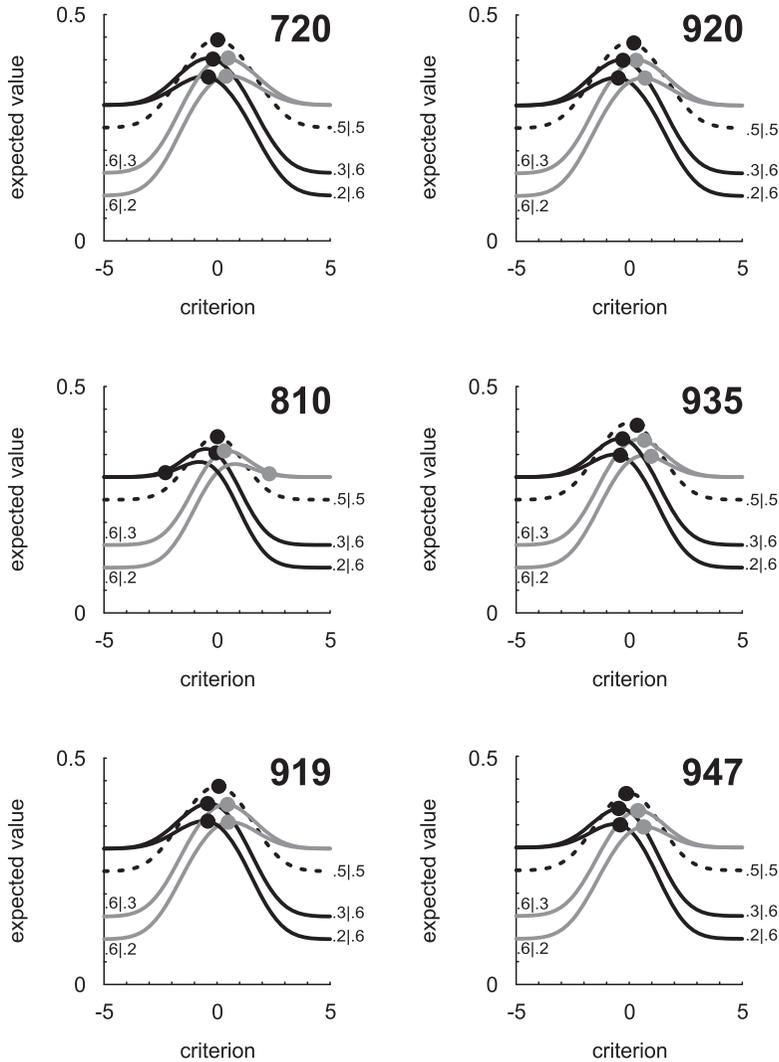
Fig. 9. Feedback functions and steady-state criterion placement for individual birds. Each panel depicts five functions, one for each contingency of reinforcement, relating criterion placement to expected payoff (reinforcers per trial). Dotted line represents symmetrical reinforcement probabilities, solid gray lines represent conditions favoring S_1, solid black lines represent conditions favoring S_2. Dots on each curve depict criterion values averaged over the last five sessions of each contingency.

unbiased decision criterion (which maximizes accuracy but, in this case, not expected value) than the ideal observer. A similar case was put forward by Maddox (2002). However, here we observed the exact opposite: At least during the initial phase of adjustment, criterion values were considerably more extreme than optimal values.

A potential reason for this overshoot may be that the differential choice allocation to options differing in reinforcement density is dependent on the discriminability of these reinforcement densities—in effect, a psychophysical problem: the discrimination of marginally different reinforcement frequencies for the two options. Assuming a constant difference limen for discrimination of reinforcement densities across the range depicted in Figures 8 and 9, the determination of which of two neighboring criterion values is better is much harder along the shallower than the steeper slope of the objective reward function.
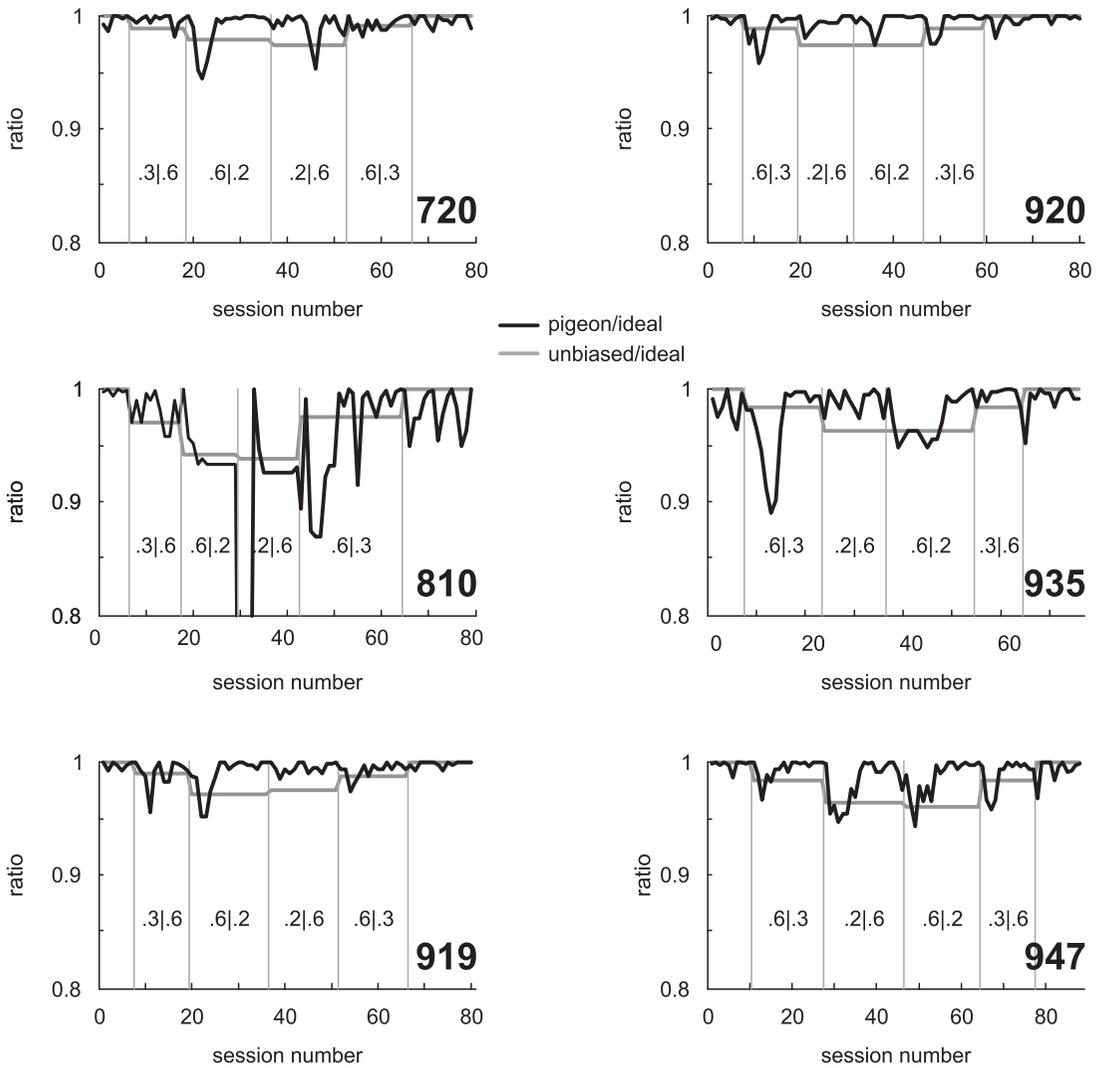
Fig. 10.   Foraging efficiency of individual birds, calculated as the expected total number of reinforcers attained with criterion values modeled for each bird relative to the expected number of reinforcers attained by an ideal observer (black line). Gray lines show the expected number of reinforcers attained by an unbiased observer having the same modeled sensitivity as each bird, divided by the expected number of reinforcers attained by an ideal observer with identical sensitivity.

This is consistent with bias being more extreme for the .6|.2 reinforcement contingencies (Figure 6).

Statistical decision theory holds that ideal observers adjust their decision criterion in such a way as to maximize overall expected value. While the birds indeed approached optimal behavior defined in that sense, the finding of relatively constant effective reinforcer densities across sessions and contingencies, despite large changes in response bias, begs

the question of whether it is really overall expected value that determines choice allocation. Alternatively, quasioptimal behavior may arise as a by-product of some other choice strategy.

To illustrate this point, consider the following: in raw numbers, the birds attained roughly 100 reinforcers per 45-min session. For a 5% loss, the reinforcement density drops from about 133 to 127 reinforcers per hour, corresponding to a drop of overall expected

value from $0.\bar{3}$ to $0.31\bar{6}$. How quickly could such a change be detected? The answer depends on the details of the algorithm that aggregates relative reinforcement frequencies, but the problem of change detection can be illustrated by a simple Bernoulli process. Assuming a null hypothesis of $EV_{null}=0.\bar{3}$ for event A and a true probability of $EV_{true}=0.31\bar{6}$, the null hypothesis can be rejected with 99% confidence after, on average, 5,368 trials. In contrast, a drop in expected value from 0.6 to 0.3, as happened in the actual experiment for a single response option, is detectable within only 30 trials. This suggests that changes in the expected value of each response option rather than changes in overall expected value drive dynamic choice allocation.

Boneau and Cole (1967) have presented a model which, with some modifications, may explain how the animals distribute their choices. Boneau and Cole noted that, in contrast to the ideal observer postulated in signal detection theory, an organism in a signal detection task cannot know the number or shape of the stimulus distributions on the internal decision axis (see Figure 12a, gray lines). Instead, the animal experiences only the sum of these distributions (Figure 12a, bold black line). Still, the animal can learn that responses to one key are more (or less) likely to yield reinforcement in the presence of one stimulus rather than in the presence of another. For a given perceived stimulus intensity (a value on the decision axis, here denoted λ), the animal can learn the probability of reinforcement after S_1 and S_2 choices. This corresponds to the probability of $p(Rf|S_1)$ * $p(S_1|\lambda)$ if the animal responded with R_1 and $p(Rf|S_2)$ * $p(S_2|\lambda)$ if the animal responded with R_2. These terms will henceforth be called ''action values'', since they represent the probability of reinforcement of an action, R_1 or R_2, conditional on correct classification of a perceived stimulus, λ. Figure 12b presents R_1 and R_2 action values as a function of λ (black and gray line, respectively), when the probability of reinforcement is equal for both types of correct responses. It is assumed that the organism always chooses the response which has the highest action value for a given perceived stimulus λ. In this formulation, the decision criterion is the point at which both action values are equal (dotted vertical line). Figure 12c depicts action values when the

reinforcement probability for correct R_1 responses is twice as high as for correct R_2 responses. In this case, the intersection of action values (the decision criterion) has moved to the right. Importantly, the decision criterion in this model—intersection of action values—is identical to the decision criterion in classical signal detection theory (see Boneau & Cole, 1967) and therefore, statistically optimal in the sense that expected value is maximal at this point.

To connect this model to the problem of change detection, consider what happens when symmetrical reinforcement contingencies (Figure 12b) are replaced by asymmetrical contingencies (Figure 12c), favoring S_1. Action values are continuously updated and thus affected immediately by each outcome; thus, R_2 action value will decrease as the relative rate of reinforcement for S_2 for each possible value of λ drops, and R_1 action value will increase. However, R_2 action value would, at first, decrease only for those values of λ experienced after the change of contingency. Assuming some generalization of action value along the abscissa (affecting action values for neighboring λ), the decision criterion would be affected earlier by values of λ close to the criterion than by those further away—a prediction easily amenable to empirical test. The speed of criterion change depends on parameters that have still to be worked out; potential candidates are learning rate and prediction error, as used in the Rescorla-Wagner model (Rescorla & Wagner, 1972).

The Boneau-Cole framework presents a possible solution to the problem of change detection: Choice may be driven by action values instead of overall expected value, as postulated in SDT. In addition, this model also makes some interesting and testable predictions: For example, when reinforcement probability is changed not for an entire response option, but only for a single stimulus, the speed of criterion adjustment should be dependent on the perceptual distance of this stimulus to the decision criterion. In the present case, changing reinforcement probability for gray values of 170 and 190 should result in a more immediate adaptation than changing reinforcement for gray values of 110 and 250.

Optimal decision criterion setting is not an infrequent finding in animal psychophysics
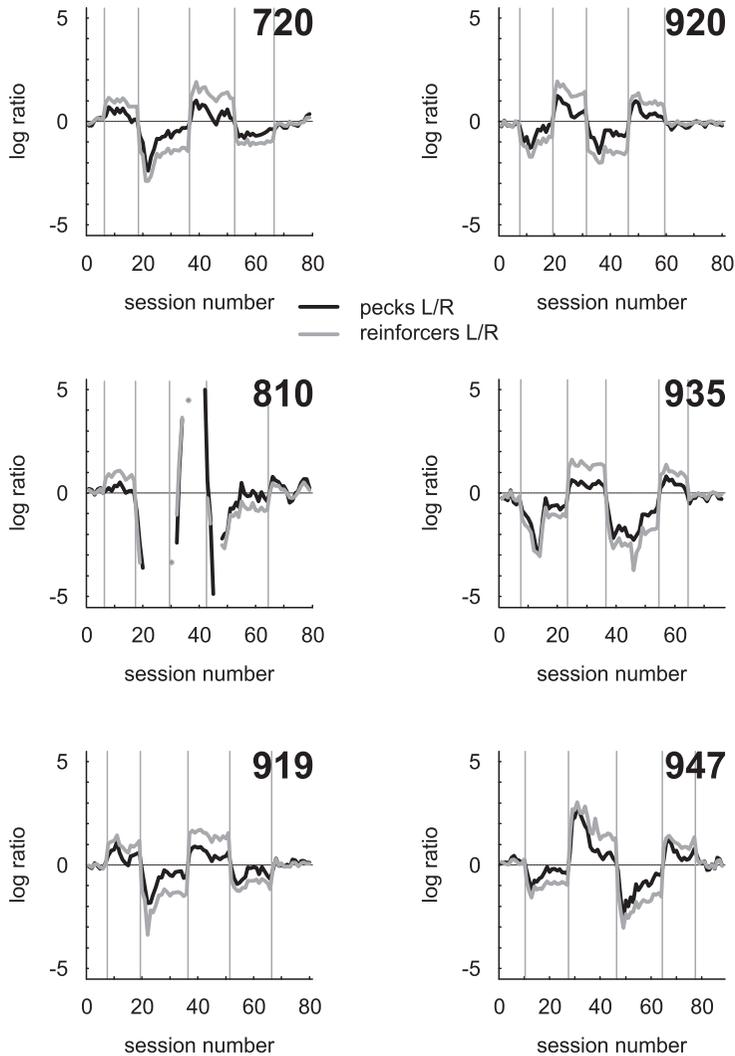
Fig. 11.   Response ratios consistently undermatch reinforcer ratios. Panels show the logarithm of ratios of left and right responses (black lines) and ratios of reinforcers obtained from responding left and right (gray lines). Absolute values for the latter are consistently larger than for the former, indicating undermatching. Missing data points for Bird 810 result from exclusive preference for one option, precluding the calculation of meaningful ratios.

(e.g. Feng et al., 2009) and other tasks (e.g. relative risk assessment; Balci, Freestone, & Gallistel, 2009). A recent study by Rorie, Gao, McClelland, & Newsome (2010) employed a perceptual categorization paradigm similar to ours, in which macaque monkeys were confronted with asymmetrical payoff schedules. However, these authors changed the reinforcement contingencies randomly across trials rather than sessions, with the actual contingency being signaled to the animal by a cue at the beginning of each trial. Accordingly, these

investigators could not investigate the dynamics of decision criterion setting as we could with our blocked design. Interestingly though, a theoretical analysis of the monkeys' behavioral data revealed that the animals tended to use thresholds biased toward the shallower objective reward function was more shallow (Feng et al., 2009). This is identical to our finding that pigeons initially adjusted their decision criterion beyond the point yielding maximum reinforcement, onto the shallower side of the objective reward surface.

a

b

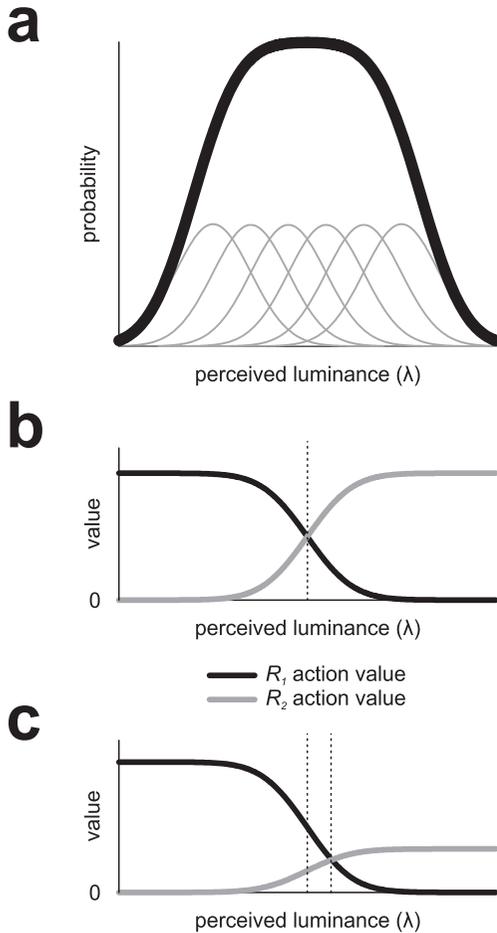— R_1 action value
— R_2 action value

c

Fig. 12. Outline of decision-theoretic model, based on Boneau and Cole (1967). (a) Discriminal distributions (gray lines) for six stimuli equidistant in perceptual space and their sum (bold black line). (b) S_1 and S_2 action values as functions of λ when reinforcement probabilities are equal. (c) S_1 and S_2 action values as functions of λ when reinforcement probability for S_1 is twice as large as for S_2.

Teichert and Ferrara (2010) employed a task in which monkeys had to categorize the speed of a moving random dot pattern as either fast or slow. They embedded a block with asymmetrical reinforcement contingencies (>600 trials) in two conditions with symmetrical contingencies (200–400 trials) in each session. Monkeys overadjusted their decision criterion such that they chose the favorable alternative more often than dictated by optimality. This is identical to our finding of an initial criterion overshoot (Figure 8). It is tempting to speculate that Teichert and

Ferrara's experiment missed the subsequent reapproach of criterion towards the optimal value because they restricted measurement in the biased condition to a single block of trials. In our experiment, the animals did not perform quasioptimally until about the eighth session; accordingly, they required a minimum of about 2,400 trials, which is considerably more than the minimum of 600 trials per block used in the study by Teichert and Ferrara.

Previous studies on free operant choice in which reinforcer ratios were varied either concentrated on steady-state behavior (see Introduction), or analyzed behavior for shorter periods of time, so that they could have missed the biphasic pattern of adaptation. Corrado, Sugrue, Seung, and Newsome (2005) subjected monkeys to a matching task (see also Sugrue, Corrado, & Newsome, 2004), with reinforcement contingencies changing several times per session. While the monkeys adjusted rapidly to the changes, the pattern of adaptation rather resembled a gradual approach to a new equilibrium with no evidence of an initial overshoot. The same holds true for other studies with similar tasks (Davison & Baum, 2000; Lau & Glimcher, 2005; Mazur, 1995). Another reason why these authors did not find the overadaptation effect is that it may become less pronounced as contingencies are switched more and more often. The speed of the adaptation process probably depends on an animal's global experience; when changes are more frequent, adaptation is more rapid (Dreyfus, 1991; Mark & Gallistel, 1994). In the dynamic matching tasks mentioned, reinforcement contingencies were changed several times per session, while our pigeons experienced only five changes in total, with each reinforcement contingency in effect for at least two weeks.

We know of no quantitative model that accounts for the pattern of dynamic choice allocation we observed. Such a model will have to account for this effect, as well as for the speed of choice allocation observed. We take our results to suggest that overall reinforcement density is unlikely to be the variable controlling adaptive choice allocation; instead, choice allocation may be driven by action values as specified in the Boneau-Cole (1967) model.

REFERENCES

Alsop, B., & Porritt, M. (2006). Discriminability and sensitivity to reinforcer magnitude in a detection task. *Journal of the Experimental Analysis of Behavior, 85*, 41–56.

Balci, F., Freestone, D., & Gallistel, C. R. (2009). Risk assessment in man and mouse. *Proceedings of the National Academy of Sciences of the United States of America, 106*, 2459–2463.

Baum, W. M. (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. M. (1981). Optimization and the matching law as accounts of instrumental behavior. *Journal of the Experimental Analysis of Behavior, 36*, 387–403.

Baum, W. M. (2010). Dynamics of choice: a tutorial. *Journal of the Experimental Analysis of Behavior, 94*, 161–174.

Boneau, C. A., & Cole, J. L. (1967). Decision theory, the pigeon, and the psychophysical function. *Psychological Review, 74*, 123–135.

Brown, G. S., & White, K. G. (2005). The optimal correction for estimating extreme discriminability. *Behavior Research Methods, 37*, 436–449.

Corrado, G. S., Sugrue, L. P., Seung, H. S., & Newsome, W. T. (2005). Linear-nonlinear-Poisson models of primate choice dynamics. *Journal of the Experimental Analysis of Behavior, 84*, 581–617.

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., & McCarthy, D. (1980). Reinforcement for errors in a signal-detection procedure. *Journal of the Experimental Analysis of Behavior, 34*, 35–47.

Davison, M., & McCarthy, D. (1987). The interaction of stimulus and reinforcer control in complex temporal discrimination. *Journal of the Experimental Analysis of Behavior, 48*, 97–116.

Davison, M., & McCarthy, D. (1989). Effects of relative reinforcer frequency on complex color detection. *Journal of the Experimental Analysis of Behavior, 51*, 291–315.

Davison, M. C., & Tustin, R. D. (1978). The relation between the generalized matching law and signal-detection theory. *Journal of the Experimental Analysis of Behavior, 29*, 331–336.

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

Feng, S., Holmes, P., Rorie, A., & Newsome, W. T. (2009). Can monkeys choose optimally when faced with noisy stimuli and unequal rewards? *PLoS Computational Biology, 5*.

Gallistel, C. R., King, A. P., Gottlieb, D., Balci, F., Papachristos, E. B., Szalecki, M., & Carbone, K. S. (2007). Is matching innate? *Journal of the Experimental Analysis of Behavior, 87*, 161–199.

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.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

Harnett, P., McCarthy, D., & Davison, M. (1984). Delayed signal-detection, differential reinforcement, and short-term memory in the pigeon. *Journal of the Experimental Analysis of Behavior, 42*, 87–111.

Herrnstein, R. J. (1970). On the law of effect. *Journal of the Experimental Analysis of Behavior, 13*, 243–266.

Lau, B., & Glimcher, P. W. (2005). Dynamic response-by-response models of matching behavior in rhesus monkeys. *Journal of the Experimental Analysis of Behavior, 84*, 555–579.

Lea, S. E. G. (1979). Foraging and reinforcement schedules in the pigeon: Optimal and non-optimal aspects of choice. *Animal Behaviour, 27*, 875–886.

MacMillan, N. A., & Creelman, C. D. (2005). *Detection theory: a user's guide*. New Jersey: Lawrence Erlbaum Associates, Inc.

Maddox, W. T. (2002). Toward a unified theory of decision criterion learning in perceptual categorization. *Journal of the Experimental Analysis of Behavior, 78*, 567–595.

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

Mazur, J. E. (1995). Development of preference and spontaneous recovery in choice behavior with concurrent variable-interval schedules. *Animal Learning & Behavior, 23*, 93–103.

McCarthy, D., & Davison, M. (1979). Signal probability, reinforcement and signal-detection. *Journal of the Experimental Analysis of Behavior, 32*, 373–386.

McCarthy, D., & Davison, M. (1980). Independence of sensitivity to relative reinforcement rate and discriminability in signal detection. *Journal of the Experimental Analysis of Behavior, 34*, 273–284.

McCarthy, D., & Davison, M. (1984). Isobias and alloiobias functions in animal psychophysics. *Journal of Experimental Psychology: Animal Behavior Processes, 10*, 390–409.

Nevin, J. A., Olson, K., Mandell, C., & Yarensky, P. (1975). Differential reinforcement and signal detection. *Journal of the Experimental Analysis of Behavior, 24*, 355–367.

Pyke, G. H., Pulliam, H. R., & Charnov, E. L. (1977). Optimal foraging: selective review of theory and tests. *Quarterly Review of Biology, 52*, 137–154.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black, & W. F. Prokasy (Eds.), *Classical conditioning II: current research and theory* (pp. 64–99). New York: Appleton-Century-Crofts.

Revusky, S. H. (1963). A relationship between responses per reinforcement and preference during concurrent VI VI. *Journal of the Experimental Analysis of Behavior, 6*, 518.

Rorie, A. E., Gao, J., McClelland, J. L., & Newsome, W. T. (2010). Integration of sensory and reward information during perceptual decision-making in lateral intraparietal cortex (LIP) of the macaque monkey. *PloS One, 5*.

Rose, J., Otto, T., & Dittrich, L. (2008). The Biopsychology-Toolbox: A free, open-source Matlab-toolbox for the control of behavioral experiments. *Journal of Neuroscience Methods, 175*, 104–107.

Stubbs, D. A., & Pliskoff, S. S. (1969). Concurrent responding with fixed relative rate of reinforcement. *Journal of the Experimental Analysis of Behavior, 12*, 887–895.

Stüttgen, M. C., & Schwarz, C. (2008). Psychophysical and neurometric detection performance under stimulus uncertainty. *Nature Neuroscience*, *11*, 1091–1099.

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

Swets, J. A. (1961). Detection theory and psychophysics: A review. *Psychometrika*, *26*, 49–63.

Teichert, T., & Ferrara, V. P. (2010). Suboptimal integration of reward magnitude and prior reward likelihood in categorical decisions by monkeys. *Frontiers in Neuroscience*, *4*, 1–13.

APPENDIX

Deriving Optimal Response Allocation with a Signal Detection Model

The effect of asymmetrical reinforcement contingencies on choice can be examined in the framework of signal detection theory. As outlined in the introduction, shifting the decision criterion in a signal detection or discrimination task can be beneficial when the payoff matrix is asymmetrical. The amount by which the decision criterion should shift to be optimal depends on the ratio of reinforcement for the two alternatives (for the two-response case, the optimal decision rule is given by equation 2). To determine the effectiveness of the choice biases observed in our data in a six-stimulus conditional discrimination task, we fitted a signal-detection-theory-based model to the psychometric data of individual birds and determined the optimal relative choice ratios.

The model assumes the existence of six Gaussian distributions on a single decision axis (instead of two distributions as in the standard SDT case—see Figure 1). The relative locations of these distributions are fixed for each animal (assuming that the transformation of physical illuminance to internal representation stays constant over the whole experiment), while the decision criterion in each session was allowed to vary.

The model was fitted to the data as follows: The fraction of left choices for each stimulus for all sessions for one animal were converted to standard scores using the probit function (the inverse of the cumulative normal distribution), resulting in 6*j values, in which j denotes the number of sessions for that animal. Fractions of 0 were corrected by substituting $1/(2N)$, where N is the number of trials for the stimulus (usually 50). Equivalently, fractions of 1 were substituted with $1 - 1/(2N)$ (see Brown and White, 2005, for a discussion of different corrections for proportions of 0 and 1).

Under the assumption that each stimulus yields a variable value on an internal decision variable, and that the variability of the values for each stimulus is described by a Gaussian probability density, the conversion of the fraction of left responses into standard scores gives the difference of the mean of a stimulus distribution from the decision criterion. Thus, a fraction of 0.1 (10% left responses for one stimulus in one session) yields a score of -1.28, meaning that the mean of the distribution is 1.28 standard scores to the left of the decision criterion. This way, 90% of the distribution is to the left of the decision criterion, 10% is to the right.

In our formulation of SDT, we assume that the centers of the six stimulus distributions on the decision axis are fixed throughout the entire experiment, and the distributions are assumed to have equal variance. Accordingly, differences in the fraction of left responses can only arise from session-wise variations in the decision criterion. Our SDT-based model tries to simultaneously find the latent scale values (i.e., the centers of the distributions) and the bias (value of the decision criterion) for each session. We assume that, in each session, the difference between the mean of a stimulus distribution and the criterion on the decision variable in each session is determined by two parameters: Firstly, the difference of the mean of a stimulus distribution and a neutral decision criterion, which applies to all sessions, and secondly, an additional shift of the decision criterion that is specific for each session, but simultaneously applies to all distributions. In formal terms:

$$d_{ij} = x_i + c_j \qquad (A1)$$

where x_i denotes the difference of the center of the distribution of stimulus i ($i \in \{1,2,3,4,5,6\}$) to a neutral decision criterion, c_j denotes the shift in the decision criterion (bias) in session j, where j can take integer values from 1 to the total number of experimental sessions for each

animal, and d_{ij} denotes the difference of the mean of the distribution of stimulus i to the decision criterion in session j. We can phrase the analysis problem as a multiple regression with dummy variable coding, where, for each of i*j rows, a variable takes the value of 1 for one of the six stimuli, and also the value of 1 for the relevant session, and all other variables take the value of 0. Accordingly, the decision criterion values would be the standard scores of the fraction of left responses for stimulus i in session j plus the value of the decision criterion for session j (applied to all six stimuli). In sum, there are i+j predictor variables and i*j observed values. This results in a linear model in which the combination of a single stimulus and a single session predicts the difference of the decision criterion to that stimulus' mean in that single session (Equation 5).

The objective reward functions were calculated as follows: For each bird's modeled stimulus distributions (Figure 7), the decision criterion was varied from −5 to +5 in steps of 0.1, and the fraction of correct responses for each stimulus was calculated for each criterion. This procedure resulted in a 101×6 matrix (101 criterion values, six associated probabilities for a correct response, one for each stimulus). Separately for each condition, each element of the matrix was multiplied with the reinforcement probability for the respective category of that stimulus. Each row of the matrix thus contained six products: probability of a correct response × probability of reinforcement for that response. These were averaged across columns (stimuli), yielding the expected number of reinforcers per trial (i.e., expected value) for each criterion value. This procedure was repeated for every contingency of reinforcement. The resulting vector with 101 elements constitutes the objective reward surface —overall expected value for each of 101 possible criterion values.