

Suboptimal criterion setting in a perceptual choice task with asymmetric reinforcement



Maik C. Stüttgen^{a,*}, Nils Kasties^{a,b}, Daniel Lengensdorf^{a,b}, Sarah Starosta^{a,b},
Onur Güntürkün^{a,b}, Frank Jäkel^c

^a Faculty of Psychology, Department of Biopsychology, University of Bochum, GAFO 05/620, Bochum, Germany

^b International Graduate School of Neuroscience, University of Bochum, Bochum, Germany

^c Institute of Cognitive Science, University of Osnabrück, Germany

ARTICLE INFO

Article history:

Received 22 November 2012

Received in revised form

18 December 2012

Accepted 18 February 2013

Keywords:

Ideal observer

Decision making

Payoff matrix

Pigeon

ABSTRACT

Performance on psychophysical tasks is influenced by a variety of non-sensory factors, most notably the magnitude or probability of reinforcement following correct responses. When reinforcement probability is unequal for hits and correct rejections, signal detection theory specifies an optimal decision criterion which maximizes the number of reinforcers. We subjected pigeons to a task in which six different stimuli (shades of gray) had to be assigned to one of two categories. Animals were confronted with asymmetric reinforcement schedules in which correct responses to five of the stimuli were reinforced with a probability of 0.5, while correct responses to the remaining stimulus were extinguished. The subjects' resultant choice probabilities clearly deviated from those predicted by a maximization account. More specifically, the magnitude of the choice bias increased with the distance of the to-be-extinguished stimulus to the category boundary, a pattern opposite to that posited by maximization. The present and a previous set of results in which animals performed optimally can be explained by a simple choice mechanism in which a variable decision criterion is constantly updated according to a leaky integration of incomes attained from both response options.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

A vast body of data supports the notion that animals, including humans, perform statistically optimally in a wide range of tasks, supporting the claim that evolution has shaped the nervous system of organisms in a way that yields maximally adaptive behavior (Pyke et al., 1977). Examples of optimal behavior include multisensory integration (Ernst and Banks, 2002), risk assessment (Balci et al., 2009), reward harvesting (Corrado et al., 2005; Navalpakkam et al., 2010), perceptual classification (Summerfield et al., 2011), visual search (Najemnik and Geisler, 2005), sensorimotor learning (Körding and Wolpert, 2004), and movement planning (Trommershäuser et al., 2005). Optimality is frequently assessed by comparing behavioral output to benchmarks computed via methods derived from statistical decision theory. Such methods have also been used to assess the reliability of sensory neural signals (Newsome et al., 1989; Stüttgen and Schwarz, 2008; Stüttgen et al., 2011a), and have even been invoked as accounts of neural processing (Deneve et al., 1999; Gold and Shadlen, 2002; Jazayeri and Movshon, 2006).

Perhaps the most widely disseminated offspring of statistical decision theory is signal detection theory (SDT; Green and Swets, 1988). SDT provides a conceptual framework for psychophysics in which the sensory and decision processes are separable. SDT posits that repeated presentations of the same physical stimulus give rise to a variable internal representation on a decision axis, which for illustration purposes can be thought of as “perceived stimulus intensity” (but note that the nature of the decision variable is more appropriately characterized as “strength of evidence”; Pastore et al., 2003). The stimulus representation is assumed to vary randomly from one presentation to the next; usually, it is assumed that the random variations conform to a normal distribution with fixed variance. If an observer is asked to categorize either of two different stimuli varying along some physical dimension, SDT assumes that the subject does so by comparing the perceived stimulus intensity on each trial (λ_t) to a criterion value c , with the decision rule:

- if $\lambda_t \geq c$, respond “high-intensity stimulus present”,
- if $\lambda_t < c$, respond “low-intensity stimulus present”.

This decision rule can be generalized to more than two stimuli and to other kinds of tasks (MacMillan and Creelman, 2005). For our present purposes, we will discuss the case of a single-interval forced choice (categorization) task with six stimuli differing in luminance.

* Corresponding author. Tel.: +49 234 32 24 323; fax: +49 234 32 14 377.

E-mail address: maik.stuettggen@rub.de (M.C. Stüttgen).

Fig. 1A shows six normal distributions (gray), each separated by 1 standard deviation from its immediate neighbors, which correspond to the hypothesized internal representations of the stimuli. Assume that the three leftmost ones are arbitrarily assigned to category S1, and the three rightmost ones are assigned to category S2. An observer is confronted with the task to respond with “S1” or “S2” on a given trial t on the basis of a single stimulus value, λ_t .

1.1. Optimization account of perceptual categorization

Statistical decision theory prescribes an optimal decision rule in which sensory evidence, a priori probability of signal occurrence, and values and costs of correct and incorrect responses are integrated. Here, we will assume that all stimuli are equiprobable and ignore costs of incorrect responses. Instead, we will focus on the effects of different values, implemented by assigning different reinforcement probabilities for correct responses following different stimuli.

Fig. 1A illustrates the simplest case in which correct responses following each stimulus are reinforced with the same probability (0.5; “symmetrical reinforcement”). The bold black line represents the “objective reward function” (ORF; Maddox, 2002), i.e. the expected number of reinforcers per trial as a function of criterion placement. The position on the x -axis for which the ORF has its maximum value corresponds to the location of the optimal criterion and is indicated by the black vertical line. In the present example where all neighboring stimuli are equidistant, the optimal criterion is located right in the middle between the means of the third and the fourth stimulus distributions. The optimal strategy dictated by statistical decision theory is to respond “S1” whenever λ_t is smaller and to respond “S2” whenever λ_t is larger than this criterion (the computations giving rise to the criterion placements in Figs. 1 and 2 are explained below and are also contained in a Matlab script provided as supplementary material).

In Fig. 1B, the six gray sigmoidal lines show the probability of reinforcement for each stimulus separately as a function of criterion placement for the same situation as in Fig. 1A. The probability of reinforcement increases with the criterion for S1 trials (three darkest curves, left) and decreases for S2 trials (three brightest curves, right), because a higher criterion value will lead to more “S1” responses and fewer “S2” responses. Reinforcement probability saturates at 0.5, because in our experiment correct responses only yield reinforcement with that probability. Since all six stimuli have the same probability (1/6) of being presented to the subject, the overall probability of obtaining reinforcement in a random trial (bold black line) is the average of the reinforcement probabilities for each stimulus; this is just another way to define the ORF.

Formally, let us denote the means of the six stimulus distributions with μ_1 to μ_6 . The probability of responding with “S1” is then $\Phi(-\mu_i + c)$ for the i th stimulus, where Φ is the standard normal cumulative distribution function. Let reinforcement $Rf_{S1} = 1$ if an S1-category stimulus was presented and the subject’s response was “S1” and led to reinforcement; let $Rf_{S1} = 0$ otherwise. Correspondingly, let $Rf_{S2} = 1$ if an S2-category stimulus was presented and the subject’s response was “S2” and led to reinforcement; let $Rf_{S2} = 0$ otherwise. Then the probabilities for reinforcement on S1- and S2-trials ($E(Rf_{S1})$ and $E(Rf_{S2})$, respectively) are

$$E(Rf_{S1}) = \sum_{i=1,2,3} \frac{1}{6} \times \Phi(-\mu_i + c) \times 0.5$$

and

$$E(Rf_{S2}) = \sum_{i=4,5,6} \frac{1}{6} \times (1 - \Phi(-\mu_i + c)) \times 0.5$$

The probability of obtaining a reinforcer in any given trial is the sum of the two: $E(Rf) = E(Rf_{S1}) + E(Rf_{S2})$. $E(Rf)$ (or, more

precisely, $E(Rf|c)$) is the objective reward function. In the following section, we will describe an experimental manipulation of reinforcement probability which yields counterintuitive predictions under a reinforcement-maximization account. Then, we will describe the predictions of a non-optimization account which builds on insights from animal learning theory.

1.2. Experimental manipulation: extinguishing responding to a single stimulus

What happens in the above scheme when the reinforcement probability for a single stimulus is reduced to 0? In our experiment, all six stimuli are still shown with the same probability, and correct responses are reinforced with probability 0.5 for all but one stimulus. Fig. 1C–H illustrates the consequences of extinguishing responding to each of the six stimuli at a time (these conditions are henceforth termed E1 to E6). It can be seen that the ORFs become asymmetrical, and that the peaks of these functions (i.e., the positions of the optimal decision criteria) move away from the neutral criterion (the latter is depicted as a dashed vertical line in all panels for comparison purposes). Obviously, the effect of extinguishing responses to a single stimulus on the position of the optimal criterion depends on the condition: Intriguingly, the criterion is almost unaffected in conditions E1 and E6 (compare Fig. 1C and H to Fig. 1B). Intuitively, since stimulus 1 is furthest away from the neutral criterion, its contribution to the ORF around the optimum is very small and nearly flat (Fig. 1A and B), and therefore extinguishing it does not change the position of the optimum much, it merely shifts its peak downwards (Fig. 1C). The distribution of stimulus 3, on the other hand, is very close to the neutral criterion at zero, and its contribution to the ORF changes a lot around the peak (Fig. 1A and B); hence, extinguishing it will lead to a larger shift of the ORF’s peak (Fig. 1E). Qualitatively, if subjects have an algorithm for setting the criterion such that it moves towards the optimal criterion, we would expect that the criteria for the six conditions ($c(E1)$ to $c(E6)$) are ordered in the following way: $c(E3) < c(E2) < c(E1) < c(E6) < c(E5) < c(E4)$ (as shown in Fig. 1C–H). Because the S1/S2 response ratio increases with increasing criterion position, we expect the same order for the S1/S2 response ratio. Assuming six distributions with equal distances between neighboring stimuli, quantitative predictions for the optimal criterion position can be computed numerically, and the results of these computations are shown in Fig. 1 (also see supplementary Matlab code).¹

However, these predictions should be highly counterintuitive for anyone familiar with animal learning theory. After all, an observer performing the task with sensitivity as depicted in Fig. 1A earns a substantial fraction (close to 40%) of overall reinforcement in trials in which stimuli 1 and 6 are presented, as these are easiest to classify. In consequence, extinction of responding to stimuli 1 and 6 entails a larger loss of reinforcers than extinction to stimuli 3 and 4. A vast body of literature shows that animals are highly sensitive to changes in the frequency of positive reinforcement and that magnitude or probability of positive reinforcement is monotonically related to choice probability (Herrnstein, 1961; Reynolds, 1961; Nevin et al., 1975; McCarthy and Davison, 1981; Corrado et al., 2005; Balci et al., 2009; Teichert and Ferrara, 2010; Stüttgen et al., 2011b). Accordingly, since subjects lose more reinforcers from category S1 in condition E1 than in E3, one would expect that S1 responses should decrease more in condition E1 than in condition E3. Similar considerations for the other conditions would lead

¹ The ordinal prediction breaks down if the distance between the stimuli, i.e. the overlap between the neighboring distributions, becomes very small (ca. $d' < 0.3$). We have set up the experiment such that the ordinal prediction holds.

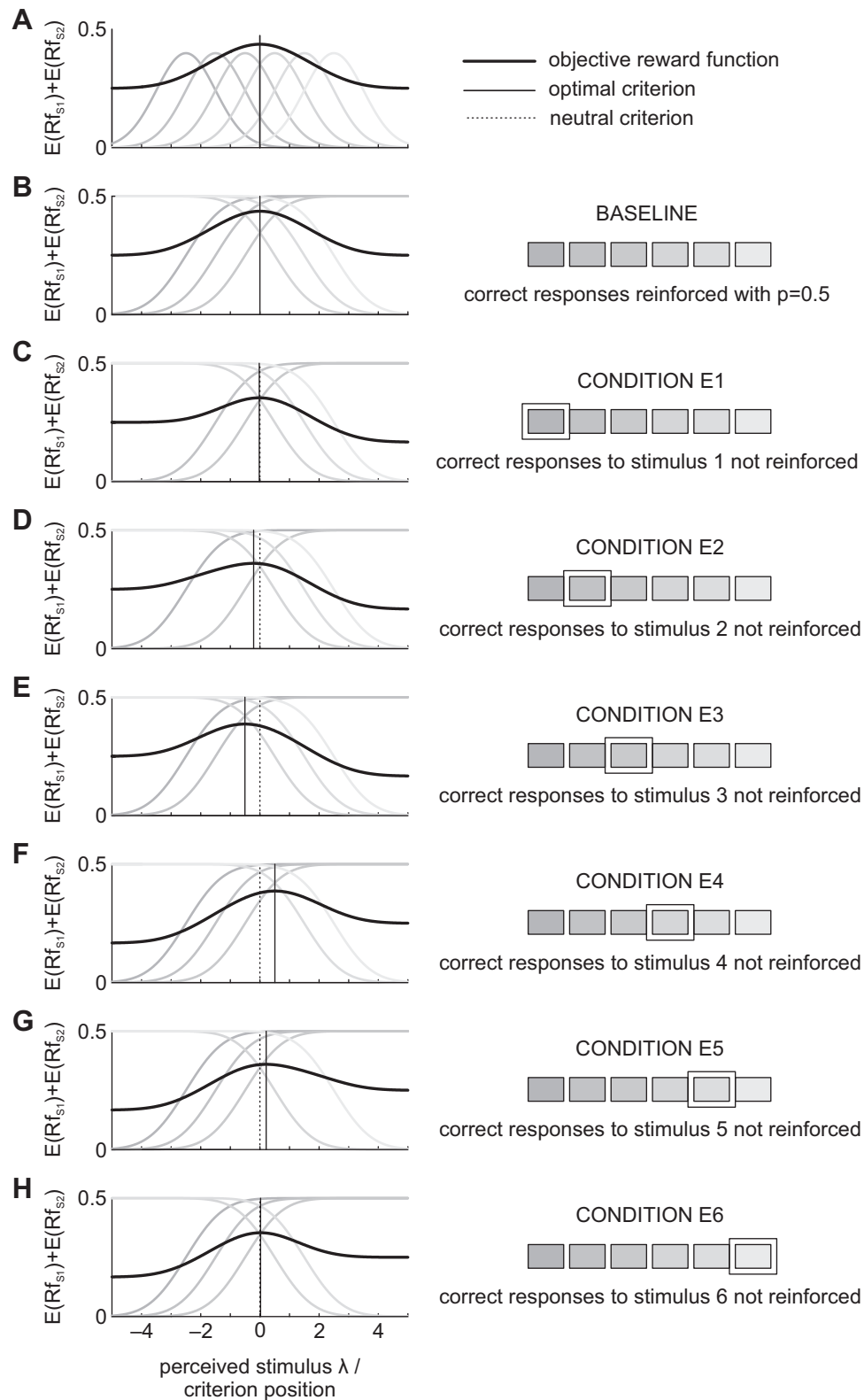


Fig. 1. Outline of signal detection theoretical model and predictions for criterion placement. (A) Panel shows six stimulus distributions (gray) on an arbitrary decision axis, corresponding to the internal representations of six stimuli. The distance between adjacent distributions is always 1 standard deviation (i.e. $d' = 1$). The three left distributions are to be categorized as “S1”; the three right distributions are to be categorized as “S2”. The solid vertical line signifies the position of the optimal (i.e. reinforcement-maximizing) criterion. The bold black line represents the objective reward function, i.e. the expected average number of reinforcers per trial as a function of criterion position, and peaks at the position of the optimal criterion. (B) As in (A), but gray lines represent the expected average number of reinforcers per trial for each stimulus separately. The six boxes on the right illustrate the to-be-categorized stimuli. (C)–(H) As in (B), but each panel is valid for one experimental condition E1 . . . E6, each of which is illustrated in right column. The stimulus whose reinforcement probability was set to 0 is highlighted. The dotted vertical lines visible in panels (C)–(H) indicate the position of the neutral (i.e. accuracy-maximizing) criterion and are shown for comparison purposes only. See Sections 1.1 and 1.2 for more details.

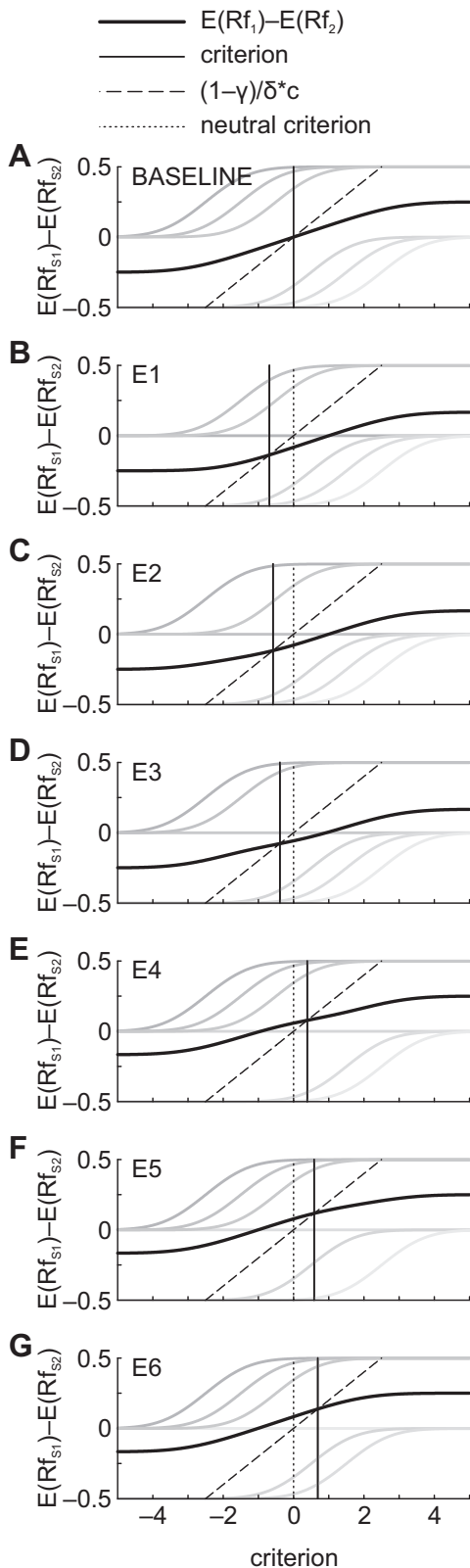


Fig. 2. Outline of an income-based model and its predictions for criterion placement. In each panel, the bold black line gives the difference between the expected number of reinforcers per trial obtained by correct S1 and S2 responses, $E(Rf_{S1}) - E(Rf_{S2})$. The gray lines represent this difference for each stimulus separately. The dashed tilted line represents $((1 - \gamma)/\delta) \times c$ as explained in Section 1.3, and is the same for all panels. The solid vertical line indicates the predicted position of the decision criterion given the particular combination of d' , γ and δ . Panel (A) gives predictions for baseline conditions when reinforcement probability for correct responses is 0.5 for all stimuli. Panels (B)–(G) give predictions for the six experimental conditions E1...E6. In all panels, the dotted vertical lines indicate the position of a neutral decision criterion.

one to expect that $c(E1) < c(E2) < c(E3) < c(E4) < c(E5) < c(E6)$ for the criteria and S1/S2 response ratios. Next, we will present a non-optimizing model that makes predictions that are more in line with intuitions from animal learning theory.

1.3. A non-optimization account of criterion setting

There are various non-optimizing algorithms one can imagine that would implement the intuition that $c(E1) < c(E2) < c(E3) < c(E4) < c(E5) < c(E6)$, for example any algorithm that leads to matching, such as melioration (Herrnstein and Vaughan, 1980). However, there is the added complication in our experiment that we have to deal with perceptual uncertainty. While there are accounts that link signal detection theory with generalized matching (Davison and Tustin, 1978; Davison and Nevin, 1999), these only deal with steady-state behavior, thus ignoring the mechanisms underlying adaptation (Stüttgen et al., 2011b). Several authors have suggested algorithms for perceptual choice tasks operating on a trial-by-trial level; the algorithms suggested by Boneau and Cole (1967) and by us (Stüttgen et al., 2011b) lead to optimization; the algorithms by Dorfman and Biderman (1971) and Treisman and Williams (1984) do not. The model we describe next is a blend between criterion setting theory (Treisman and Williams, 1984) and model 3 of Dorfman and Biderman (1971).

The model operates within the SDT framework, thus (for the case at hand) postulating six stimulus distributions and a single criterion. The central idea of the model is that the criterion is updated after each trial in which reinforcement occurred. More specifically, the criterion shifts to the right after reinforcement of an S1 response and shifts to the left after reinforcement of an S2 response; thus, whenever the animal is reinforced the criterion is shifted so that the response that was reinforced becomes more likely. Put differently, the criterion integrates the difference between the number of reinforcers obtained from S1 and S2 responses. However, in this form the model will always lead to exclusive choice of one response option, since each criterion shift will make a subsequent criterion shift in the same direction more likely and a criterion shift in the opposite direction less likely. For this reason, Dorfman and Biderman (1971) rejected this model and preferred models that also (or exclusively; Kac, 1969) learn from error trials. However, the model can be fixed by assuming that the integration of reinforcements is “leaky”, and that recent reinforcers play a bigger role than distant reinforcers in criterion setting (a similar assumption is made by criterion setting theory; Treisman and Williams, 1984).

Formally, as before, let $Rf_{S1}(t) = 1$ ($Rf_{S2}(t) = 1$) if the animal responds “S1” (“S2”) on an S1 (S2) trial t and is reinforced; otherwise $Rf_{S1}(t) = 0$ ($Rf_{S2}(t) = 0$). Note that only one of the two variables can be 1 on the same trial, but both can be zero. If the criterion on trial t is $c(t)$ then:

$$c(t + 1) = \gamma \times c(t) + \delta \times [Rf_{S1}(t) - Rf_{S2}(t)]$$

where δ is the positive step-size for the criterion shift and γ is a forgetting term (or leak factor) that can range from 0 to 1. The criterion is just a scaled version of a leaky integration of the difference of the incomes on both sides (see Section 2.5 for proof).

What is the criterion that this simple criterion learning algorithm will converge to? Since the responses are stochastic, there will be an equilibrium distribution for the criterion, but deriving this equilibrium distribution (and showing that it actually exists) is beyond the scope of this paper (but see Norman (1972) for an analysis of the Kac–Dorfman–Biderman model). Heuristically, the criterion does not change on average if the expected step away

from the neutral criterion at zero is as big as the pull back to zero, that is

$$[\gamma c - c] + \delta \times [E(Rf_{S1}) - E(Rf_{S2})] = 0$$

Hence, there is a linear relationship between the criterion and the difference of the probabilities for a reinforcement for category 1 and category 2:

$$E(Rf_{S1}) - E(Rf_{S2}) = \frac{[1 - \gamma]}{\delta} \times c$$

In Fig. 2, the difference between $E(Rf_{S1})$ and $E(Rf_{S2})$ as a function of the criterion is plotted as a black bold line. We have also plotted these differences for each stimulus separately as gray lines. The criterion that the algorithm should converge to lies on a straight line through zero (dashed black lines in Fig. 2); the slope of that line depends on the step size δ and the leakage term γ (the slope of the line in Fig. 2 was chosen arbitrarily). The criterion that we expect the animal to converge to lies at the intersection of the two black curves and is marked with a vertical line. Fig. 2B–G shows how the criterion changes when we extinguish the different stimuli. The predicted criterion-order now becomes $c(E1) < c(E2) < c(E3) < c(E4) < c(E5) < c(E6)$. Note that this ordinal prediction does not depend on the precise slope. This prediction is consistent with the intuition that the more income the animal loses from the first category, the more the animal should respond to the second category.

The design of the experiment that will be reported below does not depend on this specific model. There are other models that make the same order predictions (e.g. based on matching), but we opted for the outlined model because of its conceptual simplicity and because it directly relates to previous attempts to model psychophysical performance using an SDT framework.

2. Methods

2.1. Subjects

Subjects were four pigeons (*Columba livia*) obtained from local breeders. The subjects served on previous experiments employing choice tasks. Animals were housed individually in wire mesh cages inside a colony room with a 12-h light–dark cycle (lights on at 8 a.m.). Water was available ad libitum throughout the experiment. On weekends, food was freely available; on weekdays, food was provided only during testing in the experimental chamber. Animals were food-deprived to 85–90% of their free-feeding weight. Supplemental food to prevent further weight loss was given whenever necessary. All subjects were kept and treated according to the German guidelines for the care and use of animals in neuroscience, and all procedures were approved by an ethics committee of the State of North Rhine-Westphalia, Germany.

2.2. Apparatus

Experiments were conducted in an operant chamber (measuring roughly 35 cm along all three dimensions) with three translucent response keys arranged side by side. The response keys measured 4 cm × 4 cm and were located 20 cm above the floor. Each effective key peck produced a feedback click. The required force for activation was ~25 g. A food hopper was located below the center key. Stimuli were shown on a flat screen monitor (Philips 150P2, 15", native resolution 1024 × 768) mounted on the back of the experimental chamber. Stimuli were shades of gray with grayscale values of 140, 160, 170, 190, 200, and 220. In the following, stimuli with grayscale values of 140, 160 and 170 will be referred to as S1 or "dark", and stimuli with grayscale values of 190, 200, and 220 will be referred to as S2 or "light". Thus, the category boundary

between S1 and S2 was arbitrarily placed at grayscale value 180. The gray values were chosen from previous experience such that the most extreme values were categorized correctly in about 95% of trials in the symmetrical reinforcement condition. The chamber was positioned in a sound-attenuating shell, and white noise was provided at all times to mask extraneous sounds. Experimental hardware was controlled by custom-written Matlab (The Mathworks, Natick, MA) code (Biopsychology Toolbox, Rose et al., 2008).

2.3. Procedure

Testing was conducted every weekday. Each session lasted about 50 min and comprised 300 trials. The general paradigm is illustrated in Fig. 3. At the beginning of each trial, the center key was illuminated green. The animal could initialize the trial by responding with a single peck to the center key within 5 s. If the subject failed to respond, an initialization omission was counted, the trial was aborted, and the animal reentered the intertrial interval (4 s). Omitted trials were not repeated. On average, the animals completed more than 94% of trials in each session (all medians $\geq 98\%$). After initialization, one of the six stimuli (pseudorandom sequence) was presented on the center key for 1 s (sample phase). Subsequently, the center key turned green again (confirmation phase). A single key peck switched off the green center key and turned on the two side (choice) keys, which were illuminated orange, until the animal responded by pecking at one of them once (choice phase). If the animals failed to respond during the confirmation or the choice phase, the trial was aborted within 5 s, and the animals reentered the intertrial interval. However, trials were almost never aborted during these phases. The confirmation phase was included to ensure that the animals kept attending to the sample stimuli for the full 1 s presentation time.

Responses to the right choice key were reinforced probabilistically after presentation of any S1 stimulus, and responses to the left choice key were reinforced probabilistically after presentation of any S2 stimulus. Reinforcement consisted of 2 s access to grain provided by a food hopper located below the center key. Probability of reinforcement was 0 for incorrect responses and 0.5 for correct responses in baseline control conditions (10 sessions before and 10 sessions after the six experimental conditions). In the experimental conditions, probability of reinforcement for correct responding was reduced to 0 for one of the six sample stimuli and remained at 0.5 for all other stimuli. Throughout the entire experiment, every correct response was followed by illumination of the food hopper, regardless of whether food was provided or not. Every incorrect response was followed by a 2 s time-out during which the keys were inoperative and all houselights were turned off. The six experimental conditions (henceforth referred to as E1–E6) were maintained for 15 sessions each (up to 4500 trials per condition). A previous study (Stüttgen et al., 2011b) showed that the animals' behavior in this paradigm reliably stabilized after about 10 sessions or 3000 trials. The order in which stimuli were assigned a reinforcement probability of 0 was counterbalanced across animals (the order of testing is shown in Fig. 4 for each animal).

2.4. One-criterion-per-session model

A model based on SDT was fitted to the data. We will use this model as a standard to compare it to the criterion learning model described in the introduction. The one-criterion-per-session comparison model assumes the existence of six normal equal-variance distributions on a decision axis and a criterion which could vary across but is fixed within experimental sessions. In this way, we can observe criterion shifts over sessions even if we assume the

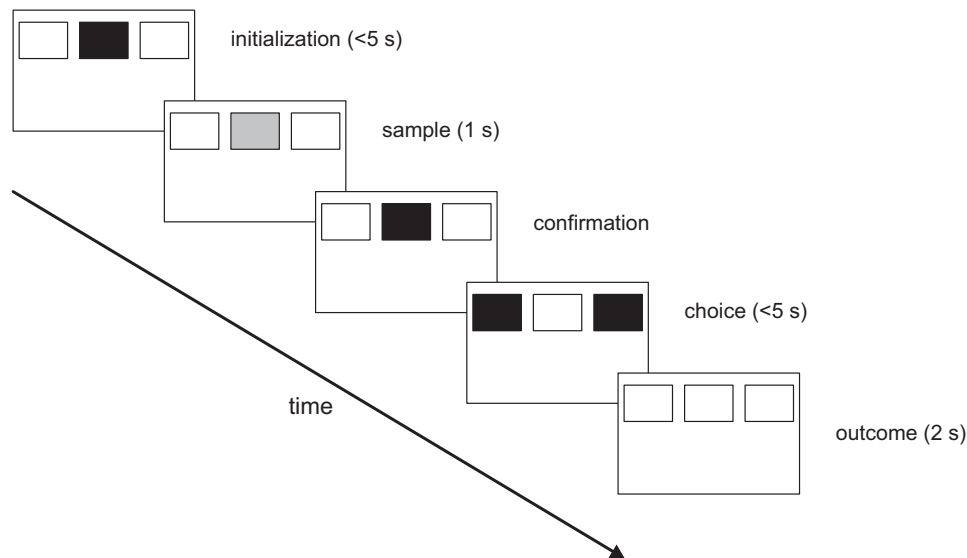


Fig. 3. Schematic outline of the categorization task. Rectangles represent three response keys. Sequence of events runs from top left to bottom right. After elapsing of the intertrial interval (ITI) of 4 s, the center key is illuminated green (initialization phase). After a single response, the center key displays one of six possible stimuli (shades of gray) for 1 s (sample phase), after which the key turns green again (confirmation phase). Following another response, the center key is turned off, and the side keys are illuminated orange (choice phase). The subject is required to indicate its decision by responding once to either choice key. If the response is correct, a food hopper is activated for 2 s according to a probabilistic schedule. If incorrect, all lights are switched off for 2 s (punishment).

criterion does not change within a session. Model parameters were fitted using maximum likelihood.² For each bird, fitted parameters included the means of the six distributions ($\mu_1, \mu_2, \dots, \mu_6$) and one criterion value (c) for each session, yielding $k = 116$ free parameters, as each bird was tested for 110 sessions. The binomial probability of responding with category 1 in a trial that presented stimulus i in session j was $\Phi(-\mu_i + c_j)$, where Φ denotes the cumulative standard normal distribution function. This model is an instance of a generalized linear model, and the convex optimization problem can easily be fitted by a standard Newton method (Dorfman, 1973). Trials in which the subject did not respond were simply ignored (5% for bird #720, 2% for #919, 1% for #920, 5% for #935). Goodness of fit was assessed by deviance (Collett, 1991). For binomial models, deviance can be thought of as the appropriate analogue of summed squared error in normal models. We compared the obtained deviance to the deviance that is expected from a parametric bootstrap using the maximum likelihood fit described in Wichmann and Hill (2001). If the obtained deviance is significantly larger than the expected deviance one speaks of overdispersion in binomial models.

To sum up, the one-criterion-per-session model describes the data set for each bird as arising from six normal distributions with fixed means and equal variance and one decision criterion for each experimental session (the same approach as in Stüttgen et al., 2011b).

2.5. Income-based criterion learning model

In order to gain a deeper understanding of the algorithm underlying adaptive choice in this task, we went on to conceive an income-based model that operates at the single trial level. Income I is simply the number of reinforcers attained from responding to each choice option (S1 and S2) over trials, $I_{S1,t} = \sum \text{Rf}_{S1,t}$

and $I_{S2,t} = \sum \text{Rf}_{S2,t}$ for choices of S1 and S2, respectively, where $\sum \text{Rf}_{Sx,t}$ denotes the sum of reinforcers from Sx responses across trials.

As the one-criterion-per-session model, the income-based model assumes the existence of six equal-variance normal stimulus distributions and a decision criterion. While the comparison model assumes one criterion for each experimental session, the criterion in the income-based model varies from trial to trial within each session. On trials where the animal is reinforced, the criterion shifts in the direction that makes the reinforced response more likely (as in the models of Kac, 1969; Dorfman and Biderman, 1971; Dorfman et al., 1975). Such a model always leads to exclusive choice (Dorfman et al., 1975); hence we added a forgetting term that constantly pulls the criterion back to a neutral criterion (Treisman and Williams, 1984). The model is already described in the introduction, but we repeat the equation here for easy reference: $c(t+1) = \gamma \times c(t) + \delta \times [\text{Rf}_{S1}(t) - \text{Rf}_{S2}(t)]$.

The income-based learning model has $k = 8$ free parameters for each bird (6 means, the leak factor γ , and the learning rate δ). Note that the criterion is just a scaled version of a leaky integration of the difference of the incomes on both sides. To see this define I_{L1} to be a leaky integration of the reinforcers for response 1 and I_{L2} correspondingly, $I_{L1}(t+1) = \gamma \times I_{L1}(t) + \text{Rf}_{S1}(t)$ and $I_{L2}(t+1) = \gamma \times I_{L2}(t) + \text{Rf}_{S2}(t)$. Then, $[I_{L1}(t+1) - I_{L2}(t+1)] = \gamma \times [I_{L1}(t) - I_{L2}(t)] + [\text{Rf}_{S1}(t) - \text{Rf}_{S2}(t)]$, and therefore $c(t) = \delta \times [I_{L1}(t) - I_{L2}(t)]$.

The probability of responding with category 1 on trial t when the stimulus was i is therefore $\Phi\{-\mu_i + c(t)\} = \Phi\{-\mu_i + \delta \times [I_{L1}(t) - I_{L2}(t)]\}$. For the very first trial in the first session we assume that income from both sides (or equivalently, the criterion) was zero before. We further assumed that the leaky incomes (or equivalently, the criterion) for the first trial of each of the other sessions are the values of the last trial of the previous sessions.

Trials in which the subject did not respond were ignored. For a fixed γ , we can fit this equal-variance signal-detection model with trial-by-trial criterion adaptation by using the same

² Actually, we used penalized maximum likelihood to increase the stability of the fitting procedure. We used L2 regularization with a small regularization parameter (10^{-5}) that does not influence the estimate in a noteworthy way.

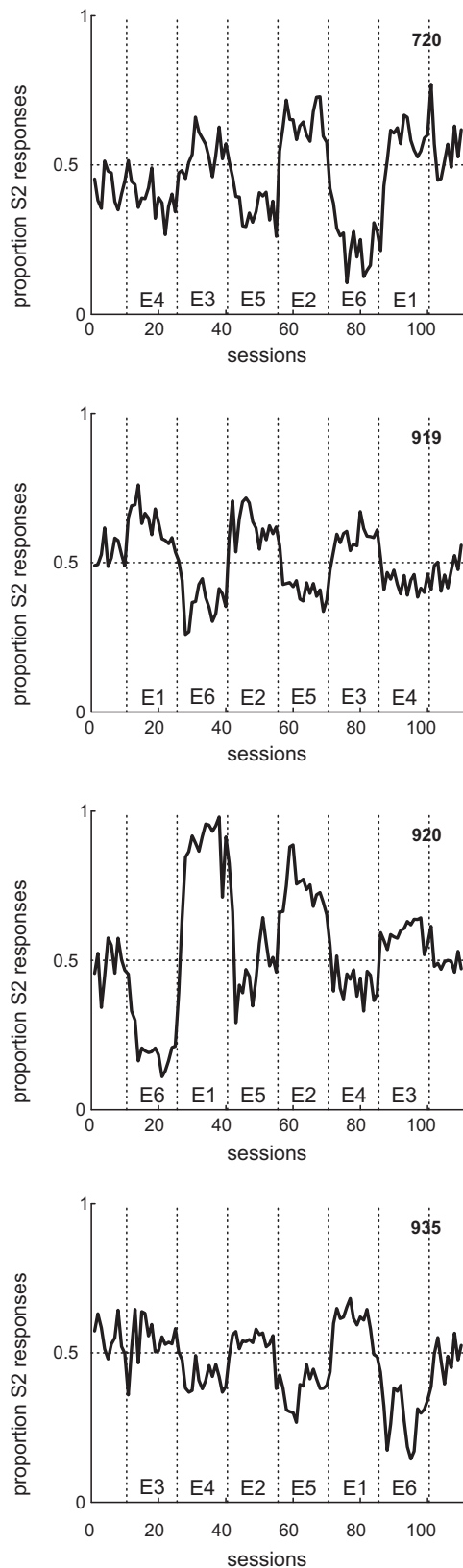


Fig. 4. Choice patterns (proportion of S2 responses) for all birds across all sessions. Dashed vertical lines separate reinforcement conditions. Numbers indicate which of the six stimuli (ranked from 1 to 6 in order of increasing luminance) did not yield reinforcement in the respective condition.

standard methods for generalized linear models as before, the only difference now is that we use the difference of integrated incomes as a predictor (Dorfman and Biderman, 1971; Dorfman, 1973). In order to optimize also over γ , we performed a one-dimensional optimization where in each substep we sought the best fit for the current value of γ .

2.6. Statistical analysis

For the ordinal analysis of the responses, we employed Spearman's rank order correlation coefficient. Goodness of fit of the models was assessed by negative log likelihood (NLL) and deviance. Models were compared using the Bayesian Information Criterion, $BIC = [2NLL + k \log(N)]$, where k is the number of free parameters (116 and 8 for one-criterion-per-session and the income-based models, respectively) and N is the number of data points (trials in this case). The BIC allows comparing goodness of fit of models with different numbers of free parameters; the better model is the one featuring a smaller BIC value. All analyses were performed using MATLAB 7.8.0 using purpose-written code.

3. Results

Fig. 4 shows the birds' choice patterns (proportion S2 responses) across all sessions. During baseline testing (symmetrical reinforcement during the first 10 sessions), the animals did not exhibit a marked preference for any response category. However, responding was heavily biased during asymmetric reinforcement schedules, such that pigeons avoided responses to the key which was associated with the non-reinforced stimulus. The magnitude of the response bias was dependent on which stimulus was unreinforced: the birds exhibited more extreme choice biases in conditions with unreinforced stimuli further away from the category boundary (E1, E6) than in conditions with unreinforced stimuli closer to the boundary (E3, E4). This impression was confirmed when constructing psychometric functions for each animal, pooling response data from the last 5 sessions of each condition (Fig. 5).

If the qualitative impression gained from Figs. 4 and 5 – larger biases for more extreme stimuli – is correct, this would run contrary to the optimization model derived from signal detection theory. The optimality model predicts that the response bias resulting from the extinction of responding to stimuli 1 and 6 should be smaller than the response bias resulting from the extinction of responding to stimuli 3 and 4 (see Section 1 and Fig. 1). More specifically, optimization predicts that the number of S2 responses should follow the pattern $E3 > E2 > E1 > E6 > E5 > E4$. Fig. 6 (left column) plots the relative frequency of S2 responses in the last five sessions of each condition, with conditions rearranged such that the data points should progressively decrease from left to right. It is evident that the predictions of the optimization account were not met; the rank-order correlation coefficients ranged from only -0.6 to -0.71 . In fact, the optimization account succeeded only in predicting the direction of the choice bias ($S2 > S1$ in conditions E1–E3, $S2 < S1$ in conditions E4–E6); however, especially in conditions E4–E6, the predicted rank order was frequently inverted (see birds #919 and #935 for the most extreme examples).

Does an account that is based on income predict the observed patterns? As explained in the introduction, pairwise comparisons of the proportions of S2 responses should follow the pattern $E1 > E2 > E3 > E4 > E5 > E6$ (see Fig. 2). The data are plotted in the right column of Fig. 6, again with the conditions arranged to reflect the predicted pattern. Although there are some deviations from the income-based predictions, rank-order correlation coefficients were

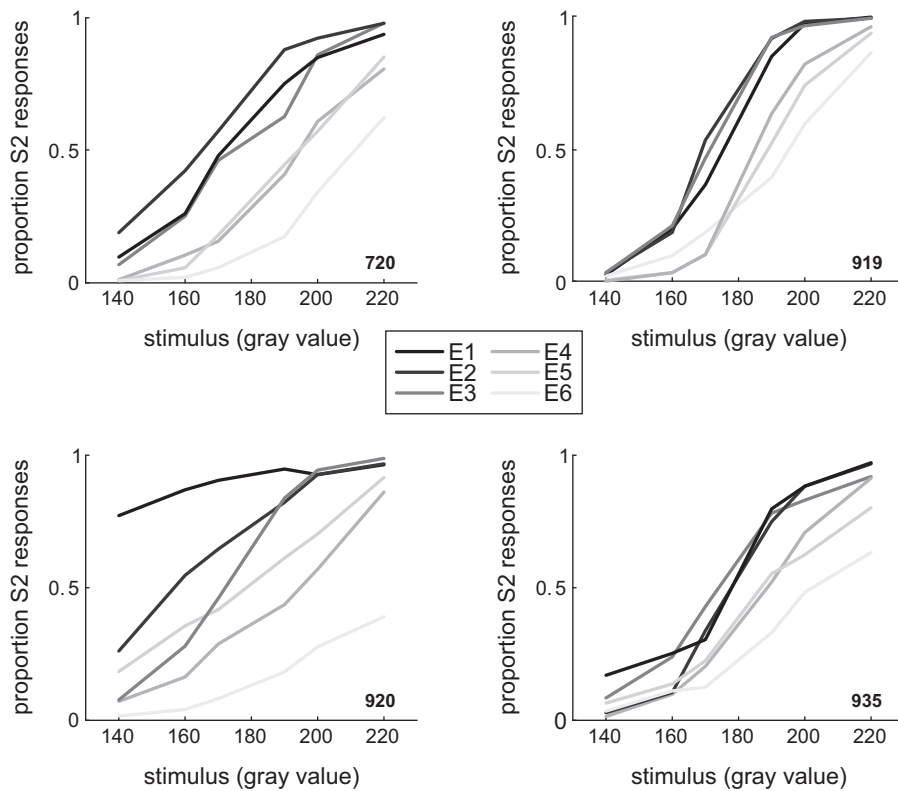


Fig. 5. Psychometric functions averaged over the last five sessions within each condition, shown separately for each bird. Shading indicates experimental condition, i.e. the stimulus for which correct responses were not reinforced (see legend).

substantially higher than for the predictions of the optimization account, ranging from -0.83 to -0.94 .³

The fit of the income-based criterion learning model is shown in Fig. 7 which plots the proportion of S2 responses aggregated for each session, separately for each bird (black: observed data replotted from Fig. 4, dark gray: model fit to the data). Visually, the fit looks reasonable. To assess whether this criterion learning model provides a statistically satisfactory fit to the data, we compared its negative log likelihood (NLL) to that of the one-criterion-per-session model (i.e., one criterion for each of the 110 sessions; see Sections 2.4 and 2.5). The NLLs and BICs for all animals are shown in Table 1. The criterion learning model has $k=8$ free parameters (6 means, the learning rate δ , and the leak factor γ) whereas the one-criterion-per-session comparison model has $k=116$ (6 means and 110 criterion values). The number of trials N is roughly 30,000 for each animal. Despite the wildly different number of free parameters, the NLLs are comparable for the two models, with the comparison model performing only slightly better. However, since that model has many more free parameters, the BIC strongly prefers the criterion learning model for all four animals.

Both models leave quite a bit of variability in the data unexplained: inspection of the fitted psychometric functions for the

Table 1

Goodness of fit for income-based learning and one-criterion-per-session models. Numbers in cells provide negative log likelihoods (columns labeled NLL) and Bayesian Information Criteria (columns labeled BIC) for the income-based criterion learning and the one-criterion-per-session comparison model.

Bird	Criterion learning model		One-criterion-per-session model	
	NLL	BIC	NLL	BIC
720	14,977.3	30,037.4	14,761.5	30,713.8
919	11,517.0	23,117.0	11,442.1	24,079.3
920	14,649.1	29,381.3	14,367.7	29,930.2
935	14,168.8	28,420.3	14,036.9	29,263.9

one-criterion-per-session model revealed that the deviance of the data points (which is comparable to variance in normal models) was about two to three times as high as expected from a binomial process (average of 2.6 for the four birds), a common finding with binomial models known as overdispersion (Collett, 1991). Obviously, this could be due to a non-stationary decision criterion, as has been noted by other authors (Treisman and Williams, 1984). Accordingly, we would expect less overdispersion with the criterion learning model. Unfortunately, deviance cannot be calculated for this model since we are dealing with Bernoulli trials and not binomial data. But as the criterion learning model has a larger negative log likelihood than the comparison model, we can conclude that there are still unidentified causes of response variability. Also, if we simulate responses from the fitted criterion learning model and then fit the one-criterion-per-session model to the simulated data we do not observe overdispersion as big as in the original data.

The criterion learning model uses a leaky integration of the incomes as a predictor for the responses. As the leaky integration

³ We opted for a test of the rank-order predictions of optimizing and the income-based models because this requires the fewest assumptions. For example, the income-based model predicts the rank order $E1 > E2 > E3 > E4 > E5 > E6$ regardless of the specific values of γ and δ . Similarly, the quantitative prediction of SDT could be moderated by a 'conservatism' bias, in which subjects adjust their response criterion only by a constant fraction of that required for optimal performance (Green and Swets, 1988). Still, the degree of conservativeness has no effect on the rank order of the predicted criterion sequence.

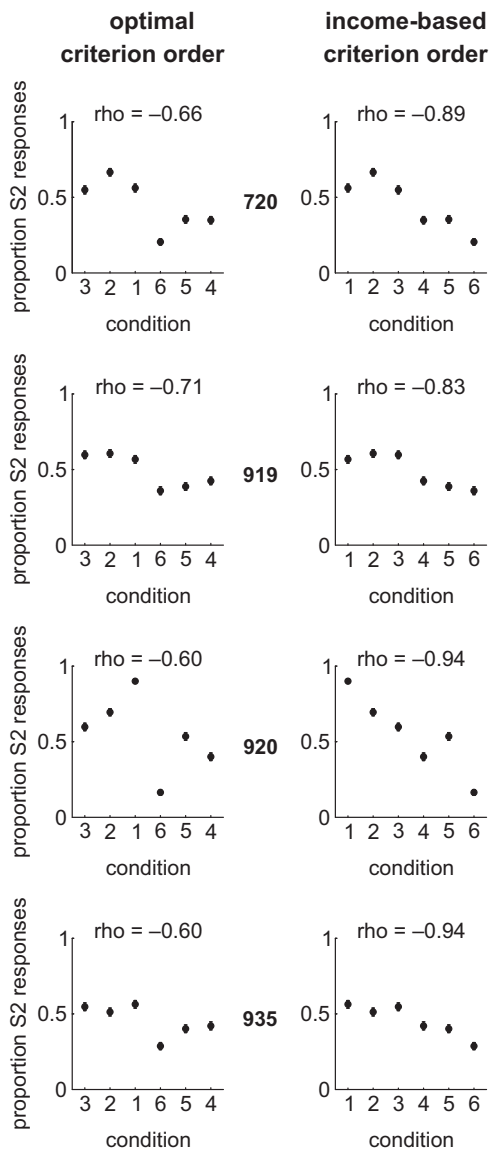


Fig. 6. Comparison of predicted and obtained response criterion order. Each row depicts the relative frequency of S2 responses during the last five sessions of each condition for an individual animal. Panels in the left column assess the rank-order fit of the optimization model, with the order of conditions arranged such that the data points should progressively decrease from left to right. Similarly, panels in the right column assess the rank-order fit of the criterion learning model, and again the conditions are arranged such that a perfect fit would require data points with monotonically decreasing value.

Each data point is based on roughly 1400–1500 trials. Vertical lines indicate exact binomial 95% confidence intervals. Goodness of fit was assessed as the degree to which the fitted response criteria showed the same rank order as predicted by the respective model and quantified using Spearman's rho, provided for each panel.

of the incomes is necessarily correlated with the local averages of the responses, it is perhaps not surprising that this model can, at least qualitatively, fit the average steady-state behavior, even though statistically not all of the variability could be accounted for. Hence, a good visual fit of average steady-state behavior does not necessarily mean that the fitted model would also reach the same steady-state behavior if the data were generated from the model (see discussion in Corrado et al., 2005). Therefore, as a sanity check, we simulated the model forward on the same sequence of stimuli that the subjects received but with the responses generated by the model. The result of one exemplary simulation run can also be seen in Fig. 7 (light gray). The steady-state behavior

that the model generates is consistent with observed behavior. This is not trivial as a model that only learns on error trials (Kac, 1969) will also fit the steady-state behavior visually well but will generate inconsistent behavior (data fit and simulations not shown).

As described above, the four birds previously participated in a similar experiment in which they were found to perform statistically optimal (Stüttgen et al., 2011b). Could optimality in the previous experiment have arisen as a by-product of the same income-based choice mechanism? We applied the fitted criterion learning model to that data set and kept the current estimates of the learning rate δ and the leak factor γ . The means of the six stimulus distributions μ_1 – μ_6 had to be estimated again because we employed a different stimulus set in the previous study. The result of this exercise can be seen in Fig. 8. The model captured the data rather well, with the exception of the 'overshoots' induced by contingency changes (see Stüttgen et al., 2011b for discussion). However, steady-state (quasi-optimal) performance was well approximated, supporting the present model as a viable account of choice behavior in our perceptual categorization task.

4. Discussion

Usually, signal detection experiments employ only two stimuli at a time. Employing six stimuli as in the present study opens up the possibility to manipulate the reinforcement probabilities for single stimuli in more complex ways. Capitalizing on this feature, we found that birds perform suboptimally when responses to single stimuli are extinguished. Moreover, the degree of suboptimality increased with the distance of the to-be-extinguished stimulus from the category boundary. This finding is seemingly at odds with data showing that animals perform near-optimally when payoff matrices are manipulated, including our own (Feng et al., 2009; Stüttgen et al., 2011b; but see Teichert and Ferrara, 2010). Importantly, the four pigeons which served as subjects for the present study also participated in an earlier experiment which was almost identical to the present with the important difference that reinforcement probabilities were identical for all stimuli within a category, ranging from 0.2 to 0.6 (Stüttgen et al., 2011b). In that study, the birds (after a few sessions) exhibited choice probabilities which maximized the number of earned reinforcers. Thus, we are left with an apparent discrepancy: in the previous experiment, animals performed optimally, in the present experiment, they did not. However, a simple process model based on earlier trial-by-trial accounts of signal detection performance (Dorfman and Biderman, 1971; Treisman and Williams, 1984) was able to reconcile the two disparate sets of data.

4.1. Learning from errors vs. learning from reinforcement

Our model builds exclusively on incomes. Dorfman et al. (1975) have argued that such models always lead to exclusive choice: since each reinforced response renders that response more likely to occur, a positive feedback loop will push the criterion towards \pm infinity, leading to consistent choice of only one response option. Therefore, these authors proposed that a model should incorporate criterion shifts for errors as well to prevent the criterion from drifting off (Dorfman and Biderman, 1971; Dorfman et al., 1975). We opted for a different solution, the constant drift of the criterion back to a neutral value (mathematically equivalent to a leaky integration of past incomes, see Section 2.4), as in Treisman's criterion setting theory (Treisman and Williams, 1984). One effect of leaky integration is that the effect of past outcomes on future choices diminishes over time, and past outcomes are forgotten at a rate that is specified

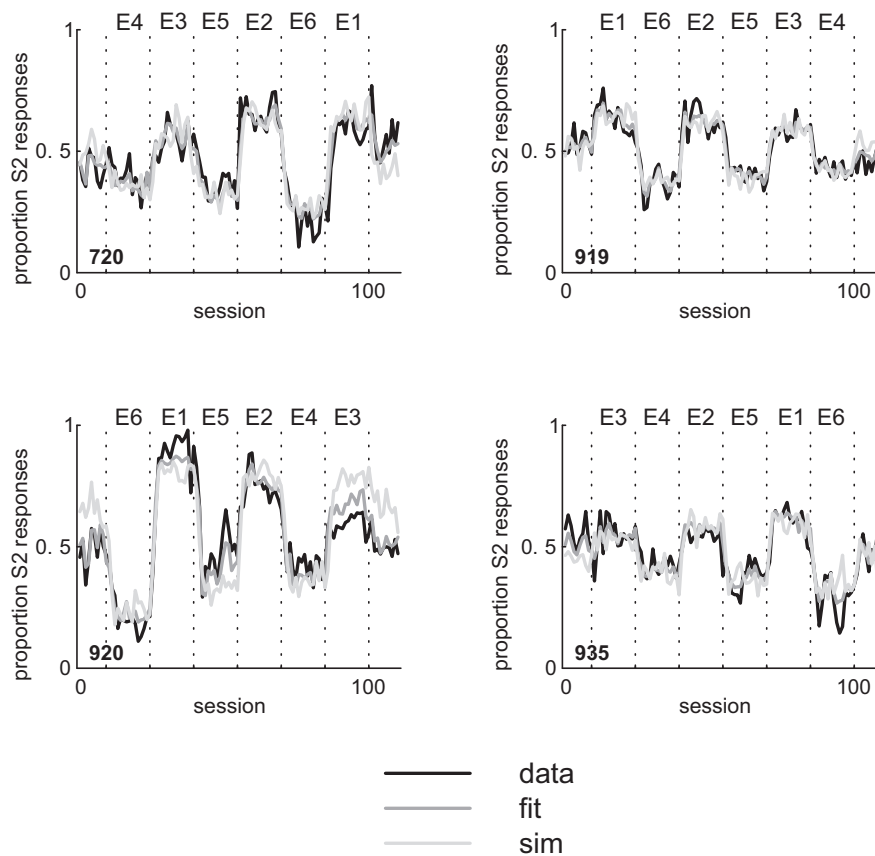


Fig. 7. Fit of the income-based learning model. Each panel shows the original response S2 fractions of one bird (black, same as in Fig. 4) along with the fitted data (dark gray) and forward-simulated data (light gray). See text for details.

by the leakage factor γ (see Section 2.5). Another effect of leaky integration is that, because of the constant drift back towards a neutral criterion, the criterion does not drift off towards $\pm\infty$. The reason we chose leaky integration over error correction as proposed by Dorfman and Biderman (1971) is that, in order to fit their data, they had to assume learning rates for errors that were several times larger than those for reinforcement. While we did not assess the relative importance of errors over reinforcement, nonsystematic investigations in our laboratory have yielded no convincing evidence that errors in this task contribute substantially to criterion setting. Also, errors were quite rare compared to reinforcements (percentage of correct responses generally exceeded 80%), so there would be less opportunity to learn from errors in this task than from reinforcement. In addition, when we fitted a model that only learns from errors it could not generate the data qualitatively when simulated forward. Nonetheless, we acknowledge that our choice of leaky integration over adding in an error correction component is somewhat arbitrary.

4.2. The effects of conditioned reinforcement

Our procedure involved a feedback signal for correct responses other than food delivery, namely the activation of the feeder light. It could be argued that the feeder light serves as conditioned reinforcer. We did not test whether the feeder light actually serves that role, but assuming it does would not affect our results, because the feeder light was equally present in all experimental conditions. Assuming conditioned and unconditioned reinforcers operate the same in choice, one can think of conditioned reinforcement as elevating the actual reinforcement magnitudes, for example from 0.5 reinforcers per trial to 0.6 reinforcers per trial.

Consequently, food omission in the experimental conditions would not reduce reinforcement magnitude to 0 but to 0.1. This would have a small quantitative but not a qualitative effect on our data – predicted decision criteria would only be somewhat less extreme.

4.3. Optimizing vs. non-optimizing choice algorithms

Animal subjects have been found to perform optimally in a wide range of behavioral tasks (Green et al., 1983; Hinson and Staddon, 1983; Sugrue et al., 2004; Corrado et al., 2005; Balci et al., 2009; Feng et al., 2009; Stüttgen et al., 2011b). While steady-state behavior is often described as being nearly optimal, the precise algorithm by which optimality is achieved is usually not specified, and the dearth of adequate process models is often noted (Boneau and Cole, 1967; Stüttgen et al., 2011a,b; Jones and Love, 2011). It is open whether optimization is explicitly factored in the choice algorithm, or whether it results as a by-product of another choice strategy such as equating returns at a short time scale (such as melioration; Herrnstein and Vaughan, 1980; Baum, 1981; Staddon and Hinson, 1983), or some sort of income-based algorithm as we used here (Gallistel et al., 2001; Corrado et al., 2005; Gallistel et al., 2007). Successful application of the income-based model to our previous results shows that quasi-optimal behavior can indeed – at least under some circumstances – arise from learning algorithms that make no explicit reference to optimality, as has been noted previously with respect to the matching law (Herrnstein and Vaughan, 1980; Vaughan, 1981; Sakai and Fukai, 2008). Importantly, conditions in our previous experiment were such that both optimality and the income-based choice mechanism predicted the same outcome in qualitative terms, namely stronger choice biases

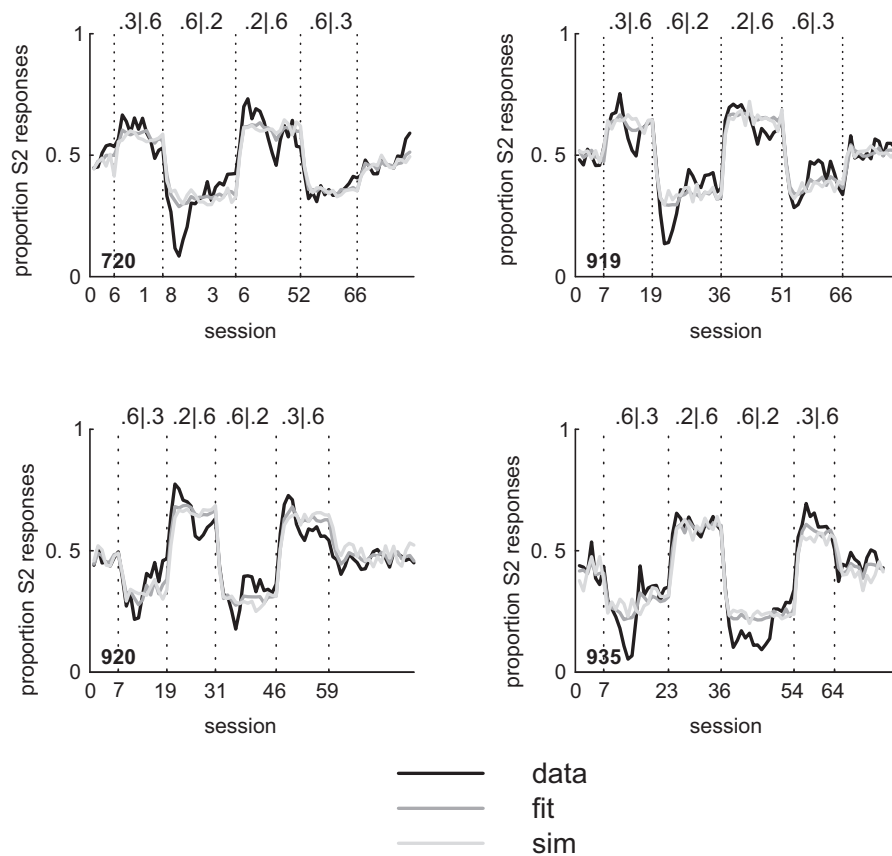


Fig. 8. Application of the income-based model to a previously published data set with asymmetrical payoff matrices (Stüttgen et al., 2011b). γ and δ parameters were taken from the fits to the present data set and applied to the earlier data set from the same birds. Conventions as in Fig. 7.

for the conditions with more extreme reinforcement ratios. Future work should examine the conditions under which the proposed choice algorithm approximates optimal behavior and investigate the degree to which error- or punishment-related criterion shifts influence adaptive choice behavior.

5. Conclusion

Our finding that animals fail to maximize reinforcement in a simple choice task poses important constraints on models for criterion setting by excluding all models based on the maximization of expected value (Boneau and Cole, 1967; Maddox, 2002; Stüttgen et al., 2011b). The income-based model presented in this paper is meant to provide a first attempt to unify two relatively separate areas of research, namely reward-based learning (mostly studied in animals) and psychophysics (mostly studied in human subjects). In addition, the model may serve a useful role in upcoming investigations in which behavioral output and single-neuron spike data are acquired simultaneously to provide hidden decision variables such as integrated incomes to be related to single-neuron spike output in decision-related brain areas (Seo and Lee, 2009; Starosta et al., 2013).

Acknowledgement

This research was supported in part by grants from the German Research Foundation to MCS (STU 544/1-1, FOR 1581 P1) and OG (SFB 874 B5, FOR 1581 P1).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.beproc.2013.02.014>.

References

- Balci, F., Freestone, D., Gallistel, C.R., 2009. Risk assessment in man and mouse. *Proc. Natl. Acad. Sci. U.S.A.* 106, 2459–2463.
- Baum, W.M., 1981. Optimization and the matching law as accounts of instrumental behavior. *J. Exp. Anal. Behav.* 36, 387–403.
- Boneau, C.A., Cole, J.L., 1967. Decision theory, the pigeon, and the psychophysical function. *Psychol. Rev.* 74, 123–135.
- Collett, D., 1991. *Modelling Binary Data*, first ed. Chapman & Hall, London.
- Corrado, G.S., Sugrue, L.P., Seung, H.S., Newsome, W.T., 2005. Linear–nonlinear–Poisson models of primate choice dynamics. *J. Exp. Anal. Behav.* 84, 581–617.
- Davison, M.C., Tustin, R.D., 1978. The relation between the generalized matching law and signal detection theory. *J. Exp. Anal. Behav.* 29, 331–336.
- Davison, M., Nevin, J.A., 1999. Stimuli, reinforcers, and behavior: an integration. *J. Exp. Anal. Behav.* 71, 439–482.
- Deneve, S., Latham, P.E., Pouget, A., 1999. Reading population codes: a neural implementation of ideal observers. *Nat. Neurosci.* 2, 740–745.
- Dorfman, D.D., 1973. Likelihood function of additive learning models – sufficient conditions for strict log-concavity and uniqueness of maximum. *J. Math. Psychol.* 10, 73–85.
- Dorfman, D.D., Biderman, M., 1971. A learning model for a continuum of sensory states. *J. Math. Psychol.* 8, 264–284.
- Dorfman, D.D., Saslow, C.F., Simpson, J.C., 1975. Learning models for a continuum of sensory states reexamined. *J. Math. Psychol.* 12, 178–211.
- Ernst, M.O., Banks, M.S., 2002. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* 415, 429–433.
- Feng, S., Holmes, P., Rorie, A., Newsome, W.T., 2009. Can monkeys choose optimally when faced with noisy stimuli and unequal rewards? *PLoS Comput. Biol.* 5, 1–15.

- Gallistel, C.R., King, A.P., Gottlieb, D., Balci, F., Papachristos, E.B., Szalecki, M., Carbone, K.S., 2007. Is matching innate? *J. Exp. Anal. Behav.* 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. *J. Exp. Psychol.: Anim. Behav. Process.* 27, 354–372.
- Gold, J.I., Shadlen, M.N., 2002. Banburismus and the brain: decoding the relationship between sensory stimuli, decisions, and reward. *Neuron* 36, 299–308.
- Green, L., Rachlin, H., Hanson, J., 1983. Matching and maximizing with concurrent ratio-interval schedules. *J. Exp. Anal. Behav.* 40, 217–224.
- Green, D.M., Swets, J.A., 1988. *Signal Detection Theory and Psychophysics*. Peninsula Publishing, Los Altos Hills.
- Herrnstein, R.J., 1961. Relative and absolute strength of response as a function of frequency of reinforcement. *J. Exp. Anal. Behav.* 4, 267–272.
- Herrnstein, R.J., Vaughan, W., 1980. Melioration and behavioral allocation. In: Staddon, J.E.R. (Ed.), *Limits to Action*. Academic Press, New York, pp. 143–176.
- Hinson, J.M., Staddon, J.E., 1983. Matching, maximizing, and hill-climbing. *J. Exp. Anal. Behav.* 40, 321–331.
- Jazayeri, M., Movshon, J.A., 2006. Optimal representation of sensory information by neural populations. *Nat. Neurosci.* 9, 690–696.
- Jones, M., Love, B.C., 2011. Bayesian fundamentalism or enlightenment? On the explanatory status and theoretical contributions of Bayesian models of cognition. *Behav. Brain Sci.* 34, 169–231.
- Kac, M., 1969. Some mathematical models in science. *Science* 166, 695–699.
- Körding, K.P., Wolpert, D.M., 2004. Bayesian integration in sensorimotor learning. *Nature* 427, 244–247.
- MacMillan, N.A., Creelman, C.D., 2005. *Detection Theory: A User's Guide*. Lawrence Erlbaum Associates, Inc., New Jersey.
- Maddox, W.T., 2002. Toward a unified theory of decision criterion learning in perceptual categorization. *J. Exp. Anal. Behav.* 78, 567–595.
- McCarthy, D., Davison, M., 1981. Towards a behavioral theory of bias in signal detection. *Percept. Psychophys.* 29, 371–382.
- Najemnik, J., Geisler, W.S., 2005. Optimal eye movement strategies in visual search. *Nature* 434, 387–391.
- Navalpakkam, V., Koch, C., Rangel, A., Perona, P., 2010. Optimal reward harvesting in complex perceptual environments. *Proc. Natl. Acad. Sci. U.S.A.* 107, 5232–5237.
- Nevin, J.A., Olson, K., Mandell, C., Yarensky, P., 1975. Differential reinforcement and signal detection. *J. Exp. Anal. Behav.* 24, 355–367.
- Newsome, W.T., Britten, K.H., Movshon, J.A., 1989. Neuronal correlates of a perceptual decision. *Nature* 341, 52–54.
- Norman, M.F., 1972. *Markov Processes and Learning Models*. Academic Press.
- Pastore, R.E., Crawley, E.J., Berens, M.S., Skelly, M.A., 2003. Nonparametric A' and other modern misconceptions about signal detection theory. *Psychon. Bull. Rev.* 10, 556–569.
- Pyke, G.H., Pulliam, H.R., Charnov, E.L., 1977. Optimal foraging – a selective review of theory and tests. *Quart. Rev. Biol.* 52, 137–154.
- Reynolds, G.S., 1961. Behavioral contrast. *J. Exp. Anal. Behav.* 4, 57–71.
- Rose, J., Otto, T., Dittrich, L., 2008. The biopsychology-toolbox: a free, open-source Matlab-toolbox for the control of behavioral experiments. *J. Neurosci. Methods* 175, 104–107.
- Sakai, Y., Fukai, T., 2008. When does reward maximization lead to the matching law? *PLoS ONE* 3, 1–7.
- Seo, H., Lee, D., 2009. Behavioral and neural changes after gains and losses of conditioned reinforcers. *J. Neurosci.* 29, 3627–3641.
- Staddon, J.E., Hinson, J.M., 1983. Optimization: a result or a mechanism? *Science* 221, 976–977.
- Starosta, S., Güntürkün, O., Stüttgen, M.C., 2013. Stimulus-Response-Outcome Coding in the Pigeon Nidopallium Caudolaterale. *PLoS ONE* 8, 1–11.
- Stüttgen, M.C., Schwarz, C., 2008. Psychophysical and neurometric detection performance under stimulus uncertainty. *Nat. Neurosci.* 11, 1091–1099.
- Stüttgen, M.C., Schwarz, C., Jäkel, F., 2011a. Mapping spikes to sensations. *Front. Neurosci.* 5, 125.
- Stüttgen, M.C., Yildiz, A., Güntürkün, O., 2011b. Adaptive criterion setting in perceptual decision-making. *J. Exp. Anal. Behav.* 96, 155–176.
- 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.
- Summerfield, C., Behrens, T.E., Koehlin, E., 2011. Perceptual classification in a rapidly changing environment. *Neuron* 71, 725–736.
- Teichert, T., Ferrara, V.P., 2010. Suboptimal integration of reward magnitude and prior reward likelihood in categorical decisions by monkeys. *Front. Neurosci.* 4, 1–13.
- Treisman, M., Williams, T.C., 1984. A theory of criterion setting with an application to sequential dependencies. *Psychol. Rev.* 91, 68–111.
- Trommershäuser, J., Gepshtein, S., Maloney, L.T., Landy, M.S., Banks, M.S., 2005. Optimal compensation for changes in task-relevant movement variability. *J. Neurosci.* 25, 7169–7178.
- Vaughan, W., 1981. Melioration, matching, and maximization. *J. Exp. Anal. Behav.* 36, 141–149.
- Wichmann, F.A., Hill, N.J., 2001. The psychometric function: I. Fitting, sampling, and goodness of fit. *Percept. Psychophys.* 63, 1293–1313.