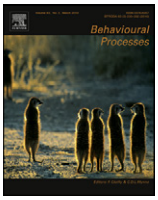
Behavioural Processes 96 (2013) 59–70

Contents lists available at SciVerse ScienceDirect

Behavioural Processes

journal homepage: www.elsevier.com/locate/behavproc



Suboptimal criterion setting in a perceptual choice task with asymmetric reinforcement

Maik C. Stüttgen[a,*], Nils Kasties[a,b], Daniel Lengersdorf[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] Insitute 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.

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.stuettgen@rub.de (M.C. Stüttgen).

0376-6357/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.beproc.2013.02.014

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 $\mathrm{Rf_{S1}}=1$ if an S1-category stimulus was presented and the subject's response was "S1" and led to reinforcement; let $\mathrm{Rf_{S1}}=0$ otherwise. Correspondingly, let $\mathrm{Rf_{S2}}=1$ if an S2-category stimulus was presented and the subject's response was "S2" and led to reinforcement; let $\mathrm{Rf_{S2}}=0$ otherwise. Then the probabilities for reinforcement on S1- and S2-trials ($E(\mathrm{Rf_{S1}})$ and $E(\mathrm{Rf_{S2}})$, respectively) are

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

and

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

The probability of obtaining a reinforcer in any given trial is the sum of the two: $E(\mathrm{Rf})=E(\mathrm{Rf_{S1}})+E(\mathrm{Rf_{S2}})$. $E(\mathrm{Rf})$ (or, more precisely, $E(\mathrm{Rf}|c)$) 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: $c(\mathrm{E3}) < c(\mathrm{E2}) < c(\mathrm{E1}) < c(\mathrm{E6}) < c(\mathrm{E5}) < c(\mathrm{E4})$ (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.