

# Can monkeys choose optimally when faced with noisy stimuli and unequal rewards?

S. Feng<sup>1</sup>, P. Holmes<sup>1,2</sup>, A. Rorie<sup>3</sup> and W.T. Newsome<sup>3</sup>

<sup>1</sup> Program in Applied and Computational Mathematics,

<sup>2</sup> Department of Mechanical and Aerospace Engineering,

Princeton University, Princeton, NJ 08544-1000, U.S.A.

<sup>3</sup> Howard Hughes Medical Institute and Department of Neurobiology,  
Stanford University, Stanford, CA 94305-2130, U.S.A.

August 21, 2008

## Abstract

We review the leaky competing accumulator model for two-alternative forced-choice decisions with cued responses, and propose extensions to account for the influence of unequal rewards. Assuming that stimulus information is integrated until the cue to respond arrives, and that firing rates of stimulus-selective neurons remain well within physiological bounds, the model reduces to an Ornstein-Uhlenbeck (OU) process which yields explicit expressions for the psychometric function that describes accuracy. From these we compute strategies that optimize the rewards expected over blocks of trials administered with mixed difficulty and reward contingencies. The psychometric function is characterized by two parameters: its midpoint slope, which quantifies a subject's ability to extract signal from noise, and its shift, which measures the bias applied to account for unequal rewards. We fit these to data from two monkeys performing the moving dots task with mixed coherences and reward schedules. We find that their behaviors averaged over multiple sessions are close to optimal, with shifts erring in the direction of smaller penalties. We propose two methods for biasing the OU process to produce such shifts.

**Keywords:** decision making, interrogation protocol, leaky accumulator, optimal behavior, Ornstein-Uhlenbeck process, psychometric function.

## Authors' summary

How do animals combine prior expectations with incoming sensory information, and can they appropriately weight these components so as to optimally shape their behavior? To investigate these questions we trained two monkeys to repeatedly indicate which one of two randomly-presented alternatives (1 and 2) appears in each of a sequence of trials. Before a noisy visual stimulus is presented on each trial, the monkey is informed of how correct identifications will be rewarded. Rewards may be equal for 1 and 2 and either high or low; or unequal, being high for 1 and low for 2 or vice versa. The four reward contingencies and the stimulus discriminabilities are randomly mixed from trial to trial.

To interpret the behavioral data we develop a decision-making model that averages incoming stimuli along with the prior reward information. This predicts a psychometric function that describes choice accuracy in terms of stimulus discriminability. Two parameters characterize this function: its *slope* quantifies a subject’s ability to extract signal from noise, and its *shift* measures the weight placed on the reward prior. Given a slope and a set of stimulus discriminabilities, we can therefore compute the shift that maximizes expected rewards over a series of many trials. Fitting the monkeys’ choice behavior to this model, we find that their performance is close to optimal: they garner 98 – 99.5% of their maximum possible rewards. This well-quantified study joins a growing body of work showing that animal foraging and human decision-making behaviors can approach optimality.

## 1 Introduction

There is increasing evidence from *in vivo* recordings in monkeys that oculomotor decision making in the brain mimics a drift-diffusion (DD) process, with neural activity rising to a threshold before movement initiation [41, 17, 31, 44]. In one well-studied task, monkeys are trained to decide the direction of motion of a field of randomly moving dots, a fraction of which move coherently in one of two possible target directions (T1 or T2), and to indicate their choice with a saccadic eye movement [6, 7, 43]. Varying the coherence level modulates the task difficulty, thereby influencing accuracy.

This paper addresses ongoing experiments on the motion discrimination task, but unlike most previous studies in which correct choices of either alternative are equally rewarded, the experiment is run under four conditions. Rewards may be high for both alternatives, low for both, high for T1 and low for T2, or low for T1 and high for T2. This design allows us to study the interaction between bottom-up (stimulus driven) and top-down (expectation driven) influences in a simple decision process. A second distinction with much previous work is that responses are delivered following a cue, rather than given freely. We idealize this as an interrogation protocol (cf. [4]), in which accumulated information is assessed at the time of the cue rather than when it passes a threshold, and we model the accumulation by an Ornstein-Uhlenbeck (OU) process. Closely related work on human decision making is reported in [14, 33].

Consistent with random walk and diffusion processes [28, 36, 30, 29, 39, 44], neural activity in brain areas involved in preparing eye movements, including the lateral intraparietal area (LIP), frontal eye field and superior colliculus [43, 26, 23, 24], exhibits an accumulation over time of the motion evidence represented in the middle temporal area (MT) of extrastriate visual cortex. Under free response conditions, firing rates in area LIP reach a threshold level just prior to the saccade [40]. Further strengthening the connection, it has recently been shown that models of LIP using heterogeneous pools of spiking neurons can reproduce key features of this accumulation process [48, 49], and that the averaged activities of sub-populations selective for the target directions behave much like the two units of the leaky competing accumulator (LCA) model of Usher and McClelland [45]. In turn, under suitable constraints, the LCA can be reduced to a one-dimensional OU process: a generalization of the simpler DD process [9, 8, 4]. This allows us to obtain explicit expressions for psychometric functions (PMFs) that describe accuracy in terms of model and experimental parameters, and to predict how they should be shifted to maximize expected returns in case of unequal rewards.

The goals of this work are to show that PMFs derived from the OU model describe animal

data well, that they can accommodate reward information and allow optimal performance to be predicted analytically, and finally, to compare animal behaviors with those predictions. Analyzing data from two monkeys, we find that, when faced with unequal rewards, both animals bias their PMFs in the appropriate directions, but by amounts larger than the optimal shifts. However, in doing so they respectively sacrifice less than 1% and 2% of their expected maximum rewards, for all coherence conditions, based on their signal-discrimination abilities (sensitivities), averaged over all session of trials. They achieve this in spite of significant variability from session to session, across which the parameters that describe their sensitivity to stimuli and reward biases show little correlation with the relationships that optimality theory predicts.

This paper extends a recent study that describes fits of behavioral data from monkeys learning the moving dots task, which also shows that DD and OU processes can provide good descriptions of psychometric functions (PMFs) [12]. A related study of humans and mice performing a task that requires time estimation [3] shows that those subjects also approached optimal behavior. The paper is organised as follows. After reviewing the experimental method in §2, in §3 we describe the LCA model and its reduction to OU and DD processes; we then propose simple models for the influence of biased rewards and display examples of the resulting psychometric functions. §4 contains the optimality analysis, and in §5 the models are fitted to data from two animals, and their performances assessed. A discussion ensues in §6.

## 2 Methods I: behavioral studies

To motivate the theoretical developments of §3, we start by briefly describing the experiment. More details will be provided, along with reports of electrophysiological data, in a subsequent publication.

### 2.1 Procedures

Two adult male rhesus monkeys, A and T (12 and 14 kg), were trained on a two-alternative, forced-choice, motion discrimination task with multiple reward contingences. Daily access to fluids was controlled during training and experimental periods to promote behavioral motivation. Prior to training, the monkeys were prepared surgically with a head-holding device [13] and a scleral search coil for monitoring eye position [25]. All surgical, behavioral, and animal care procedures complied with National Institutes of Health guidelines and were approved by the Stanford University Institutional Animal Care and Use Committee.

During both training and experimental sessions monkeys sat in a primate chair at a viewing distance of 57 cm from a color monitor, on which visual stimuli were presented under computer control. The monkeys' heads were positioned stably using the head-holding device, and eye position was monitored with a magnetic search coil apparatus (0.1° resolution; CNC Engineering, Seattle, WA). Behavioral control and data acquisition were managed by a PC-compatible computer running the QNX Software Systems (Ottawa, Canada) real-time operating system. The experimental paradigm was implemented in the NIH Rex programming environment [22]. Visual stimuli were generated by a second computer and displayed using the Cambridge Research Systems VSG (Kent, UK) graphics card and accompanying software. Liquid rewards were delivered via a gravity-fed juice tube placed near the animal's mouth, activated by a computer-controlled solenoid valve. Subsequent data analyses and computer simulations were performed using the Mathworks MATLAB (Natick, MA) programming environment.

## 2.2 Motion stimulus

The monkeys performed a two-alternative, forced-choice, motion discrimination task that has been used extensively to study both visual motion perception (e.g. [34, 35, 10]) and visually-based decision making [42, 23, 21]. The stimulus is composed of white dots, viewed through a circular aperture, on a dark computer screen. On each trial a variable proportion of the dots moved coherently in one of two opposite directions while the remaining dots flashed transiently at random locations and times (for details see [6]), and the animals reported which of two possible directions of motion was present. Discriminability was varied parametrically from trial to trial by adjusting the percentage of the dots in coherent motion: the task was easy if a large proportion of dots moved coherently (i.e. 50% or 100% coherence), but became progressively more difficult as coherence decreased. In what follows we indicate the motion direction by signing the coherence: thus +25% and -25% coherences are equally difficult to discriminate, but the coherent dots move in opposite directions. Typically, the animals viewed a range of signed coherences spanning psychophysical threshold. Animals were always rewarded for indicating the correct direction of motion, except that 0% coherence was rewarded randomly (50% probability) irrespective of their choices.

## 2.3 Experimental paradigm

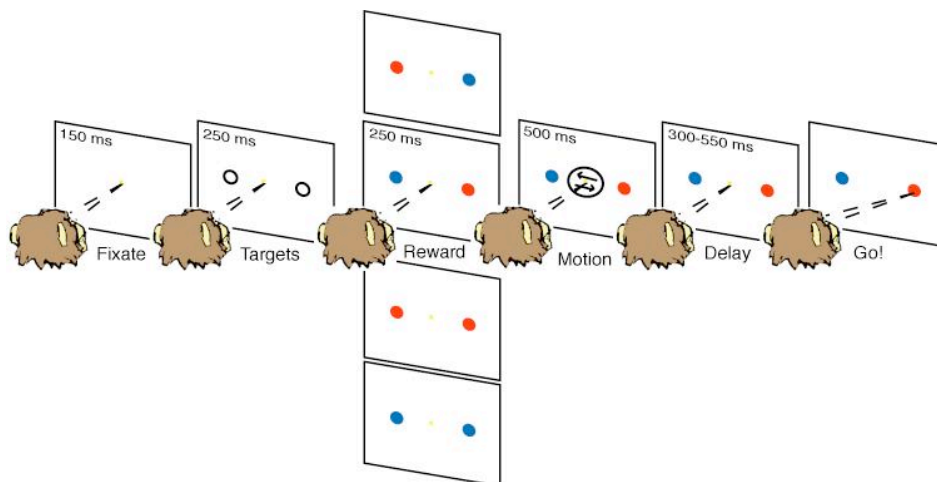


Figure 1: The motion discrimination task. Target colors cue the magnitude of rewards for correct responses, red denoting a value twice that of blue. The four panels in the reward segment show the possible reward conditions. See text for full description.

The horizontal row of panels in Fig. 1 illustrates the sequence of events comprising a typical trial, which began with the onset of a small, yellow dot that the monkey must visually fixate for 150 msec. Next, two saccade targets appeared (open gray circles)  $10^\circ$  eccentric from the visual fixation point and  $180^\circ$  apart from each other, in-line with the axis of motion to be discriminated. By convention, target 1 (T1) corresponds to positive coherence and target 2 (T2) to negative coherence. After 250 msec the targets changed color, indicating the magnitude of reward available for correctly choosing that target. A blue target indicated a low magnitude (L) reward (1 unit,

$\approx 0.12$  ml of juice), while a red target indicated a high magnitude (H) reward (2 units). There were four reward conditions overall, schematized by the column of four panels in the Reward segment of Fig. 1: 1) LL, in which both targets were blue, 2) HH, in which both were red, 3) HL, in which T1 was red and T2 blue, and 4) LH: the mirror image of HL.

The colored targets were visible for 250 msec prior to onset of the motion stimulus which appeared for 500 msec, centered on the fixation point. Following stimulus offset, the monkey was required to maintain fixation for a variable delay period (300-550 msec) after which the fixation point disappeared, cueing the monkey to report his decision with a saccade to the target corresponding to the perceived direction of motion. The monkey was given a grace period of 1000 msec to respond. If he chose the correct direction, he received the reward indicated by the color of the chosen target. Fixation was enforced throughout the trial by requiring the monkey to maintain its eye position within an electronic window ( $1.25^\circ$  radius) centered on the fixation point. Inappropriate breaks of fixation were punished by aborting the trial and enforcing a time-out period before onset of the next trial. Psychophysical decisions were identified by detecting the time of arrival of the monkey’s eye in one of two electronic windows ( $1.25^\circ$  radius) centered on the choice targets.

Trials were presented pseudo-randomly in block-randomized order. For monkey A, we employed 12 signed coherences, 0% coherence and four reward conditions, yielding 52 conditions overall. For monkey T we eliminated two of the lowest motion coherences because this animal’s psychophysical thresholds were somewhat higher than those of monkey A, giving 36 conditions overall. We attempted to acquire 40 trials for each condition, enabling us to characterize a full psychometric function for each reward condition, but because the behavioral data were obtained simultaneously with electrophysiological recordings, we did not always acquire a full set for each condition (the experiment typically ended when single unit isolation was lost). For the data reported in this paper, the number of repetitions obtained for each experiment ranged from 19 to 40 with a mean of 36. The behavioral data analyzed here consists of 35 sessions from monkey A and 25 sessions from monkey T.

### 3 Methods II: models for evidence accumulation and choice

Here we describe a simple model for two-alternative forced-choice (2AFC) tasks. Several other models are reviewed in [4], along with the relations among them and conditions under which they can be reduced to OU and DD processes.

#### 3.1 The leaky competing accumulator model

The LCA is a stochastic differential equation [1] whose states  $(x_1(t), x_2(t))$  describe the activities of two mutually-inhibiting neural populations, each of which receives noisy sensory input from the stimulus, and also, in the instantiation developed here, input derived from reward expectations. See [32, 45]. The system may be written as

$$dx_1 = [-\gamma x_1 - \beta f(x_2) + I_1(t)] dt + \sigma dW_1, \tag{1}$$

$$dx_2 = [-\gamma x_2 - \beta f(x_1) + I_2(t)] dt + \sigma dW_2, \tag{2}$$

where  $f(\cdot)$  is a sigmoidal-type activation (or input-output) function,  $\gamma$  and  $\beta$  respectively denote the strengths of leak and inhibition, and  $\sigma dW_j$  are independent white noise (Weiner) increments of r.m.s. strength  $\sigma$ . The inputs  $I_j(t)$  are in general time-dependent, since stimulus and expectation

effects can vary over the course of a trial. To fix ideas, we may suppose that the states  $(x_1(t), x_2(t))$  represent short-term averaged firing rates of LIP neurons sensitive to alternatives 1 and 2. We recognize that the decision may be formed by interactions among several oculomotor areas, but note that a partial causal role for LIP has been demonstrated [21].

Under the interrogation protocol the choice is determined by the difference  $x \stackrel{\text{def}}{=} x_1(t) - x_2(t)$ : if  $x > 0$ , T1 is chosen, and if  $x < 0$ , T2 is chosen. As explained in [4], this models the “hard limit” of a cued response, in which subjects may not answer before the cue, and must answer within a short window following it, to qualify for a reward.

### 3.2 Reduction to an Ornstein-Uhlenbeck process

In the absence of noise ( $\sigma = 0$ ) and with *constant* inputs  $I_1, I_2$ , equilibrium solutions of Eqs. (1-2) lie at the intersections of the nullclines given by  $\gamma x_1 = -\beta f(x_2) + I_1(t)$  and  $\gamma x_2 = -\beta f(x_1) + I_2(t)$ , and, depending on the values of the parameters  $\gamma, \beta, I_1, I_2$  and the precise form of  $f(\cdot)$ , there may be one, two or three stable equilibria, corresponding to low activity in both populations, high activity in  $x_1$  and low in  $x_2$ , and vice-versa. If the nullclines lie sufficiently close to each other over the activity range that encompasses the equilibria, it follows that a one-dimensional, attracting, slow manifold exists that contains both stable and unstable equilibria, and solutions that connect them [20, 9]: see Fig. 2. With  $\sigma \neq 0$  (and  $I_j(t)$  non-constant), we must appeal to the theory of stochastic center manifolds to draw a similar, probabilistic conclusion [27, 5] and [2, Chap. 7].

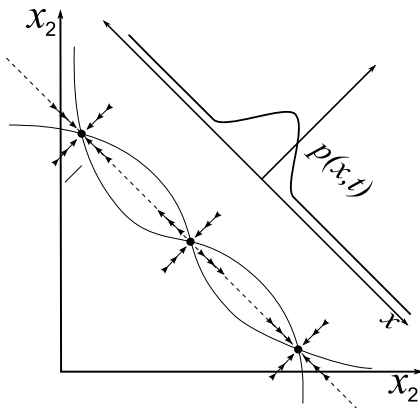


Figure 2: A typical state space of the LCA model, showing nullclines on which  $dx_j = 0$  for  $\sigma = 0$  (thin curves), fixed points (filled circles with arrows indicating stability types) and slow manifold (dashed line). Diagonal solid line represents one-dimensional state space  $x$  of reduced OU model, with associated probability distribution  $p(x, t)$  of sample paths.

To illustrate, we simplify Eqs. (1-2) by linearizing the sigmoidal function at the central equilibrium point  $(\bar{x}, \bar{x})$  in the case of equal inputs  $I_j(t) \equiv I$ , where  $\bar{x} = [-\beta f(\bar{x}) + I]/\gamma$ . Parameterizing the sigmoid so that  $df/dx(\bar{x}) = 1$ , Eqs. (1-2) become

$$dx_1 = [-\gamma x_1 - \beta x_2 + I(t)] dt + \sigma dW_1, \quad (3)$$

$$dx_2 = [-\gamma x_2 - \beta x_1 + I(t)] dt + \sigma dW_2, \quad (4)$$

and subtracting these equations yields a single scalar SDE for the activity difference  $x$ :

$$dx = [\lambda x + A(t)] dt + \sigma dW, \quad (5)$$

where  $\lambda = \beta - \gamma$ ,  $A(t) = I_1(t) - I_2(t)$  and  $dW = dW_1 - dW_2$  are independent white noise increments. Thus, if stimulus A is displayed, we expect  $A = I_1 - I_2 > 0$  and vice versa.

Eq. (5) describes an OU process, or, for  $\lambda = 0$ , a DD process. The DD process is a continuum limit of the sequential probability ratio test [4], which is optimal for 2AFC tasks in that it delivers a decision of guaranteed accuracy in the shortest possible time, or that, given a fixed decision time, it maximizes accuracy [46, 47]. The latter case is relevant to the cued responses considered here.

### 3.3 Prediction of psychometric functions

The probability of choosing alternative 1 under the interrogation protocol can be computed from the probability distribution of solutions  $p(x, t)$  of Eq. (5), which is governed by the forward Kolmogorov or Fokker-Planck equation [15]:

$$\frac{\partial p}{\partial t} = -\frac{\partial}{\partial x} [(\lambda x + A(t))p] + \frac{\sigma^2}{2} \frac{\partial^2 p}{\partial x^2}. \quad (6)$$

When the distribution of initial data is a Gaussian (normal) centered about  $\mu_0$ ,

$$p(x, 0) = \frac{1}{\sqrt{2\pi\nu_0}} \exp\left[-\frac{(x - \mu_0)^2}{2\nu_0}\right], \quad (7)$$

solutions of (6) remain Gaussian as time evolves:

$$p(x, t) = \frac{1}{\sqrt{2\pi\nu(t)}} \exp\left[-\frac{(x - \mu(t))^2}{2\nu(t)}\right], \quad \text{where} \quad (8)$$

$$\mu(t) = \mu_0 e^{\lambda t} + \int_0^t e^{\lambda(t-s)} A(s) ds \quad \text{and} \quad \nu(t) = \nu_0 e^{2\lambda t} + \frac{\sigma^2}{2\lambda} (e^{2\lambda t} - 1) \quad (9)$$

contain integrated stimulus and noise respectively. Note that  $\nu(t) > 0$  regardless of the sign of  $\lambda$ , so the square root in Eq. (11) is well-defined. In the DD limit  $\lambda = 0$   $\mu(t)$  and  $\nu(t)$  simplify to

$$\mu(t) = \mu_0 + \int_0^t A(s) ds \quad \text{and} \quad \nu(t) = \nu_0 + \sigma^2 t. \quad (10)$$

Henceforth we set  $\nu_0 = 0$ , assuming that all sample paths start from the same initial condition  $x(0) = \mu_0$ . From Eq. (10) the probability that T1 is chosen at time  $t = T$  can be computed explicitly as a cumulative normal distribution:

$$P(T) = \int_0^\infty p(x, T) dx = \frac{1}{2} \left[ 1 + \operatorname{erf}\left(\frac{\mu(T)}{\sqrt{2\nu(T)}}\right) \right]. \quad (11)$$

Here  $\operatorname{erf}(y) = (2/\sqrt{\pi}) \int_0^y \exp(-u^2) du$  denotes the error function and Eq. (11) represents a *psychometric function* (PMF) whose values rise from 0 to 1 as the argument  $(\mu/\sqrt{2\nu})$  runs from  $-\infty$  to  $+\infty$ , so multiplying it by 100 gives the expected percentage of T1 choices.

In addition to its dependence on viewing time  $T$ , the PMF also depends on the functional forms of the drift and noise terms embedded in  $\mu(t)$  and  $\nu(t)$ . In particular  $\mu(t)$  depends on the coherence or stimulus strength via  $A(t)$ , and upon prior expectations or biases that reward information might introduce, for example via  $\mu_0$ . Examples are provided in §3.4. To emphasize this we sometimes write the PMF as  $P(C, T)$  or  $P(C, T; \mu_0)$ , to denote its dependence on  $C$  and other parameters. Specifically, we shall examine two aspects of the PMF as a function of  $C$ : the *slope*  $\frac{dP(t)}{dC}$  at 50% accuracy, and the *shift*: the value of  $C$  at which  $P(C, T) = 0.5$ , or equivalently, where  $\mu(T) = 0$ .

### 3.4 Models of stimuli and reward biasing

Following [16, 18], we suppose that the part of the drift rate due to the stimulus depends linearly on coherence:  $A_{\text{stim}} = aC$ . (While power-law dependence on  $C$  has been introduced to account for behavior early in training, a linear relationship seems generally adequate for well-trained animals [18].) Here  $C \in [-1, 1]$  (between 100% leftward and 100% rightward motion coherence), as determined by the experimenter, and  $a$  is a scaling or sensitivity parameter that allows one to fit data from different subjects, or from one subject during different epochs of training [12, Fig. 14].

We propose two strategies to account for prior reward information. The first and simplest is to bias the initial condition at stimulus onset  $t = 0$ , taking  $x(0) = \mu_0 > 0$  if T1 garners a higher reward (HL) and  $x(0) = -\mu_0 < 0$  if T2 does so (LH), with  $x(0) = 0$  for equal rewards (LL and HH). In this case, from Eq. (9), the integrated drift rate and noise levels are:

$$\mu(C, t) = \mu_0 e^{\lambda t} + \frac{aC}{\lambda} (e^{\lambda t} - 1) \quad \text{and} \quad \nu(t) = \frac{\sigma^2}{2\lambda} (e^{2\lambda t} - 1), \quad \text{where } t \in [0, T], \quad (12)$$

and the decision is rendered at the end of the motion period  $t = T$ . Such biasing of initial data is optimal for the free response protocol if coherences remain fixed over each block of trials [4], but, as we shall see, other strategies can do equally well under the interrogation protocol.

Indeed, motivated by the task sequence of Fig. 1, and as suggested by J.L. McClelland (personal communication), one can alternatively assume that bias enters throughout a reward cue period of duration  $\tau$  and the ensuing motion period, as a drift term upon upon which the stimulus is additively superimposed to form a piecewise-constant drift rate:

$$A(C, t) = \begin{cases} b, & -\tau \leq t < 0, \\ b + aC, & 0 \leq t \leq T. \end{cases} \quad (13)$$

From Eqs. (9) the resulting integrated drift and noise during the motion period  $[0, T]$  are now:

$$\mu(C, t) = \frac{b}{\lambda} (e^{\lambda(t+\tau)} - 1) + \frac{aC}{\lambda} (e^{\lambda t} - 1) \quad \text{and} \quad \nu(t) = \frac{\sigma^2}{2\lambda} (e^{2\lambda(t+\tau)} - 1). \quad (14)$$

Here we set  $\mu_0 = 0$ , since  $b$  accounts for reward bias, with  $b > 0$  if T1 has higher reward,  $b < 0$  if T2 has higher reward and  $b = 0$  for equal rewards. Note that accumulation of reward information now begins at  $t = -\tau$ .

The first model assumes that reward information is assimilated during the reward cue period  $[-\tau, 0)$  and loaded into the initial accumulator state  $\mu_0$  at motion onset, after which it is effectively displaced by the stimulus. In the second strategy the reward information  $b$  continues to apply pressure throughout the motion period  $[0, T]$ . (Presumably  $\mu_0$  and  $b$  should scale monotonically, but not necessarily linearly, with reward ratio.) These represent extremes of a range of possible

strategies. More complex time-varying drift functions could be proposed to model reward expectations, waxing and waning attention to stimuli, and for the fixation, target and delay periods, but analyses of electrophysiological data (LIP firing rates), currently in progress, are required to inform such detailed modeling. Here we simply assume that the accumulation process starts at reward cue onset ( $t = 0$  or  $t = -\tau$ ) and ends at motion offset ( $t = T$ ). Moreover, as we now show, until experiments with variable stimulus and/or reward cue viewing times are run, it is impossible to distinguish between models even as simple as those described above.

The PMF (11) depends only upon the ratio  $\mu(T)/\sqrt{2\nu(T)}$  (which is one half the discriminability factor  $d'$  of [45, Eq. (7)], cf. [19]), and in Eqs. (12) and (14) reward biases appear as additive factors in the numerator  $\mu(T)$ . Thus, if all parameters other than  $C$  are fixed, and  $C$  appears linearly as assumed above, the argument of the PMF can be written in both cases in the simple form  $b_1(C + b_2)$ , so that

$$P(C, T) = \frac{1}{2} \left[ 1 + \operatorname{erf} \left( \frac{\mu(T)}{\sqrt{2\nu(T)}} \right) \right] = \frac{1 + \operatorname{erf}[b_1(C + b_2)]}{2}. \quad (15)$$

Here  $b_1$  and  $b_2$  respectively determine the slope and shift of the PMF: the slope at 50% T1 choices being  $b_1/\sqrt{\pi}$  in the units of probability of a T1 choice per % coherence, and  $b_2$  having the units of % coherence. In turn,  $b_1$  and  $b_2$  depend upon the parameters  $a, \sigma, \lambda, \mu_0, T, b$ , and  $\tau$  introduced above; for the specific cases of Eqs. (12) and (14), we respectively have:

$$b_1 = \frac{a(e^{\lambda T} - 1)}{\sigma\sqrt{\lambda}(e^{2\lambda T} - 1)}, \quad b_2 = \frac{\mu_0\lambda e^{\lambda T}}{a(e^{\lambda T} - 1)}, \quad (16)$$

$$\text{and} \quad b_1 = \frac{a(e^{\lambda T} - 1)}{\sigma\sqrt{\lambda}(e^{2\lambda(\tau+T)} - 1)}, \quad b_2 = \frac{b(e^{\lambda(\tau+T)} - 1)}{a(e^{\lambda T} - 1)}. \quad (17)$$

The ratios  $a/\sigma$  and  $\mu_0/a$  or  $b/a$  in Eqs. (16) and (17) characterize a subject's ability to extract information from the noisy stimulus, and the weight placed on reward information relative to stimulus. Experiments in which  $\tau$  and  $T$  are varied independently could in principle distinguish between these cases, but with the present data we can only fit the slope  $b_1$  and shift  $b_2$ . Nor can we determine whether the process is best described by a pure DD process with  $\lambda = 0$  and constant drift  $A$ , or an OU process with  $\lambda \neq 0$ , or, indeed, whether the drift rate varies with time. Recent experiments on human subjects with biased rewards that use a range of interrogation times [14, 33] suggests that a leaky competing accumulator model [45] is indeed appropriate, and data from those experiments may allow such distinctions to be made.

### 3.5 Examples of psychometric functions

To illustrate how PMFs depend upon the parameters describing evidence accumulation ( $a, C, \sigma, \lambda, T$ ) and reward biasing ( $b, \tau$ ), we compute examples based on the second model of §3.4. Substituting the expressions (14) in Eq. (11), we obtain:

$$P(C, T) = \frac{1}{2} \left[ 1 + \operatorname{erf} \left( \frac{b(e^{\lambda(\tau+T)} - 1) + aC(e^{\lambda T} - 1)}{\sigma\sqrt{\lambda}(e^{2\lambda(\tau+T)} - 1)} \right) \right]. \quad (18)$$

In case  $\lambda = 0$  the exponential expressions simplify (cf. Eqs. (10)), giving:

$$P(C, T) = \frac{1}{2} \left[ 1 + \operatorname{erf} \left( \frac{b(\tau + T) + aCT}{\sigma\sqrt{2(\tau + T)}} \right) \right]. \quad (19)$$

Examples of these PMFs are plotted in Fig. 3 for  $\lambda < 0$ ,  $\lambda > 0$  and  $\lambda = 0$ . Parameter values, listed in the caption, are chosen to illustrate qualitative trends. Note that the slopes of the functions are lower for  $\lambda \neq 0$  (top row) than for  $\lambda = 0$  (bottom), and lowest for  $\lambda > 0$  (top right), illustrating that the DD process  $\lambda = 0$  is optimal. Also, for fixed  $a, b, \tau$  and  $T$ , the PMFs are shifted to the left or right for  $b > 0$  and  $b < 0$  respectively, by an amount that grows as  $\lambda$  increases from negative to positive.

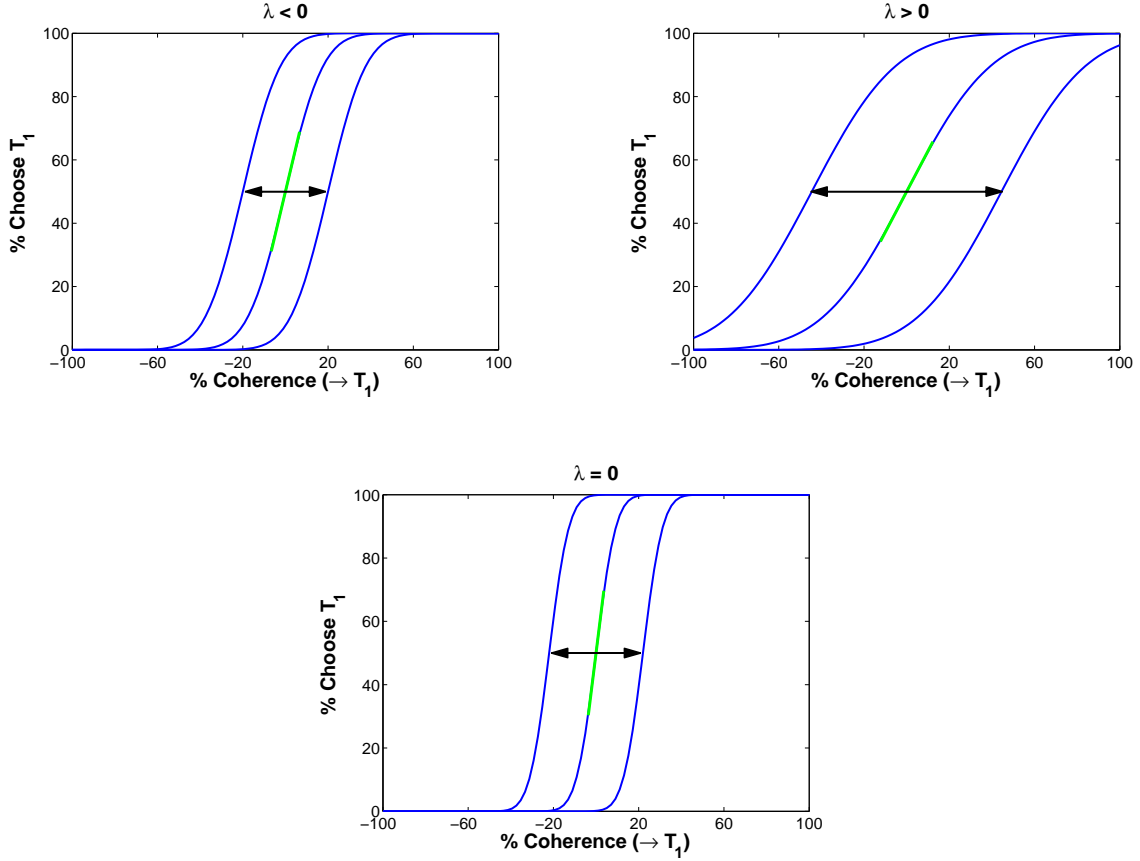


Figure 3: Psychometric functions showing fraction of T1 choices as a function of coherence  $C$  for constant reward bias  $b$  applied before and during motion period. Top left:  $\lambda = -0.2$ ; top right:  $\lambda = +0.2$ ; bottom:  $\lambda = 0$ ; each panel shows the cases  $b = +0.1, 0$  and  $-0.1$  (left to right). Remaining parameters are  $a = 0.005$ ,  $\sigma = 0.2214$ ,  $\tau = 4$  and  $T = 40$  (arbitrary time units). Green lines indicate slopes for zero bias; arrows show shifts.

To understand these trends, we recall that a stable OU process ( $\lambda < 0$ ) exhibits recency effects while an unstable one ( $\lambda > 0$ ) exhibits primacy effects [45]. In the former case information arriving early decays, while for  $\lambda > 0$  it grows, so that reward information in the pre-stimulus cue period exerts a greater influence, leading to greater shifts. Unstable OU processes also yield lower accuracy than stable processes. Specifically, the factor  $\sqrt{\exp(2\lambda(\tau \dots))}$  in Eq. (18) reflects the fact that noise accumulates during the cue period, leading to accelerating growth of solutions when  $\lambda > 0$  which the stimulus cannot repair. In general, while accuracy increases monotonically with viewing time,

it approaches a limit below 100% for any  $\lambda \neq 0$ : specifically:

$$\lim_{T \rightarrow \infty} P(C, T) = \begin{cases} \frac{1}{2} \left[ 1 + \operatorname{erf} \left( \frac{b+aC}{\sigma\sqrt{|\lambda|}} \right) \right], & \text{for } \lambda < 0, \\ \frac{1}{2} \left[ 1 + \operatorname{erf} \left( \frac{b+aCe^{-\lambda\tau}}{\sigma\sqrt{\lambda}} \right) \right], & \text{for } \lambda > 0. \end{cases} \quad (20)$$

The slopes of the PMF can clearly be increased by setting  $\lambda = 0$  and raising the sensitivity-to-noise ratio  $a/\sigma$ , but these parameters are constrained for individual subjects by physiological factors and by training. Indeed, Eckhoff et al. [12] find that  $a/\sigma$  and  $\lambda$  remain stable over relatively long periods (several sessions) for trained animals. As noted in §3.4 the present data does not allow us to estimate such “detailed” parameters. In the analysis to follow we therefore adopt the two-parameter form of Eq. (15), regarding the PMF slope  $b_1$ , which quantifies sensitivity to stimulus, as fixed, and seeking shifts in  $b_2$  that maximize the overall expected reward for that sensitivity, although this implies a causal chain that animals may not follow, as we note in §6.

## 4 Results I: optimality analysis

Given a fixed slope  $b_1$ , we now ask what is the shift  $b_2$  in the PMF that maximizes expected rewards in the case that the two alternatives are unequally rewarded. How much should the subject weight the reward information relative to that in the stimulus, in order to make optimal use of both?

### 4.1 Two motivating examples

Let  $r$  denote the reward obtained on a typical trial, namely,  $r_1$  if alternative 1 is offered and chosen, and  $r_2$  if 2 is offered and chosen. The *expected reward*  $\mathbb{E}[r]$  is obtained by multiplying each  $r_j$  by the probability that the corresponding alternative is chosen, when it appears in the stimulus. To make this explicit, first suppose that coherence is fixed from trial to trial and that the two possible stimuli  $C = +\bar{C}$  (T1) and  $C = -\bar{C}$  (T2) are equally likely. In this case

$$\mathbb{E}[r] = r_1 \frac{P(+\bar{C}; b_1, b_2)}{2} + r_2 \frac{[1 - P(-\bar{C}; b_1, b_2)]}{2}, \quad (21)$$

where we use the fact that  $P(C; b_1, b_2)$  and  $[1 - P(-C; b_1, b_2)]$  are the average proportions of correct T1 choices and T2 choices for coherences  $\pm C$  and we write the argument of  $P$  explicitly to indicate its dependence on coherence and the slope and bias parameters introduced in §3.4.

Using Eq. (15) and the fact that

$$\frac{d}{du} \operatorname{erf}(u) = \frac{2}{\sqrt{\pi}} \exp(-u^2), \quad (22)$$

we may compute the derivatives of  $P(\pm\bar{C}; b_1, b_2)$  with respect to  $b_2$  to derive a necessary condition for a maximum in  $\mathbb{E}[r]$ :

$$\frac{\partial \mathbb{E}[r]}{\partial b_2} = \frac{b_1 [r_1 \exp(-b_1^2(b_2 + \bar{C})^2) - r_2 \exp(-b_1^2(b_2 - \bar{C})^2)]}{2\sqrt{\pi}} = 0. \quad (23)$$

This implies that

$$\frac{r_1}{r_2} = \frac{\exp(-b_1^2(b_2 - \bar{C})^2)}{\exp(-b_1^2(b_2 + \bar{C})^2)} = \exp(4b_1^2 b_2 \bar{C}), \quad \text{and thus } b_2^{\text{opt}} = \frac{1}{4b_1^2 \bar{C}} \ln \left( \frac{r_1}{r_2} \right). \quad (24)$$

To verify that (24) identifies the global maximum we compute the second derivative at  $b_2 = b_2^{\text{opt}}$ :

$$\left. \frac{\partial^2 \mathbb{E}[r]}{\partial b_2^2} \right|_{b_2^{\text{opt}}} = \frac{-b_1^3 b_2 \bar{C}}{\sqrt{\pi}} [r_1 \exp(-b_1^2(b_2 + \bar{C})^2) + r_2 \exp(-b_1^2(b_2 - \bar{C})^2)] < 0. \quad (25)$$

For equal rewards  $r_1 = r_2$  we recover  $b_2^{\text{opt}} = 0$ : an unbiased PMF with  $P(0; b_1, 0) = 0.5$ , and for a fixed reward ratio,  $b_2^{\text{opt}}$  varies inversely with  $\bar{C}$ , approaching  $\infty$  as  $\bar{C} \rightarrow 0$ . In this limit the stimulus contains no information and it is best to always choose the more lavishly rewarded alternative. Fig. 4 (left panel, solid blue curves) shows examples of  $b_2^{\text{opt}}$  plotted as a function of reward ratio for fixed  $b_1$  and three different coherence levels.

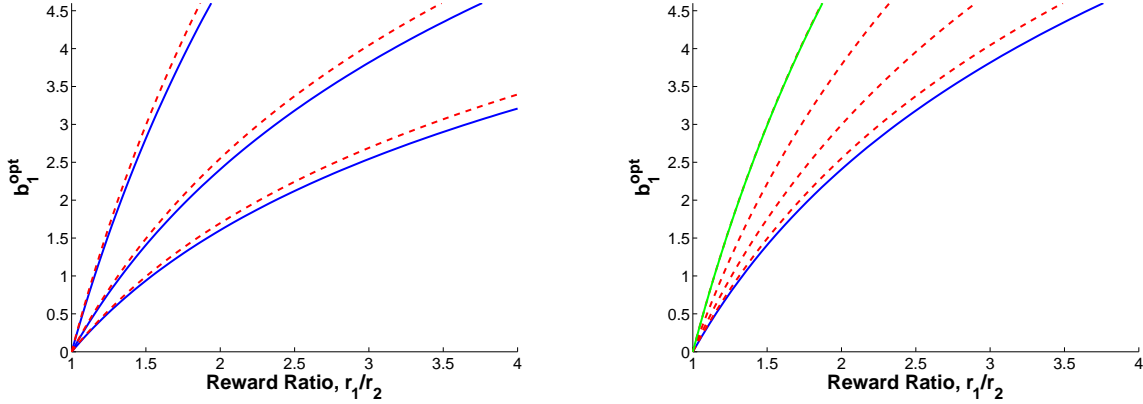


Figure 4: Optimal shifts  $b_2^{\text{opt}}$  as a function of the reward ratio  $r_1/r_2$  for fixed coherences (solid blue curves) and for coherence ranges centered on the fixed coherences (dashed red curves). Left:  $\bar{C} = 10, 20$  and  $30\%$  (top left to bottom right, solid blue), and  $[C_1, C_2] = [5, 15], [15, 25]$  and  $[25, 35]$  (top left to bottom right, dashed red). Right: Coherence bands centered on  $\bar{C} = 20\%$  (solid blue curve) with widths  $10, 20, 30$  and  $40\%$  (bottom left to top right, dashed red). Approximation of Eq. (30) shown in green. The slope  $b_1$  is fixed at  $0.06$  throughout.

Coherences are mixed during blocks of trials in the experiment of interest, so we now consider a continuum idealization in which coherences are selected from a uniform distribution over  $[C_1, C_2]$  (again positive for T1 and negative for T2). Instead of summing the weighted probabilities of correct 1 and 2 choices for  $\pm \bar{C}$ , we must now average over the entire range of coherences:

$$\mathbb{E}[r] = \frac{1}{(C_2 - C_1)} \int_{C_1}^{C_2} \left[ \frac{r_1}{2} P(+C; b_1, b_2) + \frac{r_2}{2} [1 - P(-C; b_1, b_2)] \right] dC. \quad (26)$$

Computing the derivative via the Leibniz integral rule, noting that the limits of integration do not depend on  $b_2$ , and again using Eq. (22) we find that

$$\frac{\partial \mathbb{E}[r]}{\partial b_2} = \frac{1}{2(C_2 - C_1)} \int_{C_1}^{C_2} \left[ r_1 \frac{\partial P}{\partial b_2}(+C; b_1, b_2) - r_2 \frac{\partial P}{\partial b_2}(-C; b_1, b_2) \right] dC = 0,$$

which implies that

$$\frac{r_1}{r_2} = \frac{\int_{C_1}^{C_2} \exp(-b_1^2(b_2 - \bar{C})^2) dC}{\int_{C_1}^{C_2} \exp(-b_1^2(b_2 + \bar{C})^2) dC}, \quad (27)$$

where we have cancelled common terms in the integrands that do not depend upon  $C$ . To turn these expressions into standard error function integrals we change variables by setting  $y = b_1(b_2 \pm C)$  and  $dy = \pm b_1 dC$ . Integrating Eq. (27) and cancelling further common terms yields the optimality condition:

$$\frac{r_1}{r_2} = - \left\{ \frac{\operatorname{erf}[b_1(b_2^{\text{opt}} - C_2)] - \operatorname{erf}[b_1(b_2^{\text{opt}} - C_1)]}{\operatorname{erf}[b_1(b_2^{\text{opt}} + C_2)] - \operatorname{erf}[b_1(b_2^{\text{opt}} + C_1)]} \right\}. \quad (28)$$

Setting  $C_2 = \bar{C} + \epsilon$ ,  $C_1 = \bar{C} - \epsilon$ , expanding (28) in a Taylor series and letting  $\epsilon \rightarrow 0$ , we recover the single coherence level result (24).

The expression (28) cannot be inverted to solve explicitly for the optimal starting point  $b_2^{\text{opt}}$  in terms of the other parameters, but we may use it to plot the reward ratio  $r_1/r_2$  as a function of  $b_2^{\text{opt}}$  for fixed  $a$ ,  $T$ ,  $\sigma$  and coherence range  $[C_1, C_2]$ . The axes of the resulting graph can then be exchanged to produce a plot of  $b_2^{\text{opt}}$  vs.  $r_1/r_2$  for comparison with the single coherence prediction (24). The dashed red curves in Fig. 4 (left) show optimal shifts for  $[\bar{C} - 5\%, \bar{C} + 5\%]$  centered around the three fixed coherence levels (solid blue curves). Fig. 4 (right) shows optimal shifts for coherence bands of increasing width centered around  $\bar{C} = 20\%$ . Note that the coherence bands require larger biases than fixed coherences at their centers demand (left panel), and that optimal bias increases with the width of a band centered on a given coherence (right panel). Biases, and hence optimal shifts of the PMF, increase with coherence range because the reward information is more significant for coherences close to zero, where accuracy is lowest. This fact will play a subtle role in §5.2 when we compare optimal shifts predicted for the two monkeys, one of which worked with a smaller set of coherences than the other (see §2.3).

If coherences span the range from  $C_1 = 0$  to an upper limit  $C_2$  that is sufficiently large that we may approximate

$$\operatorname{erf}[b_1(b_2^{\text{opt}} - C_2)] \approx -1 \quad \text{and} \quad \operatorname{erf}[b_1(b_2^{\text{opt}} + C_2)] \approx 1, \quad (29)$$

then (28) implies that

$$\frac{r_1}{r_2} \approx \frac{1 - \operatorname{erf}(b_1 b_2^{\text{opt}})}{1 + \operatorname{erf}(b_1 b_2^{\text{opt}})} \quad \text{or} \quad \operatorname{erf}(b_1 b_2^{\text{opt}}) \approx \frac{r_1 - r_2}{r_1 + r_2}. \quad (30)$$

(Note that  $\lim_{u \rightarrow \infty} \operatorname{erf}(u) = 1$  and  $\operatorname{erf}(u) > 0.985$  for  $u \geq 1.75$ , and that the latter condition holds for the parameters estimated for both monkeys in §5.) Eq. (30) in turn implies that, instead of the relationship  $|b_2^{\text{opt}}| \sim 1/b_1^2$  of Eq. (24) in the single coherence case, for a sufficiently broad band of coherences including zero, we have  $|b_2^{\text{opt}}| \sim 1/b_1$  or  $b_1 b_2^{\text{opt}} = \text{constant}$ . The green curve in Fig. 4 (right) shows that this simple relationship can provide an excellent approximation.

## 4.2 Optimal shifts for a finite set of coherences

In the experiment analyzed in §5 a finite set of fixed nonzero coherences  $\{\pm C_j, j = 1, \dots, N\}$  is used, along with zero coherence, each of these  $2N + 1$  conditions being presented with equal probability. Moreover, zero coherence stimuli (for which there is no correct answer) are rewarded equally probably with  $r_1$  and  $r_2$ . The expected reward on each trial is therefore:

$$\mathbb{E}[r] = \frac{1}{2N + 1} \left\{ r_1 \sum_{j=1}^N P(b_1(+C_j + b_2)) + r_2 \sum_{j=1}^N [1 - P(b_1(-C_j + b_2))] + \frac{[r_1 P(b_1 b_2) + r_2 (1 - P(b_1 b_2))]}{2} \right\}. \quad (31)$$

As in §4.1 the optimal shift is determined by seeking zeros of the derivative of (31) with respect to  $b_2$ . Excluding the normalization factor  $2N + 1$ , this leads to:

$$\frac{\partial \mathbb{E}[r]}{\partial b_2} = r_1 \sum_{j=1}^N \frac{\partial P}{\partial b_2}(b_1(+C_j + b_2)) - r_2 \sum_{j=1}^N \frac{\partial P}{\partial b_2}(b_1(-C_j + b_2)) + \frac{(r_1 - r_2)}{2} \frac{\partial P}{\partial b_2}(b_1 b_2) = 0, \quad (32)$$

from which, again appealing to Eq. (22), we obtain the expression

$$\frac{r_1}{r_2} = \frac{\sum_{j=1}^N \exp(-b_1^2(b_2 - C_j)^2) - \exp(-b_1^2 b_2^2)}{\sum_{j=1}^N \exp(-b_1^2(b_2 + C_j)^2) - \exp(-b_1^2 b_2^2)}. \quad (33)$$

As for Eq. (28) we cannot solve Eq. (33) explicitly for  $b_2$  in terms of the reward ratio and  $b_1$ , but we can again plot  $r_1/r_2$  as a function of  $b_2$  for fixed  $b_1$  values, and invert the resulting graph. See Fig. 6 in §5.2.

To get an explicit idea of how the key quantities of slope  $b_1$ , shift  $b_2$  and reward ratio  $r_1/r_2$  are related at optimal performance, we recall the relationships (24) and (30) derived in §4.1 in the special cases of a single coherence and a broad range of uniformly-distributed coherences including zero. These predict, respectively, that  $|b_2^{\text{opt}}| \sim 1/b_1^2$  and  $|b_2^{\text{opt}}| \sim 1/b_1$ . For non-uniformly distributed coherences such as those used in the experiments detailed in §5 we have found that a function of the form

$$b_2^{\text{opt}} = K b_1^{-\alpha}, \quad (34)$$

with  $K$  and  $\alpha$  suitably chosen constants that depend upon the set of coherences and the reward ratio, fits the optimal shift-sensitivity relationship very well; we shall appeal to this in analyzing some of the experimental data in §5.3. In all cases, optimal shifts increase rapidly as sensitivity ( $b_1 \sim a/\sigma$ ) diminishes.

## 5 Results II: fitting the theory to monkey data

Here we perform fits of accuracy data collected for a discrete set of coherences, namely  $C = 0, \pm 1.5\%, \pm 3\%, \pm 6\%, \pm 12\%, \pm 24\%, \pm 48\%$ , under the four reward schedules described in §2. Data from the two monkeys (A and T) are analyzed separately. As described in §2.3, T was not tested with the lowest coherences  $C = \pm 1.5\%$  and  $\pm 3\%$ . While each coherence is presented with equal probability, their spacing increases with  $C$ , so that the majority of trials occur in the center of the range around  $C = 0$ , unlike the case of uniformly-distributed coherences considered in §4.1. This will play a subtle role when we compare optimal shifts for the two animals in §5.2.

### 5.1 Fits of data averaged over multiple sessions to PMFs

Drawing on the observations in §3.4, we start by estimating average values of the parameters  $b_1$  and  $b_2$  in the psychometric function in the form (15), by collectively fitting all the data for each animal: 35 blocks of trials for A and 25 for T. We first fitted  $b_1$  and  $b_2$  separately for the four reward conditions by computing the fraction of T1 choices  $F(C_j)$  for each coherence level and minimizing the residual error:

$$\text{Err} = \sum_{j=-N}^{+N} [F(C_j) - P(C_j)]^2,$$

obtaining the values in the top two rows of Table 1. Fig. 5 shows the resulting PMFs for A (left) and T (right). We then pooled the accuracy data for equal rewards, re-fitted to determine common  $b_1$  and  $b_2$  values for conditions HH and LL for each animal, and held  $b_1$  at the resulting value while re-estimating  $b_2$  for the unequal rewards data, to obtain rows 3 and 4 of the table. The bottom two rows list values of  $b_1$  and  $b_2$  obtained when  $b_2 = 0$  is imposed in separate fits of conditions LL and HH (first two columns), and the value of  $b_1$  obtained from pooled HH and LL data with  $b_2 = 0$ , along with values of  $b_2$  for unequal rewards obtained using that same  $b_1$  value (last two columns). Fit errors are substantially higher for monkey T under the  $b_2 = 0$  constraint, due to his greater shifts for LL and HH (figures in parentheses in last row).

Subject	$b_1, b_2$ for LL	$b_1, b_2$ for HH	$b_1, b_2$ for HL	$b_1, b_2$ for LH
Monkey A	0.0508, 0.890 (0.00096)	0.0509, -0.110 (0.0011)	0.0526, 15.5 (0.0017)	0.0531, -14.0 (0.0013)
Monkey T	0.0399, -4.58 (0.00087)	0.0469, -2.87 (0.00057)	0.0415, 15.6 (0.00081)	0.0460, -17.5 (0.0018)
Monkey A	0.0508, 0.390 (0.00036)	0.0508, 0.390 (0.00036)	0.0508, 15.8 (0.0020)	0.0508, -14.3 (0.0018)
Monkey T	0.0432, -3.68 (0.00023)	0.0432, -3.68 (0.00023)	0.0432, 15.4 (0.0011)	0.0432, -17.9 (0.0024)
Monkey A	0.0508, 0 (0.0013)	0.0509, 0 (0.0012)	0.0508, 15.8 (0.0013, 0.0020)	0.0508, -14.3 (0.0013, 0.0018)
Monkey T	0.0385, 0 (0.046)	0.0460, 0 (0.023)	0.0421, 15.5 (0.033, 0.00085)	0.0421, -18.0 (0.033, 0.0030)

Table 1: Parameter values for data fits for monkeys A and T, averaged over all sessions, to the PMF (15). Upper two rows show separate fits of  $b_1$  and  $b_2$  for the four reward conditions. Middle two rows show fits for pooled LL and HH data, with resulting common  $b_1$  value held fixed across unequal reward conditions. Lower two rows show results with  $b_2$  constrained to zero for equal rewards; in columns 1 and 2 LL and HH are fitted separately, in columns 3 and 4 LL and HH data is pooled to produce  $b_1$ , and this value is fixed across unequal reward conditions. Units of  $b_1$  and  $b_2$  are increase in probability of a T1 choice per per % coherence and % coherence respectively (see §3.4). Values are given to 3 significant figures with residual fit errors (squared  $L^2$  norm) in parentheses.

In the first and least-constrained fits, Monkey A’s  $b_1$  values change across the four reward conditions by a factor of only 1.05, indicating that the predominant effect of unequal rewards is a lateral shift of the PMF, with no significant change in slope. His shifts for the HL and LH conditions are significantly different from zero and from those for HH and LL (according to one- and two-sample t tests on the underlying normal distributions  $\frac{\partial P}{\partial C} = \frac{b_1}{\sqrt{\pi}} \exp[-b_1^2(C + b_2)^2]$  with parameters listed in the top row of Table 1 and  $p < 0.01$  [11, §9.2])<sup>1</sup>. At 15.5% and -14.0% the HL and LH shifts are not significantly asymmetrical (t test,  $p = 0.77$ ), and his PMFs for equal rewards are also statistically indistinguishable from each other (t test,  $p = 0.86$ ) and from an unshifted

<sup>1</sup>Matlab codes used for data analysis, computation of statistics, and producing figures are available at [www.math.princeton.edu/~sffeng](http://www.math.princeton.edu/~sffeng).

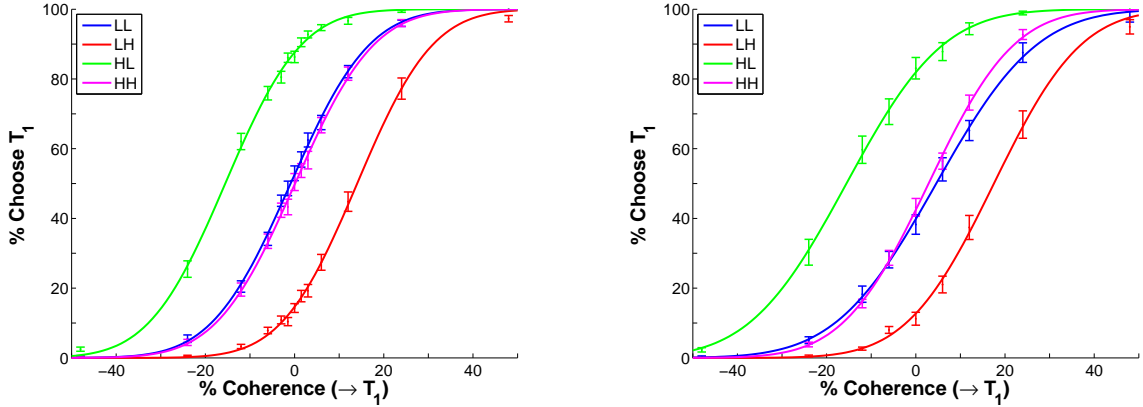


Figure 5: Fits of accuracy data from monkeys A (left) and T (right) to the PMF (15), for the four reward conditions averaged over all sessions. Bars denote standard errors. See text for details.

PMF with  $b_2 = 0$  (t tests,  $p = 0.82$ ). In contrast, Monkey T displays slopes that differ by a factor of 1.18 and shifts toward T2 of 4.58% and 2.87% respectively in the the LL and HH conditions, his slope being lower and his shift larger for LL than for HH, possibly indicating increased attention in the case of high rewards. However, his PMFs for LL and HH are also statistically indistinguishable (t test,  $p = 0.83$ ) and, in spite of the more obvious asymmetry their shifts are also not significantly different from zero (t tests,  $p = 0.44$ ). Like A's, his PMFs for the unequally rewarded conditions are significantly shifted (t tests,  $p < 0.05$ ), but again without significant asymmetry (t test,  $p = 0.85$ ).

In the optimality analysis to follow we require a common estimate of slope as a measure of the animal's sensitivity, or ability to discriminate the signal. Rows 3 and 4 of Table 1 show that shifts for the unequally rewarded conditions change by at most 0.4% when  $b_1$  is held at the common value fitted to the equal rewards data. We therefore believe that the common slope estimates  $b_1 = 0.0508$  for monkey A and  $b_1 = 0.0432$  for monkey T are suitable bases for optimality predictions. We have already noted in §2.3 that monkey T's higher psychophysical threshold led us to exclude the  $\pm 1.5\%$  and  $\pm 3\%$  coherences, and his common slope value is substantially less than that of monkey A.

Finally, we computed rows 5 and 6 of Table 1 with  $b_2$  constrained to zero in order to check that the slope parameter is not significantly affected by shifts and left/right asymmetries in the equally rewarded cases. Monkey A's slope is unchanged (to 3 significant figures) and Monkey T's distinct LL and HH slopes change by factors of only 0.96 and 0.98. Even when a common fit to LL and HH data with  $b_2 = 0$  is enforced, Monkey T's shifts for unequal rewards change by only 0.1%, and monkey A's are unchanged.

We remark that the sigmoidal or logit function

$$P_{\text{sig}}(C) = \frac{1}{1 + \exp[-b_1(C + b_2)]}, \quad (35)$$

used in the work reported in [14, 33], provides an alternative model for the PMF. Eq. (35) is somewhat simpler than the cumulative normal distribution (15), which involves the special error function, although (35) lacks a principled derivation from a choice model. We also examined fits to  $P_{\text{sig}}(C)$  and found that they were generally similar to the cumulative normal fits, but typically incurred slightly higher residual fit errors.

## 5.2 How close are the animals, on average, to optimal performance?

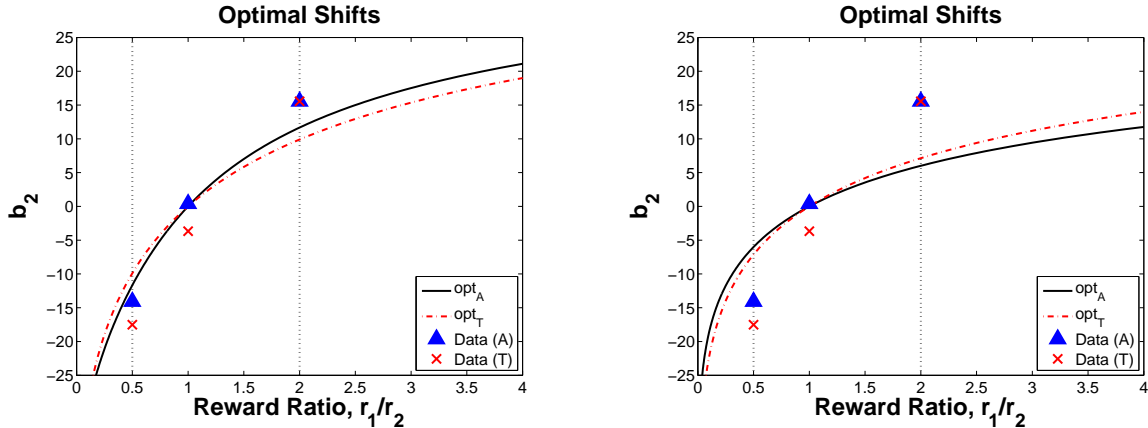


Figure 6: Optimal shifts  $b_2$  for a range of reward ratios  $r_1/r_2$  and  $b_1 = 0.0508$  (solid, black) and  $b_1 = 0.0432$  (dot-dashed, red), corresponding to slopes of PMFs fitted to equal rewards data for monkeys A and T. The vertical lines at  $r_1/r_2 = 0.5$  and  $2$  intersect the curves at the symmetrically placed optimal shifts for those reward ratios. Left panel shows predictions for the different sets of nonuniformly-distributed coherences viewed by each animal; right panel shows results for coherences distributed uniformly from  $-48\%$  to  $48\%$ : note smaller optimal shifts and reversal of order of curves for A and T in right panel. Triangles and crosses respectively indicate shifts determined from data for monkeys A and T for  $r_1/r_2 = 0.5, 1$  and  $2$  (cf. Table 1).

We took the slope values  $b_1 = 0.0508$  for A and  $b_1 = 0.0432$  for T, fitted to the pooled LL and HH equal rewards data averaged over all sessions (rows 3 and 4 of Table 1) to best represent the animals’ average sensitivities. Using these values, we then computed optimal shifts predicted by Eq. (33) for unequal reward conditions over the range  $r_1/r_2 \in [0, 4]$ , which includes the ratios  $r_1/r_2 = 2$  (HL) and  $0.5$  (LH) that were tested. We did this both for the sets of coherences viewed by A and T, and for a uniformly distributed set of coherences spanning the same range. Fig. 6 shows the resulting optimal shift curves along with the actual session-averaged shifts computed from the animals’ unequal reward data as listed in the top two rows of Table 1, and the common values for equal rewards as listed in rows 3 and 4 (triangles and crosses). Both animals “overshift” beyond the optimal values for the LH and HL conditions, T’s overshifts being greater than A’s. The figure also clearly shows T’s appreciable shift for equal rewards, in contrast to A’s nearly optimal behavior under those conditions.

Fig. 6 (left) shows that, when based on the coherences used in the experiment, monkey T’s optimal curve predicts shifts *smaller* than those for monkey A, despite T’s lower sensitivity. For a given reward ratio and the *same* set of coherences, a smaller  $b_1$  requires *greater* shifts because, as sensitivity falls, it is better to place increasing weight on the alternative that gains higher rewards, as shown in Fig. 6 (right). However, since monkey A views four low coherence stimuli ( $\pm 1.5\%$  and  $\pm 3\%$ : cf. §2.3) that T does not, his optimal shifts are additionally raised as noted in §4.1 above, outweighing his higher sensitivity. We also observe that the overall magnitudes of the optimal shifts predicted for uniformly distributed coherences are substantially smaller, being  $6.14\%$  and  $7.16\%$  for A and T respectively, in comparison with  $11.7\%$  and  $9.92\%$  for the coherences used in the experiments.

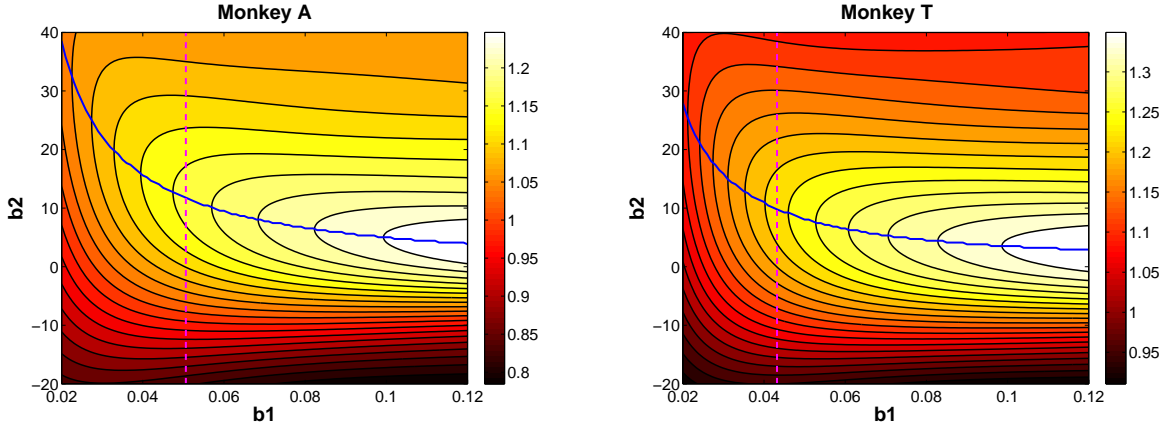


Figure 7: Contours (black curves) of expected rewards  $\mathbb{E}[r]$  for  $r_1/r_2 = 2$  for monkeys A (left) and T (right) over the  $(b_1, b_2)$ -plane, based on the coherences viewed by each animal. Vertical dashed lines indicate  $b_1$  values fitted to pooled equal rewards data. Note that gradients in  $b_2$  in either direction away from ridges of maximum expected rewards (blue curves) become smaller as  $b_1$  decreases, that gradients are smaller for overshifts in  $b_2$  than for undershifts, that this asymmetry increases as  $b_1$  decreases, and that gradients are steeper for T than for A. See text for discussion.

While the overshifts for conditions HL and LH are significant in terms of coherence, it is important to assess how dearly they cost the animals in reduced rewards. In Fig. 7 we plot expected reward functions (31) for  $r_1/r_2 = 2$  and the sets of coherences experienced by each animal (expected rewards for  $r_1/r_2 = 1/2$  are obtained by reflecting about  $b_2 = 0$ ). This reveals that, given the animals' averaged  $b_1$  values (dashed magenta lines), the second derivatives  $d^2\mathbb{E}[r]/db_2^2$  at the maxima are small, so the peaks are mild and deviations of  $\pm 10\%$  coherence from  $b_2^{\text{opt}}$  lead to reductions in expected rewards by only 2–3% from the maximum values (blue curves): an observation to which shall return below. Moreover, for unequal rewards the expected values decrease from their maxima more rapidly as  $b_2$  falls below  $b_2^{\text{opt}}$  than they do for  $b_2$  above  $b_2^{\text{opt}}$ . (The asymmetry becomes stronger as the reward ratio increases, and the curves are even functions when  $r_1 = r_2$  (not shown here).) This provides a rationale for the overshifting exhibited by the monkeys: smaller losses are incurred than in undershifting by the same amount. A similar observation appears in [4, pp 728-729], in connection with the dependence of reward rate on decision threshold in a free response task.

We conclude that, when averaged over all sessions, both animals' shifts err in the direction that is least damaging, and that neither suffers much penalty due to his overshift. Fig. 8 further quantifies this by plotting the optimal PMF curves based on the slope values  $b_1$  for pooled equal rewards ( $b_2^{\text{opt}} = 0$ ), and with the symmetric optimal shifts  $\pm b_2^{\text{opt}} \neq 0$  for the HL and LH reward conditions predicted by Eq. (33), along with bands that contain over- and under-shifted PMFs that garner 99.5% of the maximum rewards. With two exceptions ( $C = \pm 48\%$ ), monkey A's mean shifts for all conditions lie within or on the borders of these bands. Monkey T is less accurate, exhibiting substantial shifts for the HH and LL conditions and significantly overshifting for unequal rewards (especially LH); even so, his rewards lie within 99% bands with the exception of that for the LH condition, which lies within the 98% band (not shown here, but see Fig 9 below).

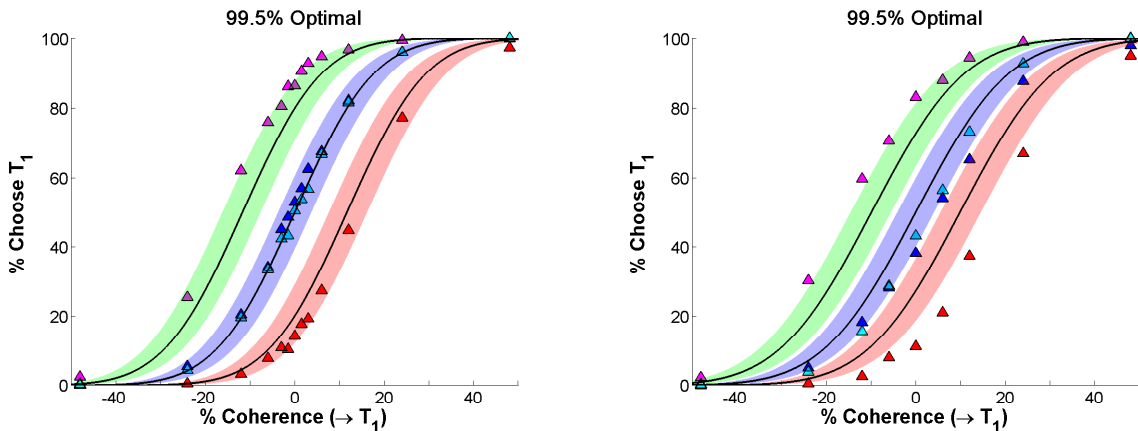


Figure 8: Optimal PMFs (black) and bands (color) in which 99.5% of maximal possible rewards are gained, compared with session-averaged HL, LL and HH, and LH data (triangles, left to right on each panel) for monkeys A (left panel) and T (right panel). See text for details.

### 5.3 Variability of behaviors in individual sessions

As Figs. 5 and 8 illustrate, when averaged over all sessions, monkeys A and T respectively come within 0.5% (except for two outlying points) and 2% of achieving maximum possible rewards, given their limited sensitivities. However, the standard errors in Fig. 5 show that their performances are quite variable. Indeed, the mean slopes  $b_1 = 0.0569$  for A and  $b_1 = 0.0491$  for T, obtained by averaging values fitted separately for each session, have standard deviations of 0.0116 and 0.0076 respectively ( $\approx 20\%$  and  $15\%$  of their means).<sup>2</sup>

Since both sensitivity, quantified by  $b_1$ , and shift ( $b_2$ ) vary substantially from session to session, we asked if these parameters exhibit any significant correlations that would indicate that the animals are tracking the ridges of maxima on Fig. 7. Specifically, from Eq. (33) we can compute values of  $b_2$  for which  $\mathbb{E}[r]$  is maximized for given  $b_1$  for reward ratios  $r_1/r_2 = 2$  (HL) and  $r_1/r_2 = 0.5$  (LH), yielding loci of optimal shifts as a function of sensitivity, and from Eq. (31) we can deduce similar loci on which fixed percentages of maximum expected rewards are realised. In Fig. 9 we compare the results of individual experimental sessions, plotted as points in the  $(b_1, b_2)$ -plane, with these curves. The asterisks indicate the mean values of  $b_1$  and  $b_2$  for each combination of animal and reward condition; the points indicate outcomes for individual sessions.

While in some cases the data seems to “parallel” the optimal performance contours (e.g., for both monkeys in condition LH and for A in conditions LL and HH), computations of Pearson’s product moment correlation ( $r$ ) between  $b_1$  and  $b_2$  reveal weak correlations that approach or exceed 0.5 only if the unequally rewarded (HL and LH) data for each animal are pooled ( $r = 0.542$ , with a 95% confidence interval  $[0.351, 0.689]$  for A;  $r = 0.458$  and  $[0.205, 0.653]$  for T). Moreover, as noted by J. Gao and J. McClelland (personal communications), these parameters are not orthogonal. In the PMF of Eq. (15),  $b_1$  accounts for how coherence scales but it is the *product*  $b_1 b_2$  that describes the effect of unequal rewards: thus, a correlation between  $b_1$  and  $b_2$  is to be expected.

The optimality theory of §4.2 allows us to perform a more telling test. While we cannot extract

<sup>2</sup>These means differ from the averages of the four  $b_1$  values in rows 1 and 2 of Table 1 because they were obtained by averaging the results of individual session fits, rather than from fits of data that was first averaged over sessions.

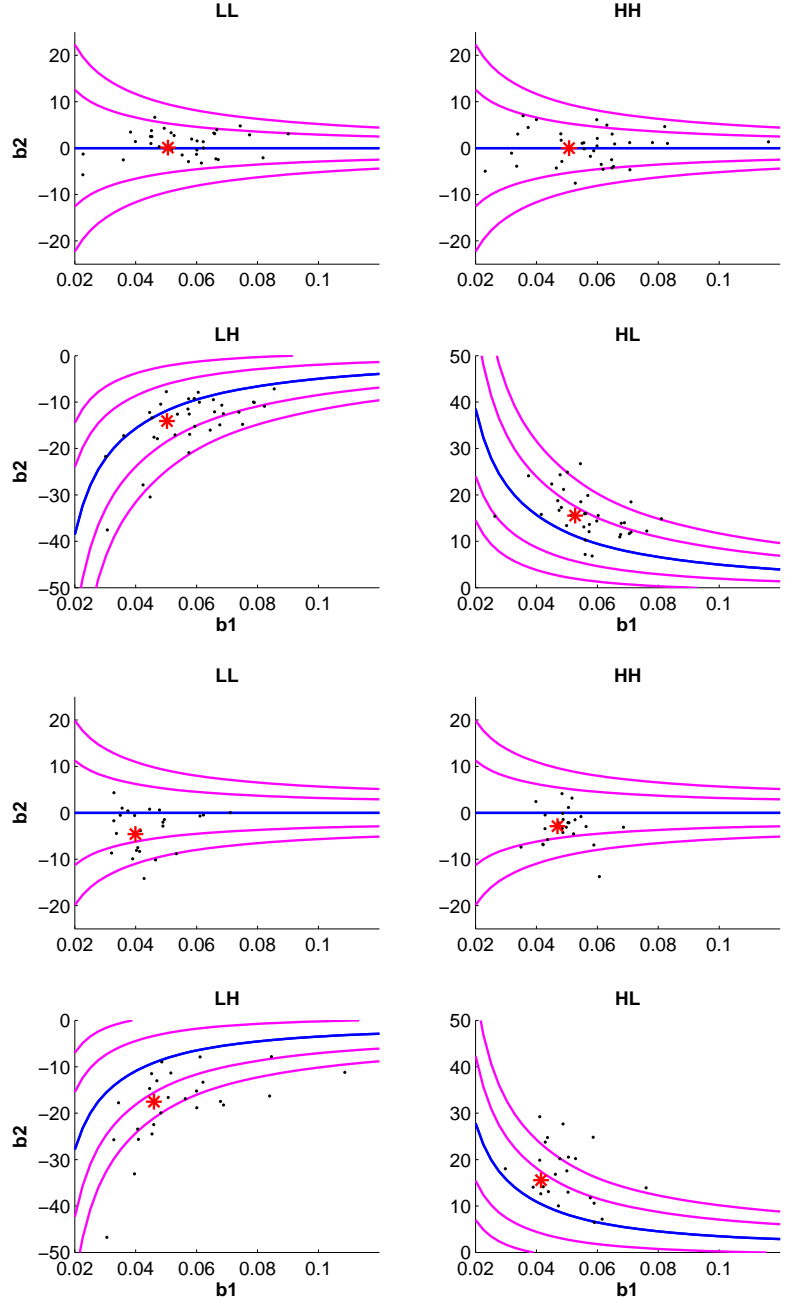


Figure 9: Slope and shift values for individual sessions and the four reward conditions, plotted as points in the  $(b_1, b_2)$ -plane for monkeys A (top four panels) and T (lower four panels). Asterisks indicate values averaged over all sessions (cf. top two rows of Table 1). Performance curves and bands show optimal  $b_2$  values for given  $b_1$  values (central blue curves) and values that gain 99% and 97% of maximum rewards are also shown (flanking magenta curves closest to and farthest from blue curves, respectively).

an exact formula for the optimal covariation of  $b_1$  and  $b_2$  implicit in Eq. (33), Eq. (34) provides an excellent approximation for the blue curves of Fig. 9, implying that individual session data should lie close to  $b_2^{\text{opt}} b_1^\alpha \approx \text{constant}$  if the animals are tracking the ridges. Fitting values of  $\alpha$  for A and T ( $\alpha = 1.26$  and  $1.30$  respectively) and comparing the HL and LH data sets with these curves gives considerably weaker correlations than those for  $b_1$  and  $b_2$  quoted above. We therefore conclude that no significantly-correlated adjustments of  $b_1$  and  $b_2$  exist, and that random scatter dominates the individual session data.

## 6 Summary and discussion

We reduce a leaky competing accumulator model to an Ornstein-Uhlenbeck (OU) process, and therefrom derive a cumulative normal psychometric function (PMF) that describes how accuracy depends upon coherence (signal-to-noise ratio) in a two-alternative forced-choice task with cued responses. The key parameters in the PMF are its *slope* at 50% accuracy, which quantifies a subject’s sensitivity to the stimulus, and its *shift*: the coherence at which 50% accuracy is realised. We compute analytical expressions describing optimal shifts that maximize expected rewards for given slopes and reward ratios. We find that this PMF can fit behavioral data from two monkeys performing a motion discrimination task remarkably well. The resulting slopes and shifts show that, faced with mixed coherences, while both animals “overshift” for unequal rewards, they nonetheless garner 98 – 99% of their maximum possible rewards (Fig. 8), and they achieve this in spite of significant variability in sensitivity and shifts from session to session. We also propose two simple methods by which the OU process could be biased by reward expectations, in order to produce such shifts. One requires a biased starting point for evidence accumulation, the other assumes a continuing bias to the drift rate that enters the OU process prior to and throughout the stimulus viewing period. As described in §3.4, the fixed viewing time experiment employed here cannot distinguish among these or other models of biasing.

Our optimality analysis presumes that the PMF slope ( $b_1$ ) has an upper bound that reflects fundamental limits on sensitivity to the visual stimulus. We then seek the unique shift ( $b_2^{\text{opt}}$ ) that maximizes expected rewards over the given coherence and reward conditions, for a fixed slope. This makes for a well-posed mathematical analysis, but it does not imply that the animal is faced with a given sensitivity and then “chooses” a shift (cf. session-to-session variability, §5.3). He might equally well choose a shift and then “accept” a sensitivity that delivers adequate rewards, perhaps by implicitly selecting a weight for the top-down reward information, and then relaxing attention to the stimuli until his reward rate reaches a predetermined level. He may even co-vary these parameters to achieve the same end. This is reminiscent of a robust-satisficing strategy that has been studied in connection with setting speed-accuracy tradeoffs [50].

A related study of optimal decision strategies in two-alternative forced-choice tasks with free responses has shown that decision thresholds can be determined for a pure drift diffusion process that optimize reward rate by setting a speed-accuracy tradeoff [4]. In that work it is necessary to assume that trials are blocked (e.g. with equal coherences  $\pm\bar{C}$ ), so that conditions remain statistically stationary during each session and one can appeal to optimality of the DD process [47]. In contrast, for cued responses only the accuracy level need be maximized, one need not assume a pure DD process, and optimization can be done in the face of mixed coherences and mixed reward contingencies. As the analyses of §§3.3-4.2 show, reduction to a one-dimensional process permits explicit calculations of PMFs and optimality conditions, and comparison with data

requires only simple two parameter fits. However, the present behavioral data lacks the reaction time distributions that allow fits that could distinguish among multiparamater variants of DD and OU models [39, 45, 38, 37].

We have taken as a utility function  $\mathbb{E}[r]$  the (normalised) value of expected rewards, implicitly assuming that two drops of juice are worth twice one drop. Subjective utility may not vary linearly with reward size: for example, at high reward ratios it may rise more slowly and saturate due to satiety. In contrast, if we suppose that two drops of juice are worth 2.5 or 3 times as much as one drop, then the shifts of both animals would lie much closer to the optimal curves of Fig. 6 (translate the HL data points horizontally from  $r_1/r_2 = 2$  to 2.5 or 3, and the LH data points from  $r_1/r_2 = 0.5$  to 0.4 or 0.33). However, a study of subjective value quantification would require investigation of a broad range of reward ratios.

The behavioral data analyzed here were obtained simultaneously with electrophysiological recordings from single neurons in the lateral intraparietal area (LIP) of the cerebral cortex, a region that is thought to play a key role in the formation of oculomotor decisions within the central nervous system [43, 40, 21]. The results presented in this paper raise important questions for our ongoing analysis of the neurophysiological data. Do decision-related neurons in LIP encode or at least reflect effects of both the reward prior and the coherence of the visual stimuli? Are the two effects present in the same proportions at the neural level as at the behavioral level (as quantified in the present paper)? Is the effect of reward bias evident as an offset in the accumulation of motion information by LIP neurons, or as a gain factor on the accumulation process, or both? These questions will be addressed in a future publication integrating neurophysiological data with the behavioral results.

**Acknowledgements** This material is based on research sponsored by the Air Force Research Laboratory, under agreement number FA9550-07-1-0537. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the Air Force Research Laboratory or the U.S. Government. We thank Juan Gao, Jay McClelland and Jonathan Cohen for insightful comments.

## References

- [1] L. Arnold. *Stochastic Differential Equations*. Wiley, New York, 1974.
- [2] L. Arnold. *Random Dynamical Systems*. Springer, Heidelberg, 1998.
- [3] F. Balci, D. Freestone, and C.R. Gallistel. Optimal risk assessment in man and mouse. Preprint, Dept. of Psychology, Rutgers University, New Brunswick, NJ, 2008.
- [4] R. Bogacz, E. Brown, J. Moehlis, P. Holmes, and J.D. Cohen. The physics of optimal decision making: A formal analysis of models of performance in two alternative forced choice tasks. *Psych. Rev.*, 113 (4):700–765, 2006.
- [5] P. Boxler. How to construct stochastic center manifolds on the level of vector fields. In L. Arnold, H. Crauel, and J.-P. Eckmann, editors, *Lyapunov Exponents*, pages 141–158. Springer, Heidelberg, 1991. Lecture Notes in Mathematics 1486.

- [6] K.H. Britten, M.N. Shadlen, W.T. Newsome, and J.A. Movshon. The analysis of visual motion: A comparison of neuronal and psychophysical performance. *J. Neurosci.*, 12(12):4745–4765, 1992.
- [7] K.H. Britten, M.N. Shadlen, W.T. Newsome, and J.A. Movshon. Responses of neurons in macaque MT to stochastic motion signals. *Visual Neurosci.*, 10:1157–1169, 1993.
- [8] E. Brown, J. Gao, P. Holmes, R. Bogacz, M. Gilzenrat, and J. Cohen. Simple networks that optimize decisions. *Int. J. Bifurcation and Chaos*, 15 (3):803–826, 2005.
- [9] E. Brown and P. Holmes. Modelling a simple choice task: Stochastic dynamics of mutually inhibitory neural groups. *Stochastics and Dynamics*, 1 (2):159–191, 2001.
- [10] G.C. DeAngelis and W.T. Newsome. Perceptual read-out of conjoined direction and disparity maps in extrastriate area MT. *PLOS Biology*, 2:394–404, 2004.
- [11] J.L. Devore. *Probability and Statistics*. Brooks/Cole, Belmont, CA, 2004. 6th edition.
- [12] P. Eckhoff, P. Holmes, C. Law, P.M. Connolly, and J.I. Gold. On diffusion processes with variable drift rates as models for decision making during learning. *New J. of Physics*, 10: doi:10.1088/1367-2630/10/1/015006, 2008.
- [13] E.V. Evarts. A technique for recording activity of subcortical neurons in moving animals. *Electroencephalog. and Clinical Neurophysiol.*, 24:83–86, 1968.
- [14] J. Gao, R.K. Tortell, and J.L. McClelland. Experimental investigation of the dynamic integration of reward and stimulus information: Theory and data. Abstract submitted for poster presentation at the Meeting of the Society for Neuroscience, Washington, DC, November, 2008.
- [15] C.W. Gardiner. *Handbook of Stochastic Methods, Second Edition*. Springer, New York, 1985.
- [16] J.I. Gold and M.N. Shadlen. Representation of a perceptual decision in developing oculomotor commands. *Nature*, 404:390–394, 2000.
- [17] J.I. Gold and M.N. Shadlen. Neural computations that underlie decisions about sensory stimuli. *Trends in Cognitive Science*, 5 (1):10–16, 2001.
- [18] J.I. Gold and M.N. Shadlen. The influence of behavioral context on the representation of a perceptual decision in developing oculomotor commands. *J. Neurosci.*, 23(2):632–651, 2003.
- [19] D.M. Green and J.A. Swets. *Signal Detection Theory and Psychophysics*. Wiley, New York, 1966.
- [20] J. Guckenheimer and P.J. Holmes. *Nonlinear Oscillations, Dynamical Systems and Bifurcations of Vector Fields*. Springer-Verlag, New York, 1983.
- [21] T.D. Hanks, J. Ditterich, and M.N. Shadlen. Microstimulation of macaque area LIP affects decision-making in a motion discrimination task. *Nat. Neurosci.*, 9:682–689, 2006.
- [22] A.V. Hays, B.J. Richmond, and L.M. Optican. A UNIX-based multiple process system for real-time data acquisition and control. In *WESCON Conference Proceedings*, volume 2, pages 1–10. Electron Conventions, El Segundo, CA, 1982.

- [23] G.D. Horwitz and W.T. Newsome. Separate signals for target selection and movement specification in the superior colliculus. *Science*, 284 (5417):1158–1161, 1999.
- [24] G.D. Horwitz and W.T. Newsome. Target selection for saccadic eye movements: prelude activity in the superior colliculus during a direction-discrimination task. *J. Neurophysiol.*, 86 (5):2543–2558, 2001.
- [25] S.J. Judge, B.J. Richmond, and F.C. Chu. Implantation of magnetic search coils for measurement of eye position: an improved method. *Vision Res.*, 20:535–538, 1980.
- [26] J.N. Kim and M.N. Shadlen. Neural correlates of a decision in the dorsolateral prefrontal cortex. *Nat. Neurosci.*, 2 (2):176–185, 1999.
- [27] E. Knobloch and K.A. Weisenfeld. Bifurcations in fluctuating systems: The center manifold approach. *J. Stat. Phys.*, 33 (3):611–637, 1983.
- [28] D.R.J. Laming. *Information Theory of Choice-Reaction Times*. Academic Press, New York, 1968.
- [29] S.W. Link. *The Wave Theory of Difference and Similarity*. Erlbaum, Hillsdale, NJ, 1992.
- [30] R.D. Luce. *Response Times: Their Role in Inferring Elementary Mental Organization*. Oxford University Press, New York, 1986.
- [31] M.E. Mazurek, J.D. Roitman, J. Ditterich, and M.N. Shadlen. A role for neural integrators in perceptual decision making. *Cerebral Cortex*, 13(11):891–898, 2003.
- [32] J.L. McClelland. On the time relations of mental processes: An examination of systems of processes in cascade. *Psych. Rev.*, 86:287–330, 1979.
- [33] J.L. McClelland, J. Gao, and Tortell R.K. Integrating reward and stimulus information in time-limited decisions. Abstract submitted for oral presentation at the Psychonomics Society Meeting, Chicago, Ill, November, 2008.
- [34] W.T. Newsome and E.B. Pare. A selective impairment of motion perception following lesions of the middle temporal visual area (MT). *J. Neurosci.*, 8:2201–2211, 1988.
- [35] M.J. Nichols and W.T. Newsome. Middle temporal visual area microstimulation influences veridical judgments of motion direction. *J. Neurosci.*, 22:9530–9540, 2002.
- [36] R. Ratcliff. A theory of memory retrieval. *Psych. Rev.*, 85:59–108, 1978.
- [37] R. Ratcliff and G. McKoon. The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation*, 20:873–922, 2008.
- [38] R. Ratcliff and P.L. Smith. A comparison of sequential sampling models for two-choice reaction time. *Psych. Rev.*, 111:333–367, 2004.
- [39] R. Ratcliff, T. Van Zandt, and G. McKoon. Connectionist and diffusion models of reaction time. *Psych. Rev.*, 106 (2):261–300, 1999.

- [40] J. Roitman and M. Shadlen. Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. *J. Neurosci.*, 22:9475–9489, 2002.
- [41] J.D. Schall. Neural basis of deciding, choosing and acting. *Nature Reviews Neuroscience*, 2:33–42, 2001.
- [42] M.N. Shadlen and W.T. Newsome. Motion perception: seeing and deciding. *Proc. National Acad. Sci. USA*, 93:628–633, 1996.
- [43] M.N. Shadlen and W.T. Newsome. Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. *J. Neurophysiol.*, 86:1916–1936, 2001.
- [44] P.L. Smith and R. Ratcliff. Psychology and neurobiology of simple decisions. *Trends in Neurosci.*, 27 (3):161–168, 2004.
- [45] M. Usher and J.L. McClelland. On the time course of perceptual choice: The leaky competing accumulator model. *Psych. Rev.*, 108:550–592, 2001.
- [46] A. Wald. *Sequential Analysis*. Wiley, New York, 1947.
- [47] A. Wald and J. Wolfowitz. Optimum character of the sequential probability ratio test. *Ann. Math. Statist.*, 19:326–339, 1948.
- [48] X-J. Wang. Probabilistic decision making by slow reverberation in cortical circuits. *Neuron*, 36:955–968, 2002.
- [49] K.F. Wong and X.J. Wang. A recurrent network mechanism of time integration in perceptual decisions. *J. Neurosci.*, 26 (4):1314–1328, 2006.
- [50] M. Zacksenhouse, P. Holmes, and R. Bogacz. Robust versus optimal strategies for determining the speed-accuracy tradeoff on two-alternative forced choice tasks. Preprint, Faculty of Mechanical Engineering, Technion – Israel Institute of Technology, Haifa 32000, Israel, 2008.