



# Optimality and Some of Its Discontents: Successes and Shortcomings of Existing Models for Binary Decisions

Philip Holmes,<sup>a,b,c</sup> Jonathan D. Cohen<sup>c,d</sup>

<sup>a</sup>*Department of Mechanical and Aerospace Engineering, Princeton University*

<sup>b</sup>*Program in Applied and Computational Mathematics, Princeton University*

<sup>c</sup>*Princeton Neuroscience Institute, Princeton University*

<sup>d</sup>*Department of Psychology, Princeton University*

Received 15 October 2012; received in revised form 25 April 2013; accepted 14 August 2013

---

## Abstract

We review how leaky competing accumulators (LCAs) can be used to model decision making in two-alternative, forced-choice tasks, and we show how they reduce to drift diffusion (DD) processes in special cases. As continuum limits of the sequential probability ratio test, DD processes are optimal in producing decisions of specified accuracy in the shortest possible time. Furthermore, the DD model can be used to derive a speed–accuracy trade-off that optimizes reward rate for a restricted class of two alternative forced-choice decision tasks. We review findings that compare human performance with this benchmark, and we reveal both approximations to and deviations from optimality. We then discuss three potential sources of deviations from optimality at the psychological level—avoidance of errors, poor time estimation, and minimization of the cost of control—and review recent theoretical and empirical findings that address these possibilities. We also discuss the role of cognitive control in changing environments and in modulating exploitation and exploration. Finally, we consider physiological factors in which nonlinear dynamics may also contribute to deviations from optimality.

*Keywords:* Accumulator; Cognitive control; Costs; Decision making; Drift-diffusion process; Exploitation; Exploration; Optimality; Robustness; Speed–accuracy trade-off

---

## 1. Introduction

In this article, we review mathematical models of simple decision-making tasks and use these to examine the extent to which human performance approaches but also deviates from optimality in such tasks. For the sake of analytic tractability, we focus on

two-alternative forced-choice (2AFC) tasks and consider linear models that are reduced forms of more biologically realistic models. Despite these restrictions, application of the models to empirical data suggests that human performance often approaches optimality but also reveals systematic ways in which it falls short, and it suggests hypotheses about why this is so. These include biased expectations, constraints on parameter estimation, and other factors—in particular the costs of cognitive control required for optimization—that are often neglected in analyses of cognitive performance.

We begin by considering how nonlinear neural network models of 2AFC decision tasks can be reduced to a simple, one-dimensional linear model describing a random walk, or drift-diffusion (DD) process. Such linear models have both behavioral and biological plausibility. They have long been used to model reaction time distributions and error rates in performance on 2AFC tasks (Laming, 1968; Ratcliff, 1978; Ratcliff, Van Zandt, & McKoon, 1999; Smith & Ratcliff, 2004; Stone, 1960). Furthermore, in studies of nonhuman primates performing such tasks, direct recordings from neurons in oculomotor regions, including the lateral intraparietal area (LIP), frontal eye fields, and superior colliculus, have shown that firing rates in these areas evolve over the course of a decision like sample paths of a DD process, rising to a threshold prior to response initiation (Gold & Shadlen, 2001; Mazurek, Roitman, Ditterich, & Shadlen, 2003; Ratcliff, Cherian, & Segraves, 2003; Ratcliff, Hasegawa, Hasegawa, Smith, & Segraves, 2006; Roitman & Shadlen, 2002; Schall, 2001; Shadlen & Newsome, 2001). Here, we review analyses showing how DD-based models predict optimal strategies for maximizing rewards (Bogacz, Brown, Moehlis, Holmes, & Cohen, 2006; Gold & Shadlen, 2002) and describe data that test such predictions. We consider possible sources for deviations from optimality that the data reveal, describe recent findings that address these possibilities, and finally discuss broader issues in the cognitive control of decision processes.

## **2. Optimal performance in simple decisions**

How to best characterize human decision-making performance, even for tasks as simple as 2AFC, has been the subject of intense investigation and controversy. The two most commonly considered types of models are (a) “high-level” (low-dimensional) neural networks—often called leaky competing accumulators (LCAs)—that describe the accumulation of evidence (Usher & McClelland, 2001); and (b) even simpler linear models implementing DD processes (Ratcliff, 1978; Ratcliff et al., 1999) that can be derived from LCAs in a special case.

The DD process is a continuum limit of the sequential probability ratio test (Wald & Wolfowitz, 1948), and for statistically stationary tasks, it yields decisions of specified average accuracy in the shortest possible time (Bogacz et al., 2006; Gold & Shadlen, 2002). We use this property to derive an optimal speed–accuracy trade-off that maximizes reward rate, and then describe experiments that reveal both approximations to and failures to optimize, prompting further analyses and experiments. Our main tool is the optimal performance curve of Fig. 1, which can be used to assess performance across conditions, tasks, and even individuals, irrespective of differences in task difficulty or timing. The reader wishing to

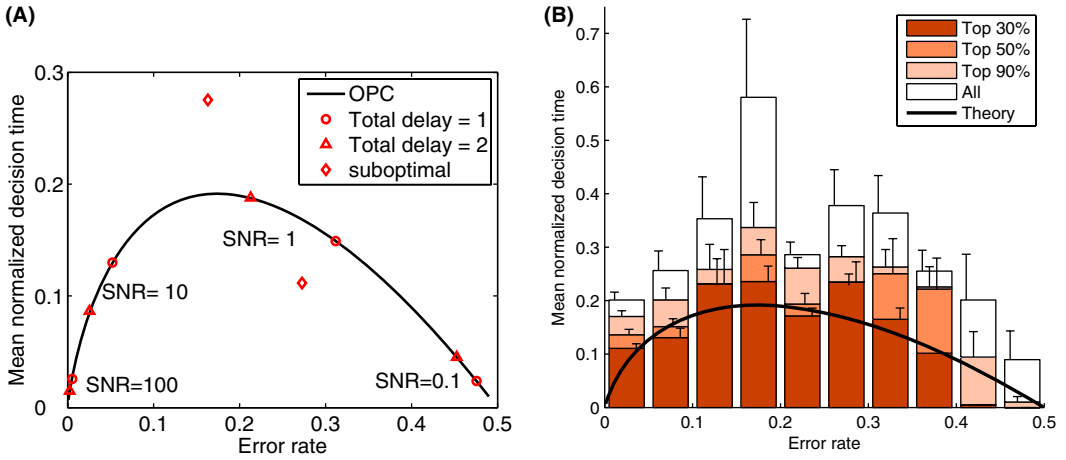


Fig. 1. (A) The optimal performance curve (OPC) of Eq. (10) relates mean normalized decision time  $\langle DT \rangle / D_{\text{tot}}$  to error-rate  $p(\text{err})$ . Triangles and circles mark hypothetical performances under eight different task conditions; diamonds mark suboptimal performances resulting from thresholds at  $\pm 1.25\theta_{\text{op}}$  for SNR = 1 and  $D_{\text{tot}} = 2$ ; both reduce RR by  $\approx 1.3\%$ . (B) OPC (curve) and data from 80 human participants (boxes) sorted according to total rewards accrued over all conditions. White: all participants; lightest: lowest 10% excluded; medium: lowest 50% excluded; darkest: lowest 70% excluded. Vertical lines show SEs. From Bogacz et al. (2006) and Zacksenhouse et al. (2010).

skip mathematical detail can find this described following Eq. (10) and refer back to Sections 2.1 and 2.2 as desired.

2.1. Leaky competing accumulators and drift-diffusion processes

In the simplest LCA model, appropriate for 2AFC tasks, two units with activity levels  $(x_1, x_2)$  represent pools of neurons selectively responsive to the stimuli (Usher & McClelland, 2001). They mutually inhibit via input–output or frequency–current ( $f-I$ ) functions that express neural activity (e.g., short-term firing rates) in terms of inputs. These include constant currents representing mean stimulus levels and i.i.d. Wiener processes modeling noise that pollute the stimuli and/or enter the local circuit from other brain regions. The dynamics are governed by the following stochastic differential equations:

$$dx_1 = [-\gamma x_1 - \beta f(x_2) + \mu_1]dt + \sigma dW_1, \tag{1}$$

$$dx_2 = [-\gamma x_2 - \beta f(x_1) + \mu_2]dt + \sigma dW_2, \tag{2}$$

where the state variables  $x_j$  denote unit activities (spike rates),  $\gamma, \beta$  are the leak and inhibition rates, and  $\mu_j, \sigma$  are the means and standard deviation of the noisy stimuli. A decision is recorded when the first activity  $x_j(t)$  exceeds a fixed threshold  $x_{j,\text{th}}$ , adjustment of which is used to tune the speed–accuracy trade-off. Similar effects may be obtained by changing baseline activity or initial conditions, mechanisms that also appear likely, and to which we

return in Section 3.2. See Grossberg (1988), Rumelhart and McClelland (1986) and Usher and McClelland (2001) for background on related connectionist networks, and Miller and Fumarola (2012) on the equivalence of different integrator models.

The function  $f(\cdot)$  characterizing neural response is typically sigmoidal:

$$f(x) = \frac{1}{1 + \exp[-g(x - b)]}, \quad (3)$$

and if the gain  $g$  and bias  $b$  are set appropriately, Eqs. (1–2) without noise ( $\sigma = 0$ ) can have one or two stable equilibria, separated by a saddle point. These correspond to two “choice attractors” in the noisy system, and if  $\gamma$  and  $\beta$  are sufficiently large, a one-dimensional, attracting curve exists that contains the three equilibria and orbits connecting them (see Brown et al., 2005; Feng, Holmes, Rorie, & Newsome, 2009; fig. 2). Hence, after rapid transients have decayed following stimulus onset, the dynamics relax to that of a nonlinear Ornstein-Uhlenbeck (OU) process (Brown et al., 2005; Roxin & Ledberg, 2008). The dominant terms are found by linearizing Eqs. (1–2) and subtracting to yield one equation for the difference  $x = x_1 - x_2$ :

$$dx = [(\mu_1 - \mu_2) + (\beta - \gamma)x]dt + \sigma dW. \quad (4)$$

In Section 3.5, we note that models of networks of spiking neurons can be reduced to nonlinear LCAs, providing a theoretical path from biophysical detail to analytically tractable linear models.

## 2.2. An optimal speed–accuracy trade-off

If leak and inhibition are balanced ( $\beta = \gamma$ ), and initial data are unbiased, appropriate when stimuli appear with equal probability and have equal reward value, Eq. (4) becomes a DD process

$$dx = A dt + \sigma dW; \quad x(0) = 0, \quad (5)$$

where  $A = \mu_1 - \mu_2$  denotes the drift rate. Responses are given when  $x$ , identified with the logarithmic likelihood ratio, first crosses a threshold  $\pm x_{\text{th}}$ ; if  $A > 0$ , then crossing of  $+x_{\text{th}}$  corresponds to a correct response and crossing  $-x_{\text{th}}$  to an incorrect one. The error rate and mean decision time, quantifying accuracy and speed, are

$$p(\text{err}) = \frac{1}{1 + \exp(2\eta\theta)} \quad \text{and} \quad \langle DT \rangle = \theta \left[ \frac{\exp(2\eta\theta) - 1}{\exp(2\eta\theta) + 1} \right]; \quad (6)$$

(see Bogacz et al., 2006, appendix; Busemeyer & Townsend, 1993). Note that the parameters  $A, \sigma$  and  $x_{\text{th}}$  reduce to two quantities  $\eta \equiv (A/\sigma)^2$  (signal-to-noise ratio [SNR], having units of inverse time), and  $\theta \equiv |x_{\text{th}}/A|$  (threshold-to-drift ratio, the first passage time for the noise-free process  $x(t) = At$ ).

If SNR and mean response-to-stimulus interval  $D_{\text{RSI}}$  remain constant across each block of trials, block durations are fixed, and rewards are insensitive to time, then optimality is achieved by maximizing *reward rate*; that is, average accuracy divided by average trial duration:

$$\text{RR} = \frac{1 - p(\text{err})}{\langle DT \rangle + T_0 + D_{\text{RSI}}}. \quad (7)$$

Here  $T_0$  is that part of the reaction time due to non-decision-related (e.g., sensory and motor) processing. As  $T_0$  and  $\eta$  also typically remain (approximately) constant for each participant, we may substitute Eqs. (6) into (7) and maximize RR for fixed  $\eta$ ,  $T_0$  and  $D_{\text{RSI}}$ , obtaining a unique threshold-to-drift ratio  $\theta = \theta_{\text{op}}$  for each pair  $(\eta, D_{\text{tot}})$ :

$$\exp(2\eta\theta_{\text{op}}) - 1 = 2\eta(D_{\text{tot}} - \theta_{\text{op}}), \quad \text{where } D_{\text{tot}} = T_0 + D_{\text{RSI}}. \quad (8)$$

Inverting the relationships (6) to obtain

$$\theta = \frac{\langle DT \rangle}{1 - 2p(\text{err})} \quad \text{and} \quad \eta = \frac{1 - 2p(\text{err})}{2\langle DT \rangle} \log \left[ \frac{1 - p(\text{err})}{p(\text{err})} \right], \quad (9)$$

the parameters  $\theta_{\text{op}}, \eta$  in Eq. (8) can be replaced by the performance measures,  $p(\text{err})$  and  $\langle DT \rangle$ , yielding a unique, parameter-free relationship describing the speed–accuracy trade-off that maximizes RR:

$$\frac{\langle DT \rangle}{D_{\text{tot}}} = \left[ \frac{1}{p(\text{err}) \log \left[ \frac{1 - p(\text{err})}{p(\text{err})} \right]} + \frac{1}{1 - 2p(\text{err})} \right]^{-1}. \quad (10)$$

Eq. (10) defines an *optimal performance curve* (OPC) (Bogacz et al., 2006), shown in Fig. 1(A). Each point  $(p(\text{err}), \langle DT \rangle / D_{\text{tot}})$  along the OPC corresponds to an optimal threshold-to-drift ratio  $\theta_{\text{op}}$  specified by Eq. (8) which, in turn, is associated with a particular decision time and error rate combination that maximizes RR: Any other threshold, lower or higher, associated with responses that are faster or slower (diamonds in Fig. 1(A)), yields smaller net rewards. Different points on the curve represent  $\theta_{\text{op}}$ s and corresponding speed–accuracy trade-offs for different values of  $\eta$  (i.e., task difficulty) and  $D_{\text{tot}}$  (i.e., task timing). Critically, the shape of the OPC *itself* is parameter free. Thus, it can be used to assess performance with respect to optimality and compare this across conditions, tasks, and even individuals, irrespective of differences in task difficulty or timing.

The OPC's shape may be intuitively understood by observing that very noisy stimuli ( $\eta \approx 0$ ) contain little information, so that, if they are equally likely, it is optimal to choose at random, giving  $p(\text{err}) = 0.5$  and  $\langle DT \rangle = 0$  (cf. SNR = 0.1 at the right of Fig. 1(A)). At its left, as  $\eta \rightarrow \infty$ , stimuli become so easy to discriminate that both  $\langle DT \rangle$  and  $p(\text{err})$  approach zero (cf. SNR = 100). For intermediate SNRs, it is advantageous to accumulate evidence for just long enough (e.g., SNRs = 1 and 10). Thus, to maximize rewards, optimal perfor-

mance is achieved by balancing speed versus accuracy, and not by systematically favoring accuracy. Human data appear to present a challenge to this model, as we now consider.

### 2.3. *Experimental evidence: Failures to optimize*

Two 2AFC experiments (Bogacz et al., 2006; Bogacz, Hu, Cohen, & Holmes, 2010) have directly tested whether humans optimize reward rate in accord with the OPC derived above. In the first, 20 participants viewed motion stimuli (Britten, Shadlen, Newsome, & Movshon, 1993) and were rewarded for each correct discrimination. The experiment was divided into 7-min blocks with different response to stimulus intervals in each block. In the second, 60 participants discriminated if the majority of 100 locations on a static display were filled with stars or empty. Blocks lasted for 4 min, and two difficulty conditions were used. In all cases, participants were instructed to maximize their total earnings, and unrewarded practice blocks were administered before testing began.

As noted above, one of the appeals of the OPC is that it is independent of the parameters defining the DD process, with  $D_{\text{tot}}$  entering only as the denominator in Eq. (10). This allows data to be pooled across all participants and task conditions. That is, findings concerning actual performance can be compared with optimal performance irrespective of task difficulty, timing, or individual differences. For the experiments of (Bogacz et al., 2006, 2010),  $\langle DT \rangle$ 's were estimated by fitting the DD model to reaction time distributions, the 0–50% error rate range was divided into 10 bins, and  $\langle DT/D_{\text{tot}} \rangle$  were computed for each bin by averaging over those results and conditions with error rates in that bin. This yields the open (tallest) bars in Fig. 1(B); the shaded bars derive from similar analyses restricted to subgroups of participants ranked by their total rewards accrued over all blocks of trials and conditions.

The top 30% group clearly performs close to the OPC, supporting the DD model as an account of human decision-making performance, and the conjecture that (at least some) participants can achieve near-optimal performance. Nonetheless, a majority achieve substantially lower total scores. Interestingly, this is due to longer decision times and greater accuracy than are required to maximize reward rate (Bogacz et al., 2010). This particular pattern of deviations from optimality raises two possibilities: (a) Participants seek to optimize some other criterion of performance, such as accuracy, in place of, or in addition to, maximizing monetary reward. (b) Participants seek to maximize reward but fall short in a systematic way due to specific constraint(s) on performance and/or the influence of other cognitive factor(s). We address these possibilities in the following sections.

### 2.4. *A preference for accuracy?*

There is a long-standing literature that suggests that humans favor accuracy over speed in reaction time tasks (e.g., Myung & Busemeyer, 1989). This could explain the observations in Fig. 1(B), insofar as decision times that are longer than optimal are typically

associated with more accurate responses. Participants may try to maximize accuracy in addition to (or even instead of) rewards, as postulated in Bohil and Maddox (2003) and Maddox and Bohil (1998). This can be formalized by combining accuracy with reward rate in at least two ways (Bogacz et al., 2006):

$$RA = RR - \frac{qp(\text{err})}{D_{\text{tot}}}, \quad \text{or} \quad RR_m = \frac{(1 - p(\text{err})) - qp(\text{err})}{\langle DT \rangle + D_{\text{tot}}}. \quad (11)$$

The first (RA) subtracts a fraction of error rate from RR of Eq. (7); the second ( $RR_m$ ) penalizes errors by reducing previous winnings. Critically, both include a free parameter  $q \in (0,1)$  that specifies the relative weight placed on accuracy. Increasing  $q$  drives the OPC upward (Bogacz et al., 2006; fig. 13), consistent with the empirical observations shown in Fig. 1(B), suggesting that participants may indeed include accuracy in their objective function. One reason for this may be that they (incorrectly) assume that errors are explicitly penalized, at least early during their experience with the task. This possibility is addressed by further empirical findings discussed below.

However, there are at least two alternative accounts of the data that preserve the assumption that participants seek primarily to optimize rewards. One is that timing uncertainty corrupts estimates of reward rate, systematically biasing performance toward longer decision times. The other is that participants factor into their estimates of reward rate the costs associated with fine-grained adjustment of parameters (such as response threshold), the advantages of which may be small relative to those costs. We consider each of these in turn and compare them with accuracy-based models in their abilities to account for empirical data.

### 2.5. Robust decisions in the face of uncertainty?

In the analyses of Sections 2.2 and 2.4, it is assumed that participants maximize an objective function for the DD model, given the exact task parameters. However, it is unlikely that participants can perfectly infer these. For example, RR depends on inter-trial delays as well as SNR. Delays, in particular, may be hard to estimate. Information-gap theory (Ben-Haim, 2006) allows parameters to lie within a bounded uncertainty set and uses a *maximin strategy* to identify parameters that optimize a worst-case scenario.

Interval timing studies (Buhusi & Meck, 2005) indicate that time estimates are normally distributed around the true duration with a standard deviation proportional to it (Gibbon, 1977). This prompted the assumption in Zacksenhouse, Bogacz, and Holmes (2010) that the estimated delay  $D_{\text{tot}}$  lies in a set  $U_p(\alpha_p; \tilde{D}_{\text{tot}}) = \{D_{\text{tot}} > 0 : |D_{\text{tot}} - \tilde{D}_{\text{tot}}| \leq \alpha_p \tilde{D}_{\text{tot}}\}$ , of size proportional to the actual delay  $\tilde{D}_{\text{tot}}$ , with presumed level of uncertainty  $\alpha_p$  analogous to the coefficient of variation in scalar expectancy theory (Gibbon, 1977). In place of the optimal threshold of Eq. (8), the maximin strategy selects the threshold  $\theta_{\text{MM}}$  that maximizes the worst RR that can occur for  $D_{\text{tot}} \in U_p(\alpha_p; \tilde{D}_{\text{tot}})$ . For uncertainties in delay, it predicts a one-parameter family of *maximin performance curves* (MMPCs) that are scaled versions of the OPC (10):

$$\frac{\langle DT \rangle}{D_{\text{tot}}} = \gamma(1 + \alpha_p) \left[ \frac{1}{p(\text{err}) \log\left(\frac{1-p(\text{err})}{p(\text{err})}\right)} + \frac{1}{1 - 2p(\text{err})} \right]^{-1}, \quad (12)$$

where  $\gamma \equiv \tilde{D}_{\text{tot}}/D_{\text{tot}}$  (Zacksenhouse et al., 2010). Like the objective functions (11) that emphasize accuracy, these curves also predict longer mean decision times than the OPC (10). Uncertain SNRs can be treated similarly, yielding families of MMPCs that differ from both the OPC Eq. (10) and MMPC for timing uncertainty Eq. (12), rising to peaks at progressively smaller  $p(\text{err})$  as uncertainty increases. An alternative strategy, also investigated in Zacksenhouse et al. (2010), yields *robust-satisficing performance curves* (RSPCs) that provide poorer fits and thus are not discussed here.

Fig. 2 shows data fits to the parameter-free OPC, the objective functions of Eq. (11), to MMPCs for timing uncertainty and SNR, and to RSPCs for timing uncertainty. While there is little difference among fits to the top 30%, data from the middle 60% and lowest 10% subgroups exhibit patterns that distinguish among the theories. Maximum likelihood computations show that MMPCs for uncertainties in delays provide the best fits, followed by RSPCs for uncertainties in delays and RA (Zacksenhouse et al., 2010). These findings indicate that optimizing reward rate given timing uncertainty leads to slower and more accurate decisions than those predicted by the OPC, which assumes perfect knowledge of delays. Greater accuracy can be considered a *consequence* of maximizing reward rate under uncertainty, and not the *objective* of optimization.

## 2.6. Experimental evidence: Practice and timing uncertainty

To test whether deviations from the OPC can be better explained by an emphasis on accuracy or by timing uncertainty, Balci and colleagues (2011) conducted a 2AFC experiment with motion stimuli encompassing a range of discriminabilities (moving dot displays with 0%, 4%, 8%, 16%, and 32% coherences [Britten et al., 1993], kept fixed in each block) and administered interval timing tests in parallel (Buhusi & Meck, 2005). Seventeen participants completed at least 13 sessions in each condition, insuring the greatest likelihood of achieving optimal performance by providing participants with extensive training and affording the opportunity to examine practice effects (Dutilh, Vandekerckhove, Tuerlinckx, & Wagenmakers, 2009; Petrov, Van Horn, & Ratcliff, 2011). Four main results emerged, three of which we now summarize; the fourth is described in Section 3.2.

First, average performance converges toward the OPC with increasing practice. Fig. 3(A) shows the mean normalized decision times (dots) for five error bins averaged over sessions 1, 2–5, 6–9, and 10–13. Performance during the final two sets of sessions is indistinguishable from the OPC for higher coherences, but decision times remain significantly above the OPC for the two lowest coherences.

Second, the accuracy-weighted objective function  $RR_m$  of Eq. (11) is superior to the OPC in fitting decision times across the full range of error rates during the early sessions,



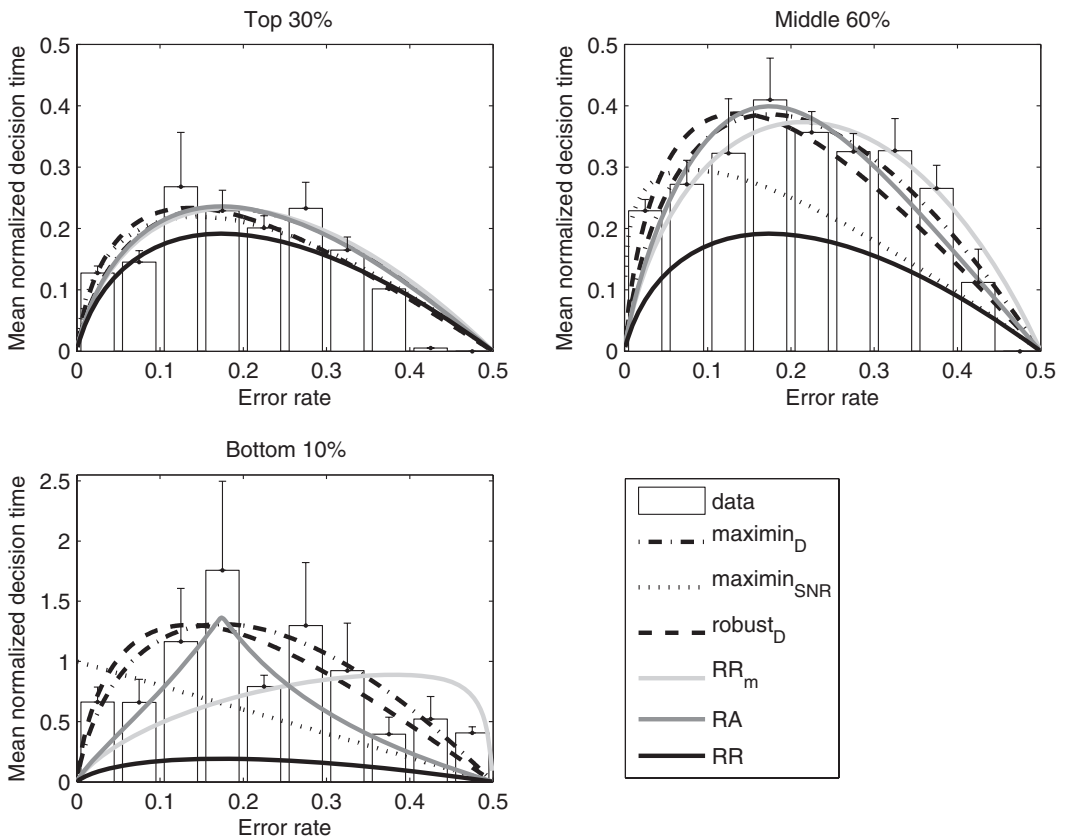


Fig. 2. Comparisons of performance curves with mean normalized decision times (with SE bars) for three groups of participants sorted by total rewards acquired. Different curves are identified by line style and gray scale in the key, in which  $\text{maximin}_D$  and  $\text{maximin}_{\text{SNR}}$  refer to maximin performance curves (MMPCs) for uncertainty in total delay Eq. (12) and noise variance respectively,  $\text{robust}_D$  to robust-satisficing performance curves (RSPCs) for uncertainty in total delay, RA and  $\text{RR}_m$  denote the accuracy-weighted objective functions of Eq. (11), and RR the OPC Eq. (10). Note different vertical axis scales in upper and lower panels. From Zacksenhouse et al. (2010).

with accuracy weight decreasing monotonically through sessions 1–9 and thereafter remaining at  $q \approx 0.2$  (data not shown here, see Balci et al., 2011; fig. 9), suggesting that participants may indeed favor accuracy early during training, but that this diminishes with practice. This is consistent with the idea that participants come to such tasks expecting that errors are associated with explicitly negative outcomes (e.g., immediate losses, as may be common in real-world settings), rather than being limited to opportunity costs (as assumed by the OPC). With practice, they may adjust their beliefs to accord with the fact that errors incur only opportunity costs.

Third, as shown in Fig. 4, deviations from the OPC are stably related to timing ability. Throughout training, the timing inaccuracy of participants, independently assessed by

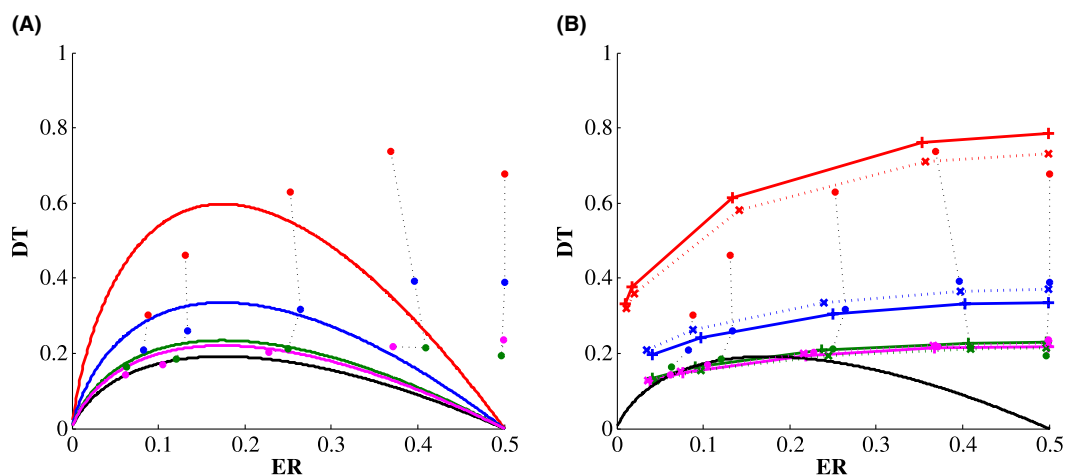


Fig. 3. Mean normalized decision times (dots) grouped by coherence vs. error proportions for sessions 1 (red), 2–5 (blue), 6–9 (green), and 10–13 (pink). (A) Performance compared with the OPC (black) and with best-fitting MMPCs for each coherence condition. With training, performance converges toward the OPC, but DTs remain relatively high at high error rates. (B) Performance compared with DD fits to single threshold for all coherences. Solid and dotted horizontal lines connect model fits and dotted vertical lines connect data points from different sessions having the same coherence. Fits connected by solid lines exclude 0 and 4% coherences; fits connected by dotted lines include all five coherences; fits from sessions 6–9 and 10–13 almost superimpose. Fits to a single threshold better capture longer DTs at high error rates relative to OPC and MMPCs. Panel (B) adapted from Balci et al. (2011).

their coefficients of variation in a peak-interval timing task (Rakitin et al., 1998), is significantly positively correlated with their distance from the OPC (Balci et al., 2011). Furthermore, this relationship provides a better account of deviations from the OPC than emphasis on accuracy (as indexed by the weight parameter  $q$  in the objective function  $RR_m$ ). This suggests that participants may choose decision parameters that maximize reward rate, taking into account their timing uncertainty, as discussed in the previous section.

Collectively, these results support the hypothesis that, given sufficient practice, humans can learn to approach optimal performance (maximizing rewards) by devaluing accuracy, with a deviation from optimality that is inversely related to timing ability. Nevertheless, MMPCs based on timing uncertainty fall short of fully capturing performance, especially for the two lowest coherences: Fig. 3(A). Thus, even after accounting for timing uncertainty, reward rates are suboptimal according to the DD model. This suggests that other factors may be involved.

One of these is persistent variability across trials. The *extended DDM* accounts for this by drawing drift rates  $A$  and starting points  $x(0)$  from distributions on each trial (Bogacz et al., 2006; Ratcliff et al., 1999), thus allowing slow and fast errors, respectively. The former may explain deviations from the OPC seen for high error rates in Fig. 3(A). However, as yet, there is no principled way to determine which factor (variability in drift vs.

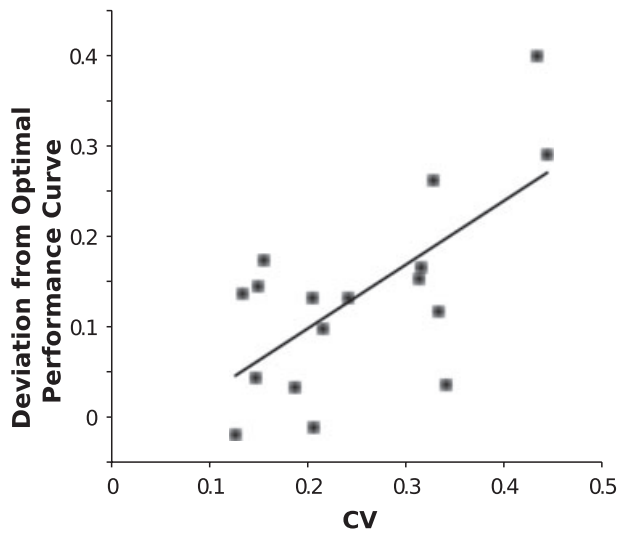


Fig. 4. Mean square deviation of decision times from optimal performance curve vs. interval timing coefficients of variation for individual participants, averaged over all coherences and sessions 2–13. Line shows linear regression fit.

starting point) should dominate. Moreover, each additional parameter brings with it a family of OPCs (as in Sections 2.4 and 2.5), and analytical expressions are not available for those derived from extended DDMs. Both Balci et al. (2011) and Simen et al. (2009) (cf. Section 3.2 below) include extended DDM fits in their optimality analyses, but we are unaware of any systematic studies of the effects of deviations from optimality due to such variability. In contrast, considering the costs of cognitive control may provide a normative account of these effects, to which we now turn.

### 3. The role of control in optimization

So far, our focus has been on optimization of performance in which task conditions are stationary: Models have assumed fixed task parameters, and data were analyzed for the most part condition-by-condition (e.g., each coherence condition was considered independently in Fig. 3). However, in both the real world and the laboratory, task conditions change and participants must adapt their decision parameters accordingly. Cognitive control has been defined as the set of processes (and underlying mechanisms) responsible for assessing task conditions and outcomes and adjusting behavior accordingly (Botvinick & Cohen, 2014). We now review work that begins to explore the role of control processes in optimization and their influence on the dynamics of performance.

### 3.1. *The cost of control*

It has long been recognized that cognitive control is associated with mental effort (Posner & Snyder, 1975). Accordingly, the effort associated with adjusting parameters as conditions change (i.e., engagement of control) may register as a cost. The fourth finding of Balci et al. (2011) supports the hypothesis that participants may take account of such costs and weigh them against benefits that would accrue from optimizing performance for each task condition. The OPCs of Fig. 3 assume that the decision threshold was optimized for the coherence (SNR) in each block. However, block-by-block adjustments would engage cognitive control that may have been perceived as effortful. Accordingly, especially if the benefits of such adjustments were small, participants may have sought a single threshold that does best over all conditions, relieving them of the need to estimate coherences and adjust thresholds from block to block.

To assess this possibility, Balci et al. (2011) computed the single threshold that was optimal when applied over all coherence conditions. Fig. 3(B) shows that the resulting curve fits the full range of performance for the later sessions (6–13), suggesting that, at least with practice, participants came to adopt a single threshold that maximized reward across all conditions. Consistent with this strategy, it was found that rewards for this single threshold differed minimally from those accrued using the individual thresholds, optimized for the different conditions. Thus, the cost–benefit trade-off appears to have weighed against exercising control on a condition-by-condition basis, and in favor of a single threshold. Hence, allowing for presumed costs of control, participants may have indeed been optimizing performance.

The analysis above highlights the need for an independent assessment of the costs of control. Recent work has begun to quantify costs associated with cognitive control (Botvinick, Huffstetler, & McGuire, 2009) and to incorporate these in formal analyses of behavioral performance (e.g., Todd, Botvinick, Schwemmer, Cohen, & Dayan, 2011). Insofar as such costs offset the rewards associated with performance, a better understanding of these, and factoring them into analyses of behavior, promises to permit more accurate assessments of the extent to which people perform optimally. That said, there is increasing evidence that, under many circumstances, people do actively adjust decision parameters in response to changes in task conditions (Gratton, Coles, Sirevaag, Eriksen, & Donchin, 1988; Henik, Bibi, Yanai, & Tzelgov, 1997; Laming, 1979; Lindsay & Jacoby, 1994; Logan & Zbrodoff, 1979; Rabbitt, 1966). Information needed to make such adjustments can come from several sources, including prior experience (i.e., expectations), the current stimulus, and assessments of the outcome of performance itself (e.g., accuracy, speed, processing conflict, confidence, etc.).

Recent studies have begun to address the mechanisms responsible for making such adjustments, with a particular focus on how they are driven by performance outcomes (Botvinick, Braver, Barch, Carter, & Cohen, 2001; Brown & Braver, 2005; Holroyd & Coles, 2002; Kiani & Shadlen, 2009; Simen, Cohen, & Holmes, 2006; Yeung, Botvinick, & Cohen, 2004). However, few of these studies have involved normative models that explicitly address the extent to which adjustments in decision parameters serve optimiza-

tion. Below, we review two areas of recent work that attempt to do this: One addresses trial-to-trial adjustments based on prior knowledge, and the other considers the dynamics of adjustment within individual trials under stimulus variability.

### 3.2. *Prior expectations and trial-to-trial adjustments*

Given prior information on the probability of observing each stimulus in a 2AFC task, a DD process can be optimized by appropriately shifting the initial condition; biased rewards favoring one response over the other can be accommodated in a similar manner (Bogacz et al., 2006). Comparisons of these predictions with human behavioral data were carried out in Simen et al. (2009). As in the experiments described above, stimulus discriminability and other conditions were fixed over each block of trials. On average, participants achieved 97%–99% of maximum reward, and some performed essentially optimally.

A related study of monkeys used a cued-response paradigm with a fixed stimulus presentation period that relieves the need for a speed–accuracy trade-off, but in which motion coherences varied randomly from trial to trial, and differences in reward contingencies for the two responses were signaled before each trial. The monkeys came within 0.5% and 2%, respectively, of maximum possible rewards by adjusting the balance between preferring the higher reward and accurately assessing the noisy stimulus (Feng et al., 2009). This implies appropriate shifts in the psychometric (accuracy vs. discriminability) function depending on coherence, which can be effected by either biasing drift rates or by shifting initial conditions. That the latter is more likely was established in Rorie, Gao, McClelland, and Newsome (2010) by fitting LIP recordings. Behavioral studies of humans discriminating low aspect ratio rectangles also revealed near-optimal shifts in initial conditions in response to changes in reward bias (Gao, Tortell, & McClelland, 2011).

In the above cases, participants were shown to perform (near) optimally based on their estimates of stimulus or reward probabilities, when explicitly informed of them. Humans also exhibit adjustment effects in response to patterns of repetitions and alternations that necessarily occur in random sequences, presumably based on a general expectation that locally observed patterns will persist or recur. Models that address RT and accuracy effects reflective of such adjustments appear in Cho et al. (2002); Gao, Wong-Lin, Holmes, Simen, and Cohen (2009); Goldfarb, Wong-Lin, Schwemmer, Leonard, and Holmes (2012); and Jones et al. (2002). These models further support the view that participants adapt by shifting the initial conditions of the decision process. While such adjustments typically degrade performance in response to truly random sequences (they are suboptimal), the ability to extract patterns in natural situations is advantageous, as it allows people to adapt prior beliefs to better match context-specific, stationary (or slowly changing) environments, once they have learned relevant statistics. This possibility is explored in Yu and Cohen (2009).

### 3.3. Stimulus variability and within-trial adjustments

The work described in the previous subsection indicates that participants make trial-to-trial adjustments in decision parameters, especially when patterns are present or apparent in the sequence of stimuli. Modeling work has also suggested that participants can adapt decision parameters *within* trials, especially when task conditions vary from trial to trial. For example, in 2AFC experiments with mixed motion coherences, it has been proposed that slow decisions, typically associated with low SNR and lack of confidence, may be accelerated by a monotonically rising “urgency” signal corresponding to increasing drift rate, a proposal supported by fits to LIP recordings (Hanks, Mazurek, Kiani, Hopp, & Shadlen, 2011). A recent analysis of DD models applied to variable discriminability stimuli shows that drift rate adjustments are suboptimal compared to setting initial conditions (van Ravenzwaaij, Mulder, Tuerlinckx, & Wagenmakers, 2012), but also that human participants appear to combine the two strategies in a similar manner for both fixed and variable discriminability, in contrast to the claims that drift rate bias should dominate in the latter case (Hanks et al., 2011), and that this is necessary for optimality (Yang et al., 2005).

The Eriksen flanker task (Eriksen & Eriksen, 1974) has also been used to study within-trial dynamics of processing and attentional control (e.g., Botvinick et al., 2001; Cohen, Servan-Schreiber, & McClelland, 1992; Gratton et al., 1988). In this task, participants carry out a standard 2AFC on a central stimulus, flanked by stimuli that either invite the same response (compatible) or its alternative (incompatible). Accuracy is lower for incompatible trials and, plotted against RT, displays a dip below 50% followed by a rise to the same asymptote as for compatible trials (see Gratton et al., 1988; fig. 1). The network model of Cohen et al. (1992) suggested that these effects could be explained by dynamic shifts of attention within each trial. In this model, separate units accumulate bottom-up perceptual evidence from the central and flanker stimuli, while subject to top-down attentional influences (Cohen et al., 1992; fig. 8), so that processing of the central stimulus eventually prevails. Subsequently, Liu, Holmes, and Cohen (2008) showed that shifting attention can be captured by a time varying drift rate  $A(t)$  in a one-dimensional OU model. Representing  $A(t)$  as a sum of exponentials, it was possible to fit accuracy/RT data and RT distributions with fewer parameters than the network of Cohen et al. (1992).

A normative interpretation has been provided by a Bayesian analysis that models task performance as dynamic updating of the joint probability distribution for stimulus identity and trial type (compatible/incompatible) (Yu, Dayan, & Cohen, 2009). Updating Bayesian posteriors may seem unrelated to the model of Cohen et al. (1992) or its reduction (Liu et al., 2008), but for Gaussian stimulus probabilities with sufficiently large variance, it can be well approximated by a set of uncoupled DD processes that allow analytical estimation of the evolving posteriors, and hence of accuracy and RT distributions (Liu, Yu, & Holmes, 2009). These also fit the data of Cohen et al. (1992), suggesting both that participants adjust decision parameters within each trial, and that the network of Cohen et al. (1992) represents a neurally plausible model of this (near-) optimal computation.

### 3.4. Exploration, exploitation, and neuromodulation

Thus far, we have focused on studies that examine performance in the context of a particular task. While the decision parameters optimal for that task may vary over stimuli, more general task features have remained constant (e.g., the domain of the stimuli, responses and rewards, the relevant set of stimulus-response mappings, etc.). Under such stationary conditions, good performance implies *exploitation* of a well-defined strategy. However, real-world conditions can change, requiring *exploration* of different possible goals (i.e., what new tasks should be performed). This typically involves forgoing immediate rewards in the service of greater ones over the longer term. Balancing this trade-off involves a higher level form of optimization. While there is no general solution to this problem, a growing body of research seeks to understand how organisms, including humans, manage the exploitation/exploration trade-off (Cohen, McClure, & Yu, 2007).

One theory proposes that the neuromodulatory brainstem nucleus locus coeruleus (LC) contributes to regulating this balance, by dynamically modulating neural responsivity through the phasic release of norepinephrine (NE; Aston-Jones & Cohen, 2005; Servan-Schreiber, Printz, & Cohen, 1990; Usher, Cohen, Servan-Schreiber, Rajkowsky, & Aston-Jones, 1999). According to this adaptive gain theory, transient increases in LC-NE activity promote exploitation by increasing the gain of task-relevant processes, and thereby optimize performance of the current task (Brown et al., 2005; Shea-Brown, Gilzenrat, & Cohen, 2008). In contrast, sustained (and therefore task-indiscriminate) increases in LC-NE activity favor off-task processes, promoting exploration. Shifts between these modes are driven by ongoing assessments of task utility, such that sustained increases favor exploitation, while prolonged decreases favor exploration (Aston-Jones & Cohen, 2005). A closely related proposal suggests that fluctuations in uncertainty can similarly drive the LC-NE system (Yu & Dayan, 2005).

These mechanisms were implemented in a reinforcement learning (RL) model of reversal learning (McClure, Gilzenrat, & Cohen, 2006). This study showed that including the LC-NE gain regulation mechanism and a simple cortical performance monitoring mechanism in a standard RL model dramatically improved the system's ability to adapt to changes in reinforcement contingencies, and it provided a qualitative fit to both behavioral data and direct recordings from the LC in monkeys performing a reversal task (Cohen, McClure, Gilzenrat, & Aston-Jones, 2005). Following reversals, the combined effects of gain modulation and RL allowed fast transitions from exploration to exploitation, whereas with RL alone, such transitions were associated with substantially higher ERs that persisted for longer. Subsequent studies in humans have used the observation that pupil diameter closely tracks LC activity (Aston-Jones & Cohen, 2005) to demonstrate that engagement of the LC-NE system conforms to predictions about its role in regulating the explore/exploit trade-off (Einhäuser, Stout, Koch, & Carter, 2008; Gilzenrat, Nieuwenhuis, & Cohen, 2010; Jepma & Nieuwenhuis, 2011; Murphy, Robertson, Balsters, & O'Connell, 2011). Further work remains to more rigorously test the extent to which this is achieved optimally, under suitably constrained conditions.

### 3.5. Physiological constraints?

Most of the theoretical work reviewed above is based on relatively abstract models. While these have provided important insights into constraints on optimality at the systems level, they do not address constraints that may arise from the underlying neural circuits. Explorations of these have begun by relating abstract models to biophysical aspects of neural function. For example, spiking-neuron models can be reduced in dimension by averaging over populations of cells (Eckhoff, Wong, & Holmes, 2011; Wong & Wang, 2006), allowing them to include the effects of synaptic time constants and neurotransmitters such as NE (Eckhoff, Wong, & Holmes, 2009).

The resulting nonlinear differential equations are more complex than the LCA Eqs. (1-2), and unlike the optimal DD process of Eq. (5), they can possess multiple stable states (Eckhoff et al., 2011, fig. 13; Wong & Wang, 2006, figs. 4 and 5). Even when there is an attracting curve, nonlinear dynamics can cause suboptimal integration, and deviations from the curve can blur the decision thresholds (van Ravenzwaaij, der Maas, & Wagenmakers, 2012; fig. 7). Adjustments in baseline activity and gain can keep accumulators in near-linear dynamical ranges (Cohen, Dunbar, & McClelland, 1990; Servan-Schreiber et al., 1990), but the nonlinear effects suggest that there are physiological obstructions to optimality, especially when task conditions span a wide range. Improvements of low-dimensional models, derived from biophysically detailed neural networks and coupled with multi-unit and multi-area recordings, may help determine whether and how such constraints cause significant deviations from optimality.

## 4. Summary and conclusions

We have reviewed leaky accumulator and drift-diffusion models of 2AFC decision making and shown how the latter can generate normative descriptions of optimal performance (Sections 2.1 and 2.2). These low-dimensional models are fast to simulate, describe behavior remarkably well (Gold & Shadlen, 2001, 2002; Smith & Ratcliff, 2004; Usher & McClelland, 2001), admit analytical study (Bogacz et al., 2006; Brown et al., 2005), and provide understanding of mechanisms. They have generated experiments that revealed failures to optimize (Section 2.3) and motivated the prediction that optimality demands accurate timing ability (Section 2.5), as well as accounting for the costs of control. Further experiments revealed that, while practice can improve participants' strategies, their asymptotic performance is indeed correlated with timing ability (Section 2.6), and that data across a wide range of stimulus discriminability were also consistent with a single, fixed decision threshold that minimized the costs of control. The explicit OPC expression (10), derived from the simple DD model, was crucial to this procedure; we doubt that such predictions could as readily emerge from computational simulations alone. Furthermore, the parameter-free OPC provides a normative benchmark against which choices of speed-accuracy trade-offs can be assessed and compared across different tasks, conditions, and individuals, irrespective of task difficulty and/or timing.



More fundamentally, tests of optimality require precise definitions of objective functions and clear understanding of how decision-making processes are parameterized to achieve them, raising questions about cognitive control, including its costs (Section 3.1) and roles in modulating behavior in changing environments (Sections 3.2 and 3.3). The trade-off between exploitation and exploration is also attracting much attention, as described in Section 3.4. Such studies of more complex behaviors than binary decisions will suggest more sophisticated objective functions and thus enable more realistic optimality assays.

The call for this special issue noted that bounds to optimal performance may be set explicitly, imposed by the task environment, by experience, or by the information-processing architecture. The statistically stationary 2AFC task, implicitly requiring maximization of reward rate, exemplifies both explicit setting, and imposition by environment (Sections 2.2 and 2.3). Modified reward rates, priors, sequential effects, and within-trial adjustments (Sections 2.4, 3.2, and 3.3) reflect experience, as do human responses to timing ability (Sections 2.5 and 2.6) and the costs of control (Section 3.1). Timing ability and control costs can also be seen as constraints imposed by neural architecture, along with those due to the nonlinear biophysics of spiking neurons (Section 3.5). All these factors appear relevant in limiting performance. Nonetheless, they suggest that, given sufficient instruction and experience and allowing for intrinsic constraints, humans can approximate optimality, at least in simple 2AFC tasks.

## Acknowledgments

This work was supported by AFOSR grants FA9550-07-1-0537 and FA9550-07-1-0528 and NIH P50 MH62196. The authors thank Fuat Balci for providing Figs. 3 and 4, and gratefully acknowledge the contributions of many collaborators, not all of whose work could be described in this article.

## References

- Aston-Jones, G., & Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: Adaptive gain and optimal performance. *Annual Review of Neuroscience*, 28, 403–450.
- Balci, F., Simen, P., Niyogi, R., Saxe, A., Holmes, P., & Cohen, J. D. (2011). Acquisition of decision making criteria: Reward rate ultimately beats accuracy. *Attention, Perception and Psychophysics*, 73(2), 640–657.
- Ben-Haim, Y. (2006). *Information gap decision theory: Decisions under severe uncertainty* (2nd ed.). New York: Academic Press.
- Bogacz, R., Brown, E., Moehlis, J., Holmes, P. & Cohen, J. D. (2006). The physics of optimal decision making: A formal analysis of models of performance in two alternative forced choice tasks. *Psychological Review*, 113(4), 700–765.
- Bogacz, R., Hu, P., Cohen, J. D., & Holmes, P. (2010). Do humans produce the speed-accuracy tradeoff that maximizes reward rate? *Quarterly Journal of Experimental Psychology*, 63, 863–891.

- Bohil, C. J., & Maddox, W. T. (2003). On the generality of optimal versus objective classifier feedback effects on decision criterion learning in perceptual categorization. *Memory and Cognition*, 31(2), 181–198.
- Botvinick, M., Braver, T., Barch, D., Carter, C., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, 108(3), 625–652.
- Botvinick, M. M. and Cohen, J. D. (2014). The computational and neural basis of cognitive control: Charted territory and new frontiers. *Cognitive Science* (in press).
- Botvinick, M. M., Huffstetler, S., & McGuire, J. (2009). Effort discounting in human nucleus accumbens. *Cognitive, Affective and Behavioral Neuroscience*, 9, 16–27.
- Britten, K. H., Shadlen, M. N., Newsome, W. T., & Movshon, J. A. (1993). Responses of neurons in macaque MT to stochastic motion signals. *Visual Neuroscience*, 10, 1157–1169.
- Brown, J. W. & Braver, T. S. (2005). Learned predictions of error likelihood in the anterior cingulate cortex. *Science*, 307(5712), 1118–1121.
- Brown, E., Gao, J., Holmes, P., Bogacz, R., Gilzenrat, M. and Cohen, J. D. (2005). Simple networks that optimize decisions. *International Journal of Bifurcation and Chaos in Applied Sciences and Engineering*, 15(3), 803–826.
- Buhusi, C. V., & Meck, W. H. (2005). What makes us tick? Functional and neural mechanisms of interval timing. *Nature*, 6, 755–765.
- Busemeyer, J. R. & Townsend, J. T. (1993). Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment. *Psychological Review*, 100, 432–459.
- Cho, R. Y., Nystrom, L. E., Brown, E. T., Jones, A. D., Braver, T. S., Holmes, P., & Cohen, J. D. (2002). Mechanisms underlying dependencies of performance on stimulus history in a two-alternative forced-choice task. *Cognitive, Affective and Behavioral Neuroscience*, 2(4), 283–299.
- Cohen, J. D., Dunbar, K., & McClelland, J. L. (1990). On the control of automatic processes: A parallel distributed processing model of the Stroop effect. *Psychological Review*, 97(3), 332–361.
- Cohen, J. D., McClure, S. M., Gilzenrat, M., & Aston-Jones, G. (2005). Dopamine-norepinephrine interactions: Exploitation vs. exploration. *Neuropsychopharmacology*, 30, S1, S28.
- Cohen, J. D., McClure, S. M., & Yu, A. J. (2007). Should I stay or should I go? How the human brain manages the tradeoff between exploitation and exploration. *Philosophical Transactions of the Royal Society of London. Series B.*, 362(1481), 933–942.
- Cohen, J. D., Servan-Schreiber, D., & McClelland, J. L. (1992). A parallel distributed processing approach to automaticity. *The American Journal of Psychology*, 105, 239–269.
- Dutilh, G., Vandekerckhove, J., Tuerlinckx, F., & Wagenmakers, E. J. (2009). A diffusion model decomposition of the practice effect. *Psychonomic Bulletin and Review*, 16(6), 1026–1036.
- Eckhoff, P., Wong, K.-F., & Holmes, P. (2009). Optimality and robustness of a biophysical decision-making model under norepinephrine modulation. *Journal of Neuroscience*, 29(13), 4301–4311.
- Eckhoff, P., Wong, K.-F., & Holmes, P. (2011). Dimension reduction and dynamics of a spiking neuron model for decision making under neuromodulation. *SIAM Journal on Applied Dynamical Systems*, 10(1), 148–188.
- Einhäuser, W., Stout, J., Koch, C., & Carter, O. L. (2008). Pupil dilation reflects perceptual selection and predicts subsequent stability in perceptual rivalry. *Proceedings of the National Academy of Sciences of the United States of America*, 105, 1704–1709.
- Eriksen, B. A. & Eriksen, C. W. (1974). Effects of noise letters upon the identification of target letters in a non-search task. *Perception and Psychophysics*, 16, 143–149.
- Feng, S., Holmes, P., Rorie, A., & Newsome, W. T. (2009). Can monkeys choose optimally when faced with noisy stimuli and unequal rewards? *PLoS Computational Biology*, 5(2), e1000284.
- Gao, J., Tortell, R., & McClelland, J. L. (2011). Dynamic integration of reward and stimulus information in perceptual decision-making. *PLoS ONE*, 6(3), e16749.
- Gao, J., Wong-Lin, K. F., Holmes, P., Simen, P., & Cohen, J. D. (2009). Sequential effects in two-choice reaction time tasks: Decomposition and synthesis of mechanisms. *Neural Computation*, 21(9), 2407–2436.

- Gibbon, J. (1977). Scalar expectancy theory and Weber's law in animal timing. *Psychological Review*, 84(3), 279–325.
- Gilzenrat, M. S., Nieuwenhuis, S., & Cohen, J. D. (2010). Pupil diameter tracks changes in control state predicted by the adaptive gain theory of locus coeruleus function. *Cognitive, Affective and Behavioral Neuroscience*, 10, 252–269.
- Gold, J. I. & Shadlen, M. N. (2001). Neural computations that underlie decisions about sensory stimuli. *Trends in Cognitive Sciences*, 5(1), 10–16.
- Gold, J. I., & Shadlen, M. N. (2002). Banburismus and the brain: Decoding the relationship between sensory stimuli, decisions, and reward. *Neuron*, 36, 299–308.
- Goldfarb, S., Wong-Lin, K. F., Schwemmer, M., Leonard, N. E., & Holmes, P. (2012). Can post-error dynamics explain sequential reaction time patterns? *Frontiers in Psychology*, 3, 213.
- Gratton, G., Coles, M. G., Sirevaag, E. J., Eriksen, C. W., & Donchin, E. (1988). Pre- and poststimulus activation of response channels: A psychophysiological analysis. *Journal of Experimental Psychology: Human Perception and Performance*, 14, 331–344.
- Grossberg, S. (1988). Nonlinear neural networks: principles, mechanisms, and architectures. *Neural Networks*, 1, 17–61.
- Hanks, T. D., Mazurek, M. E., Kiani, R., Hopp, E., & Shadlen, M. N. (2011). Elapsed decision time affects the weighting of prior probability in a perceptual decision task. *Journal of Neuroscience*, 31(17), 6339–6352.
- Henik, A., Bibi, U., Yanai, M., & Tzelgov, J. (1997). The Stroop effect is largest during first trials. Abstracts of the Psychonomic Society.
- Holroyd, C. B., & Coles, M. G. H. (2002). The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity. *Psychological Review*, 109(4), 679–709.
- Jepma, M., & Nieuwenhuis, S. (2011). Pupil diameter predicts changes in the exploration-exploitation trade-off: Evidence for the adaptive gain theory. *Journal of Cognitive Neuroscience*, 23, 1587–1596.
- Jones, A. D., Cho, R. Y., Nystrom, L. E., Cohen, J. D., Brown, E. T., & Braver, T. S. (2002). A computational model of anterior cingulate function in speeded response tasks: Effects of frequency, sequence, and conflict. *Cognitive, Affective and Behavioral Neuroscience*, 2(4), 300–317.
- Kiani, R. and Shadlen, M. N. (2009). Representation of confidence associated with a decision by neurons in the parietal cortex. *Science*, 324(5928), 759–764.
- Laming, D. (1979). Choice reaction performance following an error. *Acta Psychologica*, 43(3), 199–224.
- Laming, D. R. J. (1968). *Information theory of choice-reaction times*. New York: Academic Press.
- Lindsay, D. S., & Jacoby, L. L. (1994). Stroop process dissociations: The relationship between facilitation and interference. *Journal of Experimental Psychology: Human Perception and Performance*, 20, 219–234.
- Liu, Y., Holmes, P., & Cohen, J. D. (2008). A neural network model of the Eriksen task: Reduction, analysis, and data fitting. *Neural Computation*, 20(2), 345–373.
- Liu, Y. S., Yu, A. J., & Holmes, P. (2009). Dynamical analysis of Bayesian inference models for the Eriksen task. *Neural Computation*, 21(6), 1520–1553.
- Logan, G. D., & Zbrodoff, N. J. (1979). When it helps to be misled: Facilitative effects of increasing the frequency of conflicting stimuli in a Stroop-like task. *Memory and Cognition*, 7(3), 166–174.
- Maddox, W. T., & Bohil, C. J. (1998). Base-rate and payoff effects in multidimensional perceptual categorization. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 24(6), 1459–1482.
- Mazurek, M. E., Roitman, J. D., Ditterich, J., & Shadlen, M. N. (2003). A role for neural integrators in perceptual decision making. *Cerebral Cortex*, 13(11), 891–898.
- McClure, S. M., Gilzenrat, M. S., & Cohen, J. D. (2006). An exploration-exploitation model based on norepinephrine and dopamine activity. In Y. Weiss, B. Sholkopf, & J. C. Platt (Eds.), *Advances in neural information processing systems 18* (pp. 867–874). Cambridge, MA: MIT Press.
- Miller, K. D., & Fumarola, F. (2012). Mathematical equivalence of two common forms of firing rate models of neural networks. *Neural Computation*, 24, 25–31.

- Murphy, P. R., Robertson, I. H., Balsters, J. H., & O'Connell, R. G. (2011). Pupillometry and P3 index the locus coeruleus noradrenergic arousal function in humans. *Psychophysiology*, *48*, 1532–1543.
- Myung, I. J., & Busemeyer, J. R. (1989). Criterion learning in a deferred decision making task. *The American Journal of Psychology*, *102*(1), 1–16.
- Petrov, A. A., Van Horn, N. M., & Ratcliff, R. (2011). Dissociable perceptual-learning mechanisms revealed by diffusion-model analysis. *Psychonomic Bulletin and Review*, *18*(3), 490–497.
- Posner, M. I., & Snyder, C. R. R. (1975). Attention and cognitive control. In R. Solso (Ed.), *Information Processing and Cognition: The Loyola Symposium* (pp. 55–85). Hillsdale, NJ: Lawrence Erlbaum.
- Rabbitt, P. M. A. (1966). Errors and error-correction in choice-response tasks. *Journal of Experimental Psychology*, *71*, 264–272.
- Rakitin, B. C., Hinton, S. C., Penney, T. B., Malapani, C., Gibbon, J., & Meck, W. H. (1998). Scalar expectancy theory and peak interval timing in humans. *Journal of Experimental Psychology: Animal Behavior Processes*, *24*, 1–19.
- Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*(1), 59–108.
- Ratcliff, R., Cherian, A., & Segraves, M. A. (2003). A comparison of macaque behavior and superior colliculus neuronal activity to predictions from models of two choice decisions. *Journal of Neurophysiology*, *90*, 1392–1407.
- Ratcliff, R., Hasegawa, Y. T., Hasegawa, R. P., Smith, P. L., & Segraves, M. A. (2006). Dual-diffusion model for single-cell recording data from the superior colliculus in a brightness-discrimination task. *Journal of Neurophysiology*, *97*, 1756–1774.
- Ratcliff, R., Van Zandt, T., & McKoon, G. (1999). Connectionist and diffusion models of reaction time. *Psychological Review*, *106*(2), 261–300.
- Roitman, J., & Shadlen, M. (2002). Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. *Journal of Neuroscience*, *22*(21), 9475–9489.
- Rorie, A., Gao, J., McClelland, J. L., & Newsome, W. T. (2010). Integration of sensory and reward information during perceptual decision-making in lateral intraparietal cortex (LIP). *PLoS ONE*, *5*(2), e9308.
- Roxin, A., & Ledberg, A. (2008). Neurobiological models of two-choice decision making can be reduced to a one-dimensional nonlinear diffusion equation. *PLoS Computational Biology*, *4*(3), e1000046.
- Rumelhart, D. E., & McClelland, J. L. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition*. Cambridge, MA: MIT Press.
- Schall, J. D. (2001). Neural basis of deciding, choosing and acting. *Nature Reviews. Neuroscience*, *2*, 33–42.
- Servan-Schreiber, D., Printz, H., & Cohen, J. D. (1990). A network model of catecholamine effects: Gain, signal-to-noise ratio, and behavior. *Science*, *249*, 892–895.
- Shadlen, M. N., & Newsome, W. T. (2001). Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. *Journal of Neurophysiology*, *86*, 1916–1936.
- Shea-Brown, E., Gilzenrat, M., & Cohen, J. D. (2008). Optimization of decision making in multilayer networks: The role of locus coeruleus. *Neural Computation*, *20*, 2863–2894.
- Simen, P., Cohen, J. D., & Holmes, P. (2006). Rapid decision threshold modulation by reward rate in a neural network. *Neural Networks*, *19*, 1013–1026.
- Simen, P., Contreras, D., Buck, C., Hu, P., Holmes, P., & Cohen, J. D. (2009). Reward rate optimization in two-alternative decision making: Empirical tests of theoretical predictions. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(6), 1865–1897.
- Smith, P. L., & Ratcliff, R. (2004). Psychology and neurobiology of simple decisions. *Trends in Neurosciences*, *27*(3), 161–168.
- Stone, M. (1960). Models for choice-reaction time. *Psychometrika*, *25*, 251–260.
- Todd, M. T., Botvinick, M. M., Schwemmer, M. A., Cohen, J. D., & Dayan, P. (2011). Normative analysis of task switching. Abstract No. 194.21, Society for Neuroscience Meeting.
- Usher, M., Cohen, J. D., Servan-Schreiber, D., Rajkowsky, J., & Aston-Jones, G. (1999). The role of locus coeruleus in the regulation of cognitive performance. *Science*, *283*, 549–554.

- Usher, M., & McClelland, J. L. (2001). On the time course of perceptual choice: The leaky competing accumulator model. *Psychological Review*, *108*(3), 550–592.
- van Ravenzwaaij, D., Mulder, M. J., Tuerlinckx, F., & Wagenmakers, E.-J. (2012). Do the dynamics of prior information depend on task context? An analysis of optimal performance and an empirical test. *Frontiers in Cognitive Neuroscience*, *3*, 132.
- van Ravenzwaaij, D., der Maas, H. L. J., & Wagenmakers, E. J. (2012). Optimal decision making in neural inhibition models. *Psychological Review*, *119*(1), 201–215.
- Wald, A., & Wolfowitz, J. (1948). Optimum character of the sequential probability ratio test. *The Annals of Mathematical Statistics*, *19*, 326–339.
- Wong, K. F., & Wang, X.-J. (2006). A recurrent network mechanism of time integration in perceptual decisions. *Journal of Neuroscience*, *26*(4), 1314–1328.
- Yang, T., Hanks, T. D., Mazurek, M., McKinley, M., Palmer, J., & Shadlen, M. N. (2005). Incorporating prior probability into decision making in the face of uncertain reliability of evidence. Abstract No. 621.4, Society for Neuroscience Meeting, <http://sfn.scholarone.com>.
- Yeung, N., Botvinick, M. M., & Cohen, J. D. (2004). The neural basis of error detection: Conflict monitoring and the error-related negativity. *Psychological Review*, *111*(4), 931–959.
- Yu, A. J., & Cohen, J. D. (2009). Sequential effects: Superstition or rational behavior? In J. C. Platt, D. Koller, Y. Singer, and S. Roweis (Eds.), *Advances in neural information processing systems 21* (pp. 1873–1880). Cambridge, MA: MIT Press.
- Yu, A. J., & Dayan, P. (2005). Uncertainty, neuromodulation, and attention. *Neuron*, *46*, 681–692.
- Yu, A., Dayan, P., & Cohen, J. D. (2009). Dynamics of attentional selection under conflict: Toward a rational Bayesian account. *Journal of Experimental Psychology: Human Perception and Performance*, *35* (3), 700–717.
- Zacksenhouse, M., Bogacz, R., & Holmes, P. (2010). Robust versus optimal strategies for two-alternative forced choice tasks. *Journal of Mathematical Psychology*, *54*, 230–246.