The physics of optimal decision making: A formal analysis of models of performance in two-alternative forced choice tasks

Rafal Bogacz\textsuperscript{1,2}, Eric Brown\textsuperscript{2}, Jeff Moehlis\textsuperscript{2}, Philip Holmes\textsuperscript{2,4}, Jonathan D. Cohen\textsuperscript{1,3}

\textsuperscript{1}Center for the Study of Brain, Mind and Behavior, Princeton University, Princeton, NJ 08544, USA

\textsuperscript{2}Program in Applied and Computational Mathematics, Princeton University, Princeton, NJ 08544, USA

\textsuperscript{3}Department of Psychology, Princeton University, Princeton, NJ 08544, USA

\textsuperscript{4}Department of Mechanical and Aerospace Engineering, Princeton University, Princeton, NJ 08544, USA

RUNNING HEAD: Optimal decision making

Text pages: 96 main text (excluding references and figure legends) + 2 Appendix

Figures: 18; Tables: 1

Correspondence to:

Rafal Bogacz
Department of Computer Science, University of Bristol, Bristol BS8 1UB, UK
Tel. +44-117-954-5141
Fax. +44-117-954-5208
E-mail: R.Bogacz@bristol.ac.uk
Abstract

In this article we consider optimal decision making in two alternative forced choice (TAFC) tasks. We begin by analyzing six models of TAFC decision making, and show that all but one can be reduced to the drift diffusion model, implementing the statistically optimal algorithm (most accurate for a given speed, or fastest for a given accuracy). We prove further that there is always an optimal tradeoff between speed and accuracy that maximizes various reward functions including reward rate (percentage of correct responses per unit time) as well as several other objective functions, including ones weighted for accuracy. We use these findings to address empirical data, and make novel predictions about performance under optimality.
1. Introduction

This article concerns optimal strategies for decision making in the two-alternative forced choice (TAFC) task. We shall present and compare several decision-making models, briefly discuss their neural implementations, and relate them to one that is optimal in the sense that it delivers a decision of specified accuracy in the shortest possible time: the drift diffusion model (DDM). Further definitions of optimality, via objective functions such as reward rate, will be given below, and we shall provide explicit solutions to the speed-accuracy tradeoff for the TAFC task.

It has been known since Herrnstein's work (1961, 1997) that animals do not achieve optimality under all conditions, and in behavioral economics humans often fail to choose optimally (e.g., Kahneman & Tversky, 1984; Loewenstein & Thaler, 1989). For example, in selecting among two items of comparable value and a cash payment worth less than either, participants often opt for cash, possibly to avoid a harder decision between two more valuable but similar items (Shafir & Tversky, 1995).

Such findings raise important questions: Is human decision making always suboptimal? If it is not, under what conditions is it optimal? And, when it is suboptimal, does this reflect inherent deficiencies in decision-making mechanisms, or other sources of systematic bias? Answers to such questions may provide insights into underlying neural mechanisms. However, to consider them, one must first describe optimal performance, against which actual behavior can be compared.

Optimality has long been a key principle in the physical sciences (e.g., energy minimization and related variational methods) and it has recently begun to see application in the life sciences, including behavioral ecology (Belovsky, 1984; Dunbar, 1984; Bull et al., 1996) and neuroscience (e.g., Bialek, 1991; Olshausen & Field, 1996; Levine & Elsberry, 1997).
Optimality is also a cornerstone of standard economic theory, and its application to human decision making (e.g., the rational agent model and the principle of utility maximization). However, this has focused on optimal outcomes, and not the dynamics of decision processes. Some theories based on optimal performance have appeared in psychology as well (e.g., Edwards, 1965; Link, 1975; Carpenter & Williams, 1995; and Mozer et al., 2002). Anderson's (1990) rational analysis is perhaps the most fully developed example, and models based on it have illuminated cognitive functions including memory, categorization and reasoning (Anderson, 1990; Oaksford & Chater, 1998).

In this article, we adopt a similar approach with regard to decision making, with an additional focus on how it may be implemented in the brain. We do so within the context of a set of highly simplified decision making conditions, one in which a choice must be made between two responses based on limited information about which is correct (that is, which will be rewarded), and in which both the speed and the accuracy of the decision impact the total cumulative reward that is accrued. These conditions are often referred to as the two-alternative forced choice paradigm.

1.1. The two-alternative forced choice (TAFC) task

Choosing between two alternatives, even under time pressure and with uncertain information, is a simplification of many situations, but we focus on it for several reasons. First, it is representative of many problems faced by animals in their natural environments (e.g., whether to approach or avoid a novel stimulus). Pressures for speed and accuracy in such constrained situations may have exerted strong evolutionary influences, thereby optimizing mechanisms. Second, a wealth of human behavioral data generated since the late 19th century (e.g., Hill, 1898) has motivated formal modeling of the dynamics and response outcomes in TAFC tasks (e.g., Stone, 1960; LaBerge, 1962; Link & Heath, 1975; Link, 1975; Pike, 1966; Laming,
Finally, neuroscientists can now monitor neuronal dynamics and assess its relationship to task performance. In many cases, neural and behavioral data are converging to support formal models such as the DDM (e.g., Hanes & Schall, 1996; Shadlen & Newsome, 1996, 2001; Schall, 1998, 2001; Gold & Shadlen, 2002; Ratcliff et al., 2003b).

TAFC task models typically make three fundamental assumptions: i) evidence favoring each alternative is integrated over time; ii) the process is subject to random fluctuations; and iii) the decision is made when sufficient evidence has accumulated favoring one alternative over the other. A central question, to which we will return, is whether evidence for each alternative is integrated independently, or whether the difference in evidence is integrated. Most current theories assume that the difference in evidence drives the decision. In neural models, differences can be computed by inhibitory mechanisms, but theories vary in how inhibition is implemented, leading to different behavioral predictions. Several comparisons of theories with empirical data have appeared (e.g., Vickers et al., 1971; Smith & Vickers, 1989; Ratcliff et al. 1999; Van Zandt et al., 2000; Usher & McClelland, 2001; Smith & Ratcliff, 2004; Ratcliff & Smith, 2004), but a systematic mathematical analysis that compares models with one another and with optimal performance is lacking.

Our first goal is to conduct such a study. This is a key step if we are to decide which model best describes the data. We shall relate several existing models to a particular “standard”: the DDM (Stone, 1960; Laming, 1968; Ratcliff, 1978). We adopt the DDM as a reference because it is simple, well-characterized (e.g., Smith, 2000), has been proven to implement the optimal mechanism for TAFC decision making (e.g., Laming, 1968), and accounts for an impressive array of behavioral and neuroscientific data (e.g., Ratcliff, 1978; Hanes & Schall, 1996; Schall & Thompson, 1999; Ratcliff & Rouder, 2000; Schall, 2001; Shadlen &
Newsome, 2001; Gold & Shadlen, 2002; Thapar et al., 2003; Ratcliff et al., 2003a, 2004; Smith & Ratcliff, 2004).

1.2. The DDM for decision making in the TAFC paradigm

In applying the DDM to the TAFC we assume that the difference in the (noisy) information favoring each alternative is integrated over each trial, and a decision is reached when the resulting accumulated value crosses a critical threshold.

Discrete analogs of the DDM were first applied to decision making in World War II, via the sequential probability ratio test (SPRT), the theory being developed independently by Barnard (1946) and Wald (1947) (cf. Lehmann, 1959). (For histories, see Good (1979); Wallis (1980); DeGroot (1988).) It was subsequently shown that the DDM and its discrete analog, a directed random walk, describes accuracy and reaction times (RT) in humans performing TAFC tasks (Stone, 1960; Laming, 1968; Ratcliff, 1978). In particular, DDM first passages account for a robust feature of human performance: RT distributions are heavily skewed towards longer times (Figure 1a).

More recently, neural firing patterns have been examined in light of the DDM (e.g., Hanes & Schall, 1996; Schall & Thompson, 1999; Shadlen & Newsome, 2001; Gold & Shadlen, 2002). For example, Shadlen & Newsome (1996) have studied the activity of neurons in the lateral intraparietal (LIP) area (which code for gaze direction) in monkeys performing a TAFC task in which they respond by saccading to one of two targets. Following stimulus presentation, activity increases in neurons selective for both directions, with those selective for the correct response rising fastest on average (Figure 1c). Critically, the time at which their firing rate reaches a fixed threshold predicts the response.

As noted above and elaborated below, the DDM is optimal in the sense that it is the fastest decision maker for a given level of accuracy. This assumes a fixed decision threshold, yet the
model does not specify what this should be. Absent noise, thresholds have no effect on accuracy. However, with noisy data, accuracy increases as thresholds and decision times rise, since there is more time to average out uncorrelated noise. This produces a *speed-accuracy tradeoff*: lower thresholds produce faster but less accurate responding, while higher thresholds produce more accurate but slower responses. This speed-accuracy tradeoff is critical for interpreting the results of most psychological experiments (Pachella, 1974; Wickelgren, 1977). For example, a manipulation intended to influence task difficulty, and thereby accuracy, could instead simply induce a change in response threshold, thereby changing RTs.

A number of different theories of how humans (or animals in general) set thresholds or otherwise manage the speed-accuracy tradeoff have been proposed (Edwards 1965; Rapoport & Burkheimer, 1971; Busemeyer & Rapoport, 1988; Myung & Busemeyer, 1989; Maddox & Bohil, 1998; Gold & Shadlen, 2002). We will return to this in the second part of the article (Section 5), where we extend the DDM to show how thresholds can be chosen to optimize performance according to various criteria, and thereby make novel predictions.

1.3. Overview of the article

We have two primary goals: i) to analyze the relationship of the DDM to other decision-making models, and ii) to address the broader question of optimality, specifically identifying optimal parameters for TAFC tasks. We hope thereby to offer a unified framework in which to pursue future modeling and quantitative studies of decision-making behavior.

The article is organized as follows. In Section 2 we provide formal descriptions of the TAFC task, the SPRT and DDM, specifying how key quantities such as error rate (ER) and mean decision time (DT) depend on the parameters characterizing the DDM. We then review five other decision-making models. We analyze the relationship of these to the DDM in Section 3,
showing that all but one of them is a variant of it, at least in a limiting sense. We thereby clarify the relationships among several theories and their differing predictions, facilitating further analysis and discussion. We illustrate this by fitting DDM parameters to empirical TAFC data, and using these as a reference for comparing models throughout the remainder of the article.

We then address the question of optimality. In Section 4 we show that the parameters optimizing performance of the other decision-making models are precisely those for which the models reduce to the DDM. We review several criteria for optimality in Section 5, demonstrating that each implies a unique optimal threshold and speed-accuracy tradeoff, and illustrating their dependence on task parameters. We further show that the DDM implements the optimal decision-making procedure for all the criteria, and predict novel patterns that should emerge in empirical data as signatures of optimal performance in each case. In Section 6 we consider biased decisions and rewards, in which one alternative is more probable, or the reward associated with it more valuable, than the other. We thereby explain some published neurophysiological data, and make new quantitative predictions about behavioral strategies in these cases.

We finally discuss extensions to the DDM that might provide more complete accounts of decision-making behavior, arguing that it offers an attractive framework for further theoretical analysis and the design of empirical studies. Throughout, we restrict our formal presentation to the most important equations, and where possible we provide intuitive explanations. Further mathematical details are available in an online Appendix.
2. Background

2.1. Behavioral and neurophysiological data from the TAFC

In a common version of the TAFC task, participants must identify the direction of a coherently-moving subset of dots embedded in a random motion field (Britten et al., 1993). Critical parameters under experimenter control include: i) stimulus fidelity or difficulty, which can be manipulated by varying the fraction of coherently moving dots; ii) whether participants are allowed to respond freely or responses are cued or deadlined; and iii) the delay between response and the next stimulus.

In addition to their use in the study of behavior, TAFC tasks are also used widely in neurophysiological studies, in which direct recordings are made from brain areas involved in task performance (e.g., Shadlen & Newsome, 1996, 2001). Figure 1b shows a representation of typical firing rates observed in area MT of monkeys trained on the moving dots task (MT is involved in motion processing). When a stimulus with coherent leftward motion is presented, the firing rate of an MT neuron selective for leftward motion typically exceeds that of one selective for rightward motion (Britten et al., 1993) – the grey curve in the figure is more often above the black one. However, both firing rates are noisy, hence decisions based on instantaneous activities of MT neurons would be inaccurate, reflecting uncertainty inherent in the stimulus and its neural representation.

Figure 1c shows activity patterns of neurons in area LIP (involved in eye movement control): they clearly separate as time increases. LIP neurons are believed to integrate the noisy MT outputs over each trial, leading to more accurate decisions. Neural integration mechanisms have been studied extensively in the context of control of motor processes (e.g., Cannon et al., 1983; Seung, 1996; Seung et al., 2000; Aksay et al., 2001; Koulakov et al., 2002; Goldman et al., 2003; Major et al., 2004).
2.2. The decision problem

Motivated by the above example, we formalize the TAFC decision problem by postulating two neuronal populations whose activities (firing rates) provide evidence for the two alternatives. We denote their mean activities during a given trial by $I_1$ and $I_2$, and assume that each experiences independent random fluctuations with the same standard deviation, $c$. The goal is to identify which of $I_1$ and $I_2$ is greatest, and the optimality question becomes: For given signal and noise levels $I_1$, $I_2$ and $c$, what strategy provides the most accurate and fastest decisions on average? More precisely, there are two questions: (i) which strategy yields the lowest expected ER at a given (fixed) decision time, and (ii) which strategy yields the shortest expected RT for a given ER?

These questions correspond to two paradigms under which TAFC tasks can be run. The first typically involves fixed duration stimuli, after which participants are expected to answer, usually on presentation of a signal to respond, thus constraining their RTs. We shall refer to the “hard limit” case, in which decisions are supposed to be made instantly at the cue, as the interrogation paradigm. The second addresses a free-response paradigm under which participants respond in their own time. Since in the latter both ERs and RTs vary (participants implicitly choose a speed-accuracy tradeoff), one may assess optimality by asking which among strategies that yield the same ER, gives the shortest RT. These paradigms represent the extremes of a continuum of more general deadlining tasks, in which responses can be made at any time before a fixed deadline.

We now formalize the questions posed above, which both hinge on determining whether the difference of the means $I_1 - I_2$ is positive or negative. Let us denote by $Y$ the random variable corresponding to the difference in activity of two neuronal populations, where the activity of each represents accumulated evidence for one of the two alternatives. We suppose that
successive samples of $Y$ within each trial are drawn from one of two probability distributions with densities $p_1(y)$ and $p_2(y)$ with means $\mu_1$ and $\mu_2$. Hence we must decide which of the hypotheses $H_1 (I_1 - I_2 = \mu_1 > 0)$ or $H_2 (I_1 - I_2 = \mu_2 < 0)$ is correct.

The answer to the question (i) above is given by the procedure of Neyman & Pearson (1933). To decide from which distribution a random sequence $Y = y_1, y_2, \ldots, y_n$ is drawn, we calculate the likelihood ratio of the sample $Y$ given the hypotheses $H_1$ and $H_2$:

$$\frac{\frac{p_{1n}}{p_{2n}}} = \frac{p_1(y_1)p_1(y_2)\ldots p_1(y_n)}{p_2(y_1)p_2(y_2)\ldots p_2(y_n)}. \quad (2.1)$$

Here $p_1(y_i)$ and $p_2(y_i)$ denote the probabilities of each observation $y_i$ assuming that hypothesis $H_1$ or $H_2$ is true respectively. The ratio $p_1(y_i)/p_2(y_i)$ is greater than 1 if the observation $y_i$ supports $H_1$ (i.e., if $y_i$ is more likely under $H_1$ than $H_2$) and less than 1 if it supports $H_2$. Since all observations are assumed independent, $p_{1n}$ and $p_{2n}$ denote the probabilities of observing the sequence of observations $y_1, y_2, \ldots, y_n$ under $H_1$ or $H_2$ respectively. Hypothesis $H_1$ (or $H_2$) is accepted if the ratio of Equation 2.1 is less than $Z$ (or greater than $Z$), where $Z$ is a constant determined by the desired level of accuracy for one of the hypotheses\(^1\). Neyman & Pearson (1933) showed that, for fixed sample size $n$, setting $Z = 1$ delivers the most likely hypothesis and minimizes the total error probability. Hence this procedure minimizes ER for fixed decision time, thus is optimal for the interrogation paradigm. (Here and throughout we do not consider any explicit cost for acquiring evidence.)

The answer to the question (ii) above is provided by the sequential probability ratio test (SPRT) of Barnard (1946) and Wald (1947). Here, the goal is to decide as soon as a stream of

\(^1\) The Neyman-Pearson procedure underlies many statistical tests used in psychology to evaluate an experimental hypothesis on a fixed number of experimental samples, including the t-test.
incoming data reaches a predetermined level of reliability. Again assuming that samples are
drawn at random from one of two distributions with densities \( p_1(y) \), \( p_2(y) \), the running product
of the likelihood ratios is calculated as in Equation 2.1, but now observations continue only as
long as the product lies within predefined boundaries \( Z_2 < Z_1 \):

\[
Z_2 < \frac{p_{1n}}{p_{2n}} < Z_1.
\] (2.2)

Thus, after each measurement one updates the likelihood ratio, thereby assessing the net
weight of evidence in favor of \( H_1 \) over \( H_2 \). When the ratio first exceeds \( Z_1 \) or falls below \( Z_2 \),
sampling ends and either \( H_1 \) or \( H_2 \) is accepted; otherwise sampling continues\(^2\). The SPRT is
optimal in the following sense: Among all fixed or variable sample decision methods that
guarantee fixed error probabilities, SPRT requires on average the smallest number of samples
to render a decision (Wald & Wolfowitz, 1948). Thus, for a given ER, SPRT delivers the
fastest RT. In online Appendix A.1 we provide more precise statements and generalizations
to decisions between alternatives with unequal prior probabilities.

The SPRT is equivalent to a random walk with thresholds corresponding to the alternative
choices, as one can see by taking logarithms in Equations 2.1 and 2.2:

\[
\log Z_2 < \log \frac{p_1(y_1)}{p_2(y_1)} + \ldots + \log \frac{p_1(y_n)}{p_2(y_n)} < \log Z_1. \quad (2.3)
\]

Denoting the logarithm of the likelihood ratio by \( l^n \), Equation 2.1 implies that \( l^n \) is additively
updated after each observation:

\(^2\) In context of statistical analysis of experimental data, the SPRT would correspond to checking the condition of
Equation 2.2 after each sample collected and stopping the experiment if Equation 2.2 is not satisfied (i.e., a pre-
specified confidence level is reached) rather than waiting for a fixed number of samples and then evaluating the
hypotheses for significance (i.e., probability).
\[ I^n = I^{n-1} + \log \frac{p_1(y_n)}{p_2(y'_n)}, \tag{2.4} \]

(cf. Gold & Shadlen, 2002). The SPRT is therefore equivalent to a random walk starting at \( I^0 = 0 \) and continuing until \( I^n \) reaches the threshold \( \log Z_1 \) or \( \log Z_2 \). Moreover, as discrete samples are taken more frequently and one approaches sampling of a continuous variable, the SPRT converges on the DDM and the discrete log likelihood ratio \( I^n \) becomes a continuous time-dependent variable \( x(t) \), after a change of scale. Details are given in online Appendices A.1.3-1.4.3 As shown in Section 2.3, this continuum limit yields explicit formulae and key quantitative predictions.

In the following sections we review six mathematical models of TAFC decision making, starting with the DDM as developed by Ratcliff (1978; Ratcliff et al., 1999). We then discuss the Ornstein-Uhlenbeck model (Busemeyer & Townsend, 1993), since it provides a link between the DDM and the biologically-motivated models that follow (Usher & McClelland, 2001; Ditterich et al., 2003; Wang, 2002). These employ different forms of inhibition to compute differences in signals associated with each alternative. We also discuss the well-established and widely-used race model (LaBerge, 1962; Vickers, 1970; Smith & Vickers, 1989; Smith & Van Zandt, 2000; Logan, 2002; Logan & Bundesen, 2003). All six models are represented as simplified stochastic differential equations in which only essential elements are retained. In particular, for ease of exposition and mathematical tractability, we describe only linearized systems, although many of our observations extend to nonlinear models. Explicit comparisons among linear, piecewise-linear, and nonlinear (sigmoidal) models show that

---

3 In the online Appendices A.2-A.3 we also provide direct arguments, independent of the SPRT, suggesting that the DDM is optimal in all these senses.
linearized models often capture the key dynamics and parameter dependencies (Brown & Holmes, 2001; Brown et al., 2005; cf. Usher and McClelland, 2001).

### 2.3. Drift Diffusion model (DDM)

As indicated above, in the DDM (Ratcliff, 1978) one accumulates the difference between the amounts of evidence supporting the two hypotheses. We denote the accumulated value of this difference at time $t$ by $x(t)$, and assume that $x = 0$ represents equality in the amounts of integrated evidence. We shall consider two versions of the DDM, the first being a continuum limit of the random walk model (Laming, 1968) that we refer to as the pure DDM; and the second or extended DDM a generalized model in which drift rates and starting points may vary across trials (e.g., Ratcliff & Rouder, 1998).

In the pure DDM we start with unbiased evidence and accumulate it according to:

$$dx = Adt + cdW, \quad x(0) = 0.$$  

(Biased decisions are treated in Section 6.) In Equation 2.5, $dx$ denotes the change in $x$ over a small time interval $dt$, which is comprised of two parts: the constant drift $Adt$ represents the average increase in evidence supporting the correct choice per time unit. In terms of Section 2.2, $A > 0$ if $H_1$ is correct for the trial in question, and $A < 0$ if $H_2$ is correct. The second term $cdW$ represents white noise, which is Gaussian distributed with mean 0 and variance $c^2dt$.

Hence $x$ grows at rate $A$ on average, but solutions also diffuse due to the accumulation of noise. Neglecting boundary effects, the probability density $p(x,t)$ of solutions of Equation 2.5 at time $t$ is normally distributed with mean $At$ and standard deviation $c\sqrt{t}$ (Gardiner, 1985):

$$p(x,t) = N(At,c\sqrt{t}).$$  

We model the interrogation paradigm by asking if, at the interrogation time $T$, the current value of $x$ lies above or below zero. If $H_1$ applies, a correct decision is recorded if $x > 0$ and
an incorrect one if \( x < 0 \). The average ER is therefore the probability that a typical solution \( x(T) \) of Equation 2.5 lies below zero at time \( T \), which is obtained by integrating the density \( p(x,T) \) of Equation 2.6 from \(-\infty\) to 0:

\[
ER = \Phi\left(-\frac{A}{c} \sqrt{T}\right), \text{ where } \Phi(y) = \int_{-\infty}^{y} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} \, du.
\]  

(2.7)

(Here, \( \Phi \) is the normal standard cumulative distribution function.)

In the free-response paradigm, the decision is made when \( x \) reaches one of two fixed thresholds, positive or negative. If both alternatives are equally probable, the thresholds are symmetric (\( \pm z \)). Figure 2a shows examples of the evolution of \( x \) during separate trials (sample paths). Although on average \( x \) moves toward the correct threshold, noise causes it to hesitate and “backtrack,” and on some trials it crosses the incorrect threshold.

Solutions of a first passage problem for the pure DDM yield simple expressions revealing how ER and mean decision time DT depend on drift \( A \), noise \( c \), and decision threshold \( z \) (Feller, 1968; cf. Ratcliff, 1978):

\[
ER = \frac{1}{1 + e^{\frac{Az}{c^2}}},
\]

\[
DT = \frac{z}{A} \tanh\left(\frac{Az}{c^2}\right).
\]

(2.8)

(2.9)

(See online Appendix A.2.1.1). Here DT denotes the mean decision time: the fraction of the RT associated with the decision process. We denote the remaining fraction of RT (e.g., due to sensory or motor processes unrelated to the decision process) by \( T_0 \). Thus RT = DT + \( T_0 \).

Since Equations 2.8 and 2.9 involve ratios of \( A \), \( z \) and \( c^2 \), ER and DT do not change if these parameters are scaled by the same constant. All the models reviewed here share this property.

Since the variable \( x \) of the DDM is proportional to the log likelihood ratio, the DDM implements the Neyman-Pearson procedure (with \( Z=1 \)) for the interrogation paradigm and the
SPRT for the free-response paradigm. Hence it is the optimal decision maker for both paradigms.

The extended DDM (e.g., Ratcliff, 1978; Ratcliff et al., 1999) includes two additional elements that improve its fit to experimental data. Intertrial differences in stimulus fidelity or attention can be modeled by allowing drift to vary by selecting $A$ from a normal distribution with mean $m_A$, and standard deviation $s_A$ on each trial (Ratcliff, 1978). Drift variability produces longer DTs for errors than correct responses (Ratcliff & Rouder, 1998, 2000)\(^4\), since most errors occur on trials with low drift which typically have longer DTs. Secondly, the initial value $x(0)$ can be chosen from a uniform distribution ranging from $-s_x$ to $s_x$ (with $s_x < z$). This may reflect subjects’ sensitivity to local frequency variations in presentation of specific stimuli that occur even when they are equally probable overall (e.g., Soetens et al., 1984; Cho et al., 2002) or premature sampling: the integration of noise from sensory neuronal populations before the stimulus onset (Laming, 1978). Starting point variability produces shorter DTs for errors than correct trials (Ratcliff & Rouder, 1998) because these errors occur more often on trials for which the process begins closer to the threshold for the incorrect alternative, and hence crosses this threshold with a relatively short DT.

The expression for ER of the extended DDM in the interrogation paradigm involves an integral that cannot be analytically evaluated, although it can be derived in the presence of drift variability (see Ratcliff (1978), where an equation for $d'$ (d-prime) as a function of interrogation time is also given). Expressions for ER and DT of the extended DDM in the

\(^4\) Throughout this paper we assume that drift is fixed within a trial, and drift variations occur only from trial-to-trial. The case of varying drift within the trial is analyzed in Brown et al. (2005); cf. Ratcliff & Rouder (2000), Ratcliff (1980).
free-response paradigm also involve integrals that cannot be analytically evaluated, and so we generally resort to numerical simulations, although asymptotic approximations are derived online Appendix A.4.1. An efficient numerical method for calculating ER and the distribution of DTs has recently been developed (Tuerlickx, 2004).

## 2.4. Ornstein-Uhlenbeck model

The Ornstein-Uhlenbeck (O-U) model (Busemeyer & Townsend, 1993) differs from the DDM in that a third term, linear in $x$, is added:

$$dx = (\lambda x + A)dt + cdW, \quad x(0) = 0. \quad (2.10)$$

The rate of change of $x$ now also depends on its current value, with a magnitude controlled by the additional parameter $\lambda$, and $x$ can accelerate or decelerate toward a threshold, depending on the sign of $\lambda$. In the following discussion, we again assume that $A > 0$ and $H_1$ is correct.

First suppose $\lambda < 0$. For $x = -A/\lambda (>0)$, $dx$ is zero on average; this value of $x$ corresponds to an equilibrium or fixed point for the noise-free process. When $x < -A/\lambda$, $dx$ is on average positive and $x$ typically grows; when $x > -A/\lambda$, $dx$ is on average negative and $x$ tends to decrease. The fixed point is therefore an attractor. More precisely, for this stable O-U process, the time-dependent probability density of $x$ converges to a stationary normal distribution centered on $x = -A/\lambda$ with standard deviation $c/\sqrt{2\lambda}$ (Gardiner, 1985; Holmes et al., 1996).

$$p(x) = \mathcal{N}\left(-\frac{A}{\lambda}, \frac{c}{\sqrt{2\lambda}}\right). \quad (2.11)$$

Derivations and full expressions for the time-dependent distribution are given in online Appendix A.3.
Now suppose $\lambda > 0$. When $x < -A/\lambda (<0)$, $dx$ is on average negative and $x$ usually decreases, while for $x > -A/\lambda$, it usually increases. Hence $x$ is repelled from the fixed point, more strongly the further away it is, and the mean of $x$ accelerates away from $x = -A/\lambda$ in an unstable O-U process. Thus, the corresponding probability density has exponentially growing mean and variance, with no stationary limit. It is also described by the general expressions developed in online Appendix A.3.

Summarizing, $\lambda < 0$ causes attraction towards the fixed point proportional to the distance of the current state from it, while $\lambda > 0$ causes repulsion. The intuition for these behaviors with respect to decision making is as follows (cf. Figure 6 below). In the stable case all solutions approach and tend to remain near the fixed point, which lies nearer the correct threshold, so they typically slow down before crossing it, corresponding to conservative behavior. The unstable case corresponds to riskier behavior: solutions on the “correct” side of the fixed point accelerate towards the correct threshold, giving faster responses, but solutions on the “incorrect” side accelerate towards the incorrect threshold, possibly producing more errors. For $\lambda = 0$, the O-U model simplifies to the pure DDM.

Busemeyer & Townsend (1993) proposed that rewards for correct responses should increase $\lambda$, while punishments for errors should decrease it. They also note that negative $\lambda$ produces a recency or decay effect over the course of a trial, since later inputs influence accumulated evidence more than earlier ones, while positive $\lambda$ produces a primacy effect (earlier inputs have more influence).

The ER expression for the O-U model in the interrogation paradigm generalizes that for the DDM (Equation 2.7). For interrogation at time $T$ we have (Busemeyer & Townsend, 1992; derived in online Appendix A.3: Equations A.88 and A.92):
$$ER(T) = \Phi \left( -\frac{A}{c} \sqrt{\frac{2(e^{\lambda T} - 1)}{e^{2\lambda T} + 1}} \right).$$

(2.12)

The expressions for ER and DT in the free-response paradigm were derived by Busemeyer & Townsend (1992); we give them as Equations A.55 and A.56 in online Appendix A.2, along with asymptotic approximations A.64, and A.65 that better reveal parameter dependencies.

### 2.5. Race (inhibition-free) model

The DDM and O-U model are one-dimensional: a single integrator accumulates the net evidence. However, many models, including physiologically motivated ones, use separate integrators for the evidence in support of each alternative, and are therefore inherently two- (or more) dimensional. Here and below we review four such models, pictured in Figure 3, and describe when they can and cannot be approximately reduced to one-dimensional descriptions. We again simplify by considering only linearized models.

The simplest is the race model in which accumulators for each of the two alternatives integrate evidence independently. At least three discrete variants exist, differing in how time and evidence are discretized: the recruitment model (La Berge, 1962), the accumulator model (Vickers, 1970), and the Poisson counter model (Pike, 1966). Here we describe a continuous-time analog (see Figure 3a). We again denote the mean rates of incoming evidence by $I_1$ and $I_2$ and assume they are subject to independent white noise processes of root mean square (RMS) strength $c$. The integrators $y_1$ and $y_2$ accumulate evidence according to

$$\begin{cases}
    dy_1 = I_1 dt + c dW_1 \\
    dy_2 = I_2 dt + c dW_2
\end{cases}, \quad y_1(0) = y_2(0) = 0. \quad (2.13)$$

We model the interrogation paradigm by assuming that at time $T$ the choice is made in favor of the alternative with higher $y_i$. In free-response mode, as soon as either unit exceeds a pre-assigned threshold, the model is assumed to make a response. Again we assume equal
probabilities of the two alternatives, and hence equal decision thresholds $y_1 = y_2 = Z$ (capital $Z$ distinguishes this case from the threshold of one dimensional models, denoted by $z$). Instead of deriving ER and DT expressions for the race model here, we show how they relate to the ERs and DTs of the other models in Section 4. Anticipating those results, we note that the race model cannot be reduced to the DDM and is therefore not optimal in the sense described above.

Next we describe three biologically-motivated models that relate more closely to the DDM and O-U models than to the race model, although they also employ separate accumulators.

### 2.6. Mutual Inhibition Model

Figure 3b shows the architecture of an abstract neural network (connectionist) model described by Usher & McClelland (2001). We refer to this as the mutual inhibition model in the remainder of this article, to distinguish it from others to be discussed below. Its four units represent the mean activities of neuronal populations: two input units represent populations providing evidence in support of the alternative choices (e.g., groups of left- and right-movement sensitive MT neurons, cf. Section 2.1), and two decision units represent populations integrating the evidence (e.g., LIP neurons).

The decision units are leaky integrators with activity levels $y_i$ and each accumulates evidence from an input unit with mean activity $I_i$ and independent white noise fluctuations $dW_i$ of RMS strength $c$. The decision units also mutually inhibit each other via a connection of weight $w$.

Information is therefore accumulated according to:

\[
\begin{align*}
    dy_1 &= (-ky_1 - wy_2 + I_1)dt + cdW_1, \\
    dy_2 &= (-ky_2 - wy_1 + I_2)dt + cdW_2, \\
    y_1(0) &= y_2(0) = 0.
\end{align*}
\]  

(2.14)

Here $k$ denotes the decay rate of activity (the leak terms $-ky_i$ cause the activity to decay to zero in the absence of inputs) and $-wy_i$ represents mutual inhibition. The parameters $k$, $w$ and
those of the two models described below are all assumed to be non-negative. The activity scale is chosen so that zero represents baseline activity in the absence of inputs, hence integration starts from \( y_1(0) = y_2(0) = 0 \).

The values of \( y_1 \) and \( y_2 \) describing the state of this model may be represented as a point on a phase plane with horizontal and vertical axes \( y_1 \) and \( y_2 \), and the evolution of activities during the decision process may be visualized as a path in this plane. Figure 2c shows an example corresponding to the individual time courses of \( y_1 \) and \( y_2 \) of Figure 2b.

The major simplification of the model presented here in comparison to that of Usher & McClelland (2001) is the removal of non-linearities. In Equations 2.14 the accumulation rates of \( y_1 \) and \( y_2 \) depend linearly on their present values. To account for the fact that neural firing rates are never negative, Usher & McClelland (2001) assume that \( y_1 \) and \( y_2 \) are transformed via a threshold-linear activation function \( f(y) \) with \( f(y) = y \) for \( y \geq 0 \) and \( f(y) = 0 \) for \( y < 0 \); Equations 2.14 linearize this function by ignoring thresholding below 0. However, as noted above, in our analysis \( y_i = 0 \) corresponds to (positive) baseline activity so that \( y_i < 0 \) need not imply negative activity. As suggested by Cohen et al. (1990), this sensitive (linear) range of the activation function is precisely where one expects decision units in the focus of attention to operate. Furthermore, for suitable parameter values, Equations 2.14 yield good estimates of the ER and DT distributions of the fully nonlinear system, as shown elsewhere (Usher & McClelland, 2001; and see Brown & Holmes, 2001; Brown et al., 2005 for explicit comparisons among the dynamics of fully nonlinear, piecewise-linearized, and linearized mutual inhibition models).

### 2.7. Feedforward inhibition model

Shadlen & Newsome (2001) (cf. Ditterich et al., 2003; Mazurek et al., 2003) proposed a feedforward inhibition model for area LIP (Figure 3c). It differs from the mutual inhibition
model in that the units are perfect integrators with no leak, and they receive inhibition from “crossed” inputs rather than inhibiting each other. The accumulation rates are independent of $y_i$:

$$\begin{align*}
\frac{dy_1}{dt} &= I_1 dt + cdW_1 - u(I_2 dt + cdW_2) \\
\frac{dy_2}{dt} &= I_2 dt + cdW_2 - u(I_1 dt + cdW_1)
\end{align*}$$

where the crossed inputs imply that both noise sources influence both accumulators. Here $u$ denotes the weight of feedforward inhibitory connections: in the version of Ditterich et al. (2003) these have the same weight as excitatory connections (i.e., $u=1$). As in the mutual inhibition model, we assume that decisions are rendered whenever the activity of either unit reaches the threshold $Z$.

2.8. Pooled inhibition model

Wang (2002) developed a biophysically-based model of area LIP, simulating individual spiking neurons rather than averaged representations of neural populations as in the abstract connectionist type models described above. The neural and synaptic properties (e.g., membrane capacitance, leak conductance, etc. and decay times of AMPA and NMDA currents) were based on anatomical and neurophysiological observations.

The model, shown in Figure 3d (cf. Wang, 2002), includes two pools of neurons representing the alternative choices. These receive stimulus inputs and compete with each other as explained below. At the population level, this architecture has much in common with the mutual inhibition model (compare Figures 3b and 3d), but they differ as follows.

First, the decision neurons have self-excitatory recurrent projections, denoted by $v$ in Figure 3d. These are necessary to enable individual neurons, whose membrane voltages decay on a millisecond scale, to integrate information on the decision timescale of hundreds of milliseconds. (They also allow the system to actively maintain stimulus representations, a
property used to address short-term memory phenomena in other contexts). Second, neurons representing the alternatives do not directly inhibit each other (since real excitatory neurons do not send inhibitory connections), but both populations excite a common pool of inhibitory neurons via connections of weight $w'$, which then inhibit all decision neurons equally via recurrent connections of weight $w$.

Precise relationships between detailed neural network models such as this and the DDM are elusive, since it is not yet entirely clear when a population of neurons may be approximated by a noisy connectionist unit (although progress is being made in this direction; e.g., Ermentrout, 1994; Omurtag et al., 2000; Shelley & McLaughlin, 2002; Brunel & Latham, 2003; Renart et al., 2003; Wong & Wang, 2005; and X-J. Wang, personal communication). Here we assume that such a reduction holds, so that, denoting the activities of the decision groups by $y_1$ and $y_2$ and of the inhibitory pool by $y_3$, the model may be written as:

$$
\begin{align*}
\frac{dy_1}{dt} &= (-ky_1 - wy_3 + vy_1 + I_1)dt + cdW_1, \\
\frac{dy_2}{dt} &= (-ky_2 - wy_3 + vy_2 + I_2)dt + cdW_2, \\
\frac{dy_3}{dt} &= (-k_{inh}y_3 + w'(y_1 + y_2))dt.
\end{align*}
$$

Again, as above (but unlike Wang), we have linearized all terms in these equations. As we show in Section 3, this “reduced” or “averaged” network version of Wang’s (2002) model may be viewed as a more biologically realistic implementation of the Usher-McClelland (2001) mutual inhibition model.

---

5 Some of the work referenced above indicates that the averaged synaptic conductances are the relevant dynamical variables for the reduced network equations; here, for consistency with the rest of the article, we express (2.16) in terms of the activities of the underlying populations (Wilson & Cowan, 1972).
3. Relationships among the models

A number of previous reports have noted similarities between the DDM and biologically-inspired connectionist and “averaged” neural network models (e.g., Usher & McClelland, 2001; Wang, 2002; Seung, 2003; Ratcliff & Smith, 2004). In this section we extend those observations by conducting a formal analysis of the relationships among the models, and we derive precise correspondences among their variables and parameters. The relationships we find are summarized in Figure 4, which shows that, for appropriate parameter ranges, all the models described in Section 2 except the race model can be reduced to the optimal DDM.

Recently, Ratcliff & Smith (2004) studied the specific relationships among the DDM, O-U, race, and mutual inhibition models. They used a “model mimicry” approach, in which they generated data by simulating one model, fitting a second to these data, and measuring the closeness of this fit. Our study complements these previous efforts. While our analytical approach limits us to simplified models, it allows us to explain why, and for what parameter ranges, certain models are equivalent.

The relationships we find, summarized in Figure 4, in many cases are similar to those that have been revealed by other approaches. For example, similar to Ratcliff & Smith (2004), we find that: the DDM and the O-U model with $\lambda$ close to 0 mimic each other (as designated by the connection between these models in Figure 4 with label $\lambda = 0$), the DDM does not mimic the race model (there is no connection between them in Figure 4), and the mutual inhibition model can mimic the DDM, O-U and race models for different parameter values (arrows pass from the mutual inhibition model to the DDM, O-U and race models in Figure 4).

In the remainder of this section, we describe analyses of the specific relationships among models. In several instances, we present illustrative simulations using parameters fit to TAFC behavioral data. Therefore we begin, in Section 3.1, with a description of a fit of the DDM to
such data. Section 3.2 shows that for particular parameter ranges, the mutual inhibition model can be approximated by either the O-U model or the DDM. Section 3.3 describes the relationship between the feedforward inhibition model and DDM. Section 3.4 shows that the pooled inhibition model can also be approximated by the DDM for certain parameter values, but that there are subtle differences in interpretation of the role of inhibition.

3.1. TAFC experiment and fit of DDM

To illustrate our analysis and verify its relevance, we must choose representative parameter values. We do this using data from an experiment examining human performance in a TAFC similar to the one used by Shadlen & Newsome (2001), and identifying the parameters of the DDM that best fit the behavior of a representative participant (with mean ER and RT that were within one standard deviation of the means of full set of participants). The full set of data from this experiment is available for download at: http://www.cs.bris.ac.uk/home/rafal/optimal/data.html. A complete description of the experiment will be the subject of a future report. Here, we briefly describe the experimental methods (a more detailed description of the experimental methods is provided in Appendix B) and our methods of parameter estimation for the DDM, after which we return to the comparative analysis of models.

3.1.1. Method

Twenty adult participants performed a task in which they had to indicate whether the predominant direction of movement of an array of dots on a display screen was leftward or rightward by pressing a corresponding key. Task difficulty (i.e., motion coherence) was kept constant throughout the experiment. Participants were paid 1 cent for each correct choice. After each response participants were required to release the key, and there was a delay $D$ before presentation of the next stimulus ($D$ was kept constant within each block, but varied
across blocks). On some blocks an additional delay $D_p$ was imposed after error responses (see below). Trials were blocked by delay condition. This was manipulated to test predictions made by the DDM (discussed below in Section 5.2). There were four delay conditions: (1) $D=0.5s$; (2) $D=1s$; (3) $D=2s$; and (4) $D=0.5s$ and $D_p=1.5s$ (in the first three conditions $D_p=0$).

3.1.2. Estimating parameters of the DDM

We estimated the parameters of the extended DDM using a combination of the weighted least squares fitting method (Ratcliff & Tuerlinckx, 2002) and the method of Bogacz & Cohen (2004). Since we expected participants to adopt different decision criteria in different delay conditions (as described in Section 5), we assumed that thresholds $z$ differed across conditions. We further assumed that all the other parameters of the DDM were the same for a given participant in all conditions, since the stimuli were exactly the same in all blocks (the motion of dots had the same coherence). Thus, for each participant we estimated the following parameters: $m_A$ (mean drift rate), $s_A$ (standard deviation of the drift rate), $c$ (noise), $s_x$ (standard deviation of the starting point), $T_0$ (non-decision related response time) and $z_1, z_2, z_3, z_4$ (decision thresholds for the four delay conditions).

Following Ratcliff & Tuerlinckx (2002), for each delay condition $d$ we divided the RT distributions for correct and error trials into five quantiles $q$: 0.1, 0.3, 0.5, 0.7, 0.9, denoted as $RTC_{q,d}^{th}$, $RTE_{q,d}^{th}$ and $RTC_{q,d}^{ex}$, $RTE_{q,d}^{ex}$, with superscripts $th$ and $ex$ indicating theoretical (extended DDM) and experimental statistics. We denote the error rates given by the extended DDM and observed in the experiment by $ER_d^{th}$ and $ER_d^{ex}$ respectively.

The subplex optimization algorithm (Rowan, 1990) was used to find parameters minimizing the cost function describing the weighted difference between ERs and RT distributions for the model and from the experiment (Ratcliff & Tuerlinckx, 2002):
$$Cost = \sum_{d=1}^{4} \alpha_{ER,d} \left( ER_d^{th} - ER_d^{ex} \right)^2 + \sum_{q=1}^{5} \alpha_{RTC,d,q} \left( RTC_d^{th} - RTC_d^{ex} \right)^2 +$$

$$+ \sum_{q=1}^{5} \alpha_{RTE,d,q} \left( RTE_d^{th} - RTE_d^{ex} \right)^2$$

(3.1)

In the above equation, $\omega$’s denote the weights of the fitted statistics. We would like to choose the weight of a given statistic close to $1 / (\text{the estimated variance of this experimental statistic})$, as described by Bogacz & Cohen (2004). For ER, this can be estimated as follows: Assume, for a given condition, that a participant has a probability of making an error equal to $p_e$ and there are $n$ trials in this condition. Then the experimental error rate $ER^{ex}$ comes from the binomial distribution with mean $p_e$ and variance $p_e(1-p_e)/n$. We take $\omega_{ER} = n^{p_e(1-p_e)}$. For simplicity we estimate $p_e$ as the mean ER of the participant averaged across all delay conditions. This averaging across conditions is done to avoid dividing by 0 in blocks in which the participant did not make any errors, and also because the differences in $ER^{ex}$ across conditions for single participants are small in comparison to differences between participants.

The estimation of variance for quantiles of the RT distribution was computed using the method of Maritz & Jarrett (1978).

Figure 5 compares the ER and RT quantiles of the extended DDM fitted to experimental data from a sample participant by plotting Quantile Probability Functions, as in Ratcliff & Smith (2004). The decision process of this participant exhibited both variability of drift and starting point (estimated parameter values: $m_A=1, s_A=0.31, s_X=0.14, c=0.33$). The extended DDM fits the essential properties of the data: In Figure 5a, the participant emphasised speed over accuracy; thus the threshold was low ($z_1=0.16$, just outside the range of starting points $s_x=0.14$) and the effect of variability of starting point was prevalent, producing shorter RTs for errors than correct responses (Ratcliff & Rouder, 1998). In Figure 5d, the participant emphasised accuracy over speed; thus the threshold was high ($z_4=0.26$) and the effect of
variability of drift was prevalent, producing longer RTs for errors than for correct responses (Ratcliff & Rouder, 1998).

### 3.2. Mutual inhibition, O-U and DDM

Usher & McClelland (2001) observed that the dynamics of the difference \( y_1 - y_2 \) between activities of the decision units in the (linearized) mutual inhibition model is described by the O-U model in general and, for certain “balanced” parameters, by the DDM as well. Here, we provide an analytic interpretation of these observations and establish that when the decay and inhibition parameters of the mutual inhibition model are equal and large, the mutual inhibition, O-U and DDM models become equivalent.

#### 3.2.1. Reduction of the mutual inhibition model to the pure DDM

As illustrated in Section 2.3, the behavior of the mutual inhibition model may be visualized by plotting states on the phase plane. Figure 2c shows a representative path in state space: initially the activities of both decision units increase due to stimulus onset, but as the units become more active, mutual inhibition causes the activity of the ‘weaker’ unit to decrease and the path moves toward the threshold for the more strongly activated unit (i.e., the correct choice).

To understand how these dynamics of the mutual inhibition model are influenced by its two critical parameters (decay and inhibition), Figure 6 shows its vector fields for three different ranges of these parameters. Each arrow shows the average direction in which the state moves from the point indicated by the arrow’s tail, and its length corresponds to the speed of movement (i.e., rate of change) in the absence of noise. In Figure 6, as for most other simulations described in this article, we set \( I_1 > I_2 \), that is, we assume that the first alternative is the correct one (the opposite case is obtained simply by reflecting about the diagonal \( y_1 = y_2 \)).
Note that in all three panels of Figure 6 there is a line (an eigenvector), sloping down and to the right, to which system states are attracted: The arrows point towards this line from both sides. The orientation of this line represents an important quantity: the difference in activity between the two decision units. Note that the evolution along the line differs for different values of decay and inhibition, as does the strength of attraction toward the line, and its location in the phase plane. Most of the interesting dynamics determining decisions occur along this line, and hence we refer to it as the decision line. Therefore, it is easier to understand these in terms of new coordinates rotated clockwise by 45° with respect to the $y_1$ and $y_2$ coordinates, so that one of the new axes is parallel to the decision line. These new coordinates are shown in Figures 6b and 7a, denoted by $x_1$ (parallel to the decision line) and $x_2$ (perpendicular to it). The transformation from $y$ to $x$ coordinates is given by (cf. Grossberg, 1988; Seung, 2003):

$$\begin{align*}
    x_1 &= \frac{y_1 - y_2}{\sqrt{2}}, \\
    x_2 &= \frac{y_1 + y_2}{\sqrt{2}}.
\end{align*}$$

Equations 3.2 derive from the geometry shown in Figure 7a: $x_1$ describes the difference between activities of the two decision units, while $x_2$ describes the sum of their activities. The square root of two in the denominators of Equations 3.2 is a normalization factor, included to ensure that $y$ and $x$ coordinates have the same scale.

In deciding between two alternatives, it is natural that the difference between the activities of the units selective for the alternatives should be a useful descriptor of the decision process. However, the new coordinates do more than merely emphasize this point. They allow us to factor the two Equations 2.14 that describe the mutual inhibition decision process into two decoupled processes, separating the evolution of the difference in the activity of the two units ($x_1$) from the change in their overall (summed) activity ($x_2$). If we can show that the latter has
minimal impact on the decision process, then we can reduce the description of this process from one that is two-dimensional to a simpler one that is one-dimensional. As we will show, for certain parameters this one-dimensional description reduces to the DDM.

To transform Equations 2.14 into the new coordinates, we first calculate the derivative (rate of change) of \( x_1 \). Substituting Equations 2.14 into the first of Equations 3.2, we obtain:

\[
dx_1 = \frac{dy_1 - dy_2}{\sqrt{2}} = \left( -k \frac{y_1 - y_2}{\sqrt{2}} + w \frac{y_1 - y_2}{\sqrt{2}} + \frac{I_1 - I_2}{\sqrt{2}} \right) dt + \frac{1}{\sqrt{2}} (cdW_1 - cdW_2). \tag{3.3}
\]

We assumed earlier that the noise processes for the input units are independent. Since the standard deviation of the sum (or difference) of two independent random variables is equal to the square root of the sum of their variances, the noise process in \( x_1 \) may be written:

\[
\frac{1}{\sqrt{2}} (cdW_1 - cdW_2) = \frac{\sqrt{c^2 + c'^2}}{\sqrt{2}} dW_v = cdW_v. \tag{3.4}
\]

In Equation 3.4, \( dW_v \) again denotes a noise process with mean equal to 0 and an RMS strength of 1. Substituting Equation 3.4 and the definition of \( x_1 \) from Equation 3.2 into Equation 3.3, we obtain Equation 3.5. Following analogous calculations for \( x_2 \), we have:

\[
dx_1 = \left( (w-k)x_1 + \frac{I_1 - I_2}{\sqrt{2}} \right) dt + cdW_v, \tag{3.5}
\]

\[
dx_2 = \left( (-k-w)x_2 + \frac{I_1 + I_2}{\sqrt{2}} \right) dt + cdW_v. \tag{3.6}
\]

Equations 3.5 and 3.6 are uncoupled; that is, the rate of change of each \( x_i \) depends only on \( x_i \) itself (this was not the case for the decision units in Equations 2.14). Hence, the evolution of \( x_1 \) and \( x_2 \) may be analyzed separately, and in fact each is described by an O-U process that depends on the parameters of \( k \) (decay) and \( w \) (inhibition).

We first consider the dynamics in the \( x_2 \) direction, corresponding to the summed activity of the two decision units. Equation 3.6 for the \( x_2 \) process always gives a stable O-U process since...
\(-k - w < 0\) (recall that \(k\) and \(w\) are both constrained always to be non-negative). This corresponds to attraction to the line in all of the panels in Figure 6, implying that \(x_2\) approaches a limiting value as time increases. The rate of this (exponential) approach is \(-\lambda_2 = k + w\), and it is kept constant in the three cases of Figure 6 by setting \(k + w = 2\). From Equation 2.11, the stationary standard deviation from the decision line in the \(x_2\) direction is equal to:

\[
\text{stdev}(x_2) = \frac{c}{\sqrt{2(k + w)}}.
\] (3.7)

Turning to the \(x_1\) process, Equation 3.5 involves a drift term that is proportional to the difference between the inputs \(I_1\) and \(I_2\). As we discussed in Section 2.4, this process may be stable or unstable, depending upon the relative magnitudes of \(k\) and \(w\), in a manner comparable to the effects of the parameter \(\lambda\) in the O-U model. This dependence is due to the fact that the dynamics of \(x_1\) are described in Equation 3.5 by an O-U process with coefficient \(\lambda = w - k\). When decay is larger than inhibition, then \(\lambda < 0\), and there is an attractor for the \(x_1\) dynamics (see Section 2.4), as shown in Figure 6a. When decay is smaller than inhibition, then \(\lambda > 0\), and there is repulsion from the fixed point in the \(x_1\) direction, as shown in Figure 6c. The fixed point is a saddle in this case.

Since \(|k+w|\) (for \(x_2\) in Equation 3.6) > \(|w-k|\) (for \(x_1\) in Equation 3.5) for all positive values of \(k\) and \(w\), the average state of the system approaches the decision line faster (and often considerably faster) than it moves along it (e.g., see Figure 2c). Moreover, if the denominator \(\sqrt{2(k + w)}\) in Equation 3.7 is large relative to noise strength \(c\), the states/solutions remain tightly bunched around this line as they continue to evolve. This is an important observation, as it indicates that the decision process divides into two phases: an initial phase in which the
activity of both units increases quickly, and there is rapid equilibration to a neighborhood around the decision line; followed by slower movement along the line, governed by an O-U process in which the difference between the activities of the two units grows as one of them prevails and the other subsides.

Most relevant to the current discussion, when decay equals inhibition, the term \((w - k) x_1\) in Equation 3.5 disappears. In this case, the dynamics of \(x_1\) reduce to the pure DDM (cf. Equation 2.5). The vector field for this case is shown in Figure 6b. When decay and inhibition are both reasonably strong (as in Figure 6b), the attraction toward the decision line dominates diffusion along it. Hence, typical paths migrate quickly toward the decision line and then move relatively slowly along (or near) it. In this case, the dynamics of the two-dimensional decision process closely approximate those of the one dimensional DDM (see also Brown et al., 2005).

Solutions to the full two-dimensional system (Equations 3.5-3.6) are described by the joint probability density of \(x_1\) and \(x_2\). Since these are independent processes (neglecting threshold effects), this is simply the product of the two one-dimensional Gaussian densities for the O-U processes \(x_1\) and \(x_2\). Critically, when decay equals inhibition and both assume sufficiently large values relative to noise strength, the mutual inhibition model’s behavior should closely resemble that of the pure DDM. We will refer to a mutual inhibition model in which decay is equal to inhibition as balanced.

3.2.2. Relating parameters of the mutual inhibition model to the pure DDM

In this section, we find the parameters describing the diffusion process along the decision line in the mutual inhibition model. As above, the transformation is geometric, and so requires the same simple trigonometry used above. From Equation 3.5 we can see that the RMS noise of the approximating DDM is equal to \(c\), and the drift is given by:
Calculation of effective thresholds for the diffusion process requires a little more work. If the density of solutions is sufficiently tight about the decision line, we need only find the points where this line reaches the thresholds $y_1 = Z$ and $y_2 = Z$ (D and C in Figure 7b). More specifically, we need to find the distances $\pm z$ to those points from the intersection of the diagonal ($x_1 = 0$) with the decision line at point A. Since the diagonal and decision line are oriented at $\pm 45^\circ$ to the $y_2$ axis, the triangles ABC and ABD in Figure 7b are isosceles and the distances AC, AD and AB are all equal to $z$. The decision line is located at the equilibrium for the stable O-U process:

$$x_2^* = \frac{I_1 + I_2}{\sqrt{2(k+w)}}.$$  (3.9)

Thus, again from Figure 7b and using the right isosceles triangle OBZ, the length of AB is equal to the difference of the lengths OB and OA, or

$$z = \sqrt{2Z } - \frac{I_1 + I_2}{\sqrt{2(k+w)}}.$$  (3.10)

### 3.2.3. Relating parameters of the mutual inhibition model to the extended DDM

The extended DDM includes variability of drift and starting point across trials. In order to relate this model to the mutual inhibition model, these two types of variability need to be incorporated into the latter. Therefore, we define the extended mutual inhibition model in which the mean inputs to the integrators $I_i$ are chosen on each trial from a normal distribution with mean $m_{II}$ and standard deviation $s_I$ and the starting points of variables $y_i(0)$ are drawn from a uniform distribution between $-s_y$ and $s_y$. 

\[ A = \frac{I_1 - I_2}{\sqrt{2}}. \] (3.8)
First, let us consider the case of statistically independent $I_1$, $I_2$, $y_1(0)$, and $y_2(0)$. In this case the sum $I_1+I_2$ differs from trial to trial, and since this sum enters Equation 3.10 relating the thresholds of the mutual inhibition model and pure DDM, there is no deterministic relationship between the thresholds of the extended mutual inhibition model and DDM (note however, that if $k, w \to \infty$, the second term in Equation 3.10 disappears, and the relationship simply becomes $z = \sqrt{2} Z$). Furthermore, the distribution of starting points $x_1(0)$ projected onto the decision line $x_2$ has a triangular distribution (since it is given by the sum of two independent, uniformly distributed random variables), so the diffusion along the decision line cannot be described by the extended DDM, which assumes uniformly distributed initial states.

It is, however, straightforward to establish the equivalence between the extended versions of the mutual inhibition model and DDM if we make the following additional assumptions: $I_1+I_2 = m_{I1} + m_{I2} = \text{constant}$, and $y_1(0)+y_2(0) = 0$. The first assumption states that the total input to the integrators is constant across trials (as in the Poisson counter model of Smith & Van Zandt, 2000). Although this assumption may not be satisfied exactly, adaptation in sensory cortices (Blakemore & Campbell, 1969) is known to reduce the neuronal response if the stimulus is strong and enhance it when the stimulus is weak. The second assumption states that the starting points of the two integrators are anti-correlated. Although as yet there is no physiological evidence for such anti-correlation in cortical populations, there are indications of this in the superior colliculus (Roger Ratcliff, personal communication).

Under the above assumptions, $I_2 = m_{I1} + m_{I2} - I_1$, and hence the drift along the decision line is equal to:

$$ A = \frac{I_1 - I_2}{\sqrt{2}} = \frac{2I_1 - m_{I1} - m_{I2}}{\sqrt{2}}. \quad (3.11) $$

Therefore, the relationships between the parameters are
and the relationship between the thresholds is

\[ z = \sqrt{2Z} - \frac{m_{t1} + m_{t2}}{\sqrt{2(k + w)}}. \]  

Similarly, under the above assumptions, the starting point of the diffusion along the decision line is:

\[ x_1(0) = \frac{y_1(0) - y_2(0)}{\sqrt{2}} = \sqrt{2}y_1(0). \]  

Therefore the relationship between parameters \( s_x \) and \( s_y \) is: \( s_x = \sqrt{2} s_y. \)

The above relationships are verified in the simulations described in the next section.

3.2.4. ERs and DTs in the balanced mutual inhibition model and DDM.

As argued above, the balanced mutual inhibition model resembles the DDM. However, these models are not identical, because the mutual inhibition model describes a two-dimensional process, including movement along the \( x_2 \) axis (corresponding to the summed activity level of the decision units), while the DDM describes a one dimensional process involving movement only along the \( x_1 \) axis (corresponding to the difference in activity levels). As a way of assessing how well the DDM approximates the balanced mutual inhibition model, we compare ERs and DTs produced by the extended versions of each model when their parameters are constrained according to the relationships derived in Section 3.2.3 and fit to experimental data described in Section 3.1.

As anticipated, the approximation of the extended mutual inhibition model by the extended DDM is more accurate for larger decay and inhibition \((k + w)\). This is due to the fact that as the parameter \( \lambda_2 \) \((- (k + w); \) see Equation 3.6) becomes more negative, attraction along the
$x_2$ direction toward the decision line is faster, and solutions become more tightly distributed around it (via Equation 3.7). This is evident in Figure 8, which compares the ER and DT of the extended versions of the mutual inhibition model and DDM (with equivalent parameters). Figure 8a illustrates the case of high decay and inhibition, in which the solutions are tightly distributed about the decision line, and ER and DT of the two models are almost exactly the same. In the case of lower decay and inhibition shown in Figure 8b, the solutions are more broadly distributed about the decision line, leading to greater discrepancy between the ER and DT of the two models. Figure 8c illustrates the case in which the decay and inhibition are further decreased so that most solutions do not reach the decision line before exceeding the threshold, resulting in substantially different ER and DT for the two models.

Two additional points are worth noting with respect to these results. First, movement towards the decision line may be facilitated by the high transient activity of sensory “input” neurons observed just after stimulus onset, as illustrated in Figure 1b (Britten et al., 1993; Schall, 2001). Smith (1995) also suggested the existence of the transient on the basis of psychophysical data. Second, if decay and inhibition are very low (or the signal is very large), then according to Equation 3.10 the threshold of the DDM may become negative, a condition for which our reduction to one dimension is ill-defined.

In summary, when decay and in particular inhibition in the mutual inhibition model are very low, there is little competition between the two decision units: They integrate evidence essentially independently, and hence the mutual inhibition model cannot be approximated by the DDM. However, when decay and inhibition are greater, the balanced mutual inhibition model can be approximated by diffusion along the decision line, as described by the DDM.
3.3. Feedforward inhibition model and DDM

In this section we show that the feedforward inhibition model reduces to DDM when the parameter $u$ describing the weight of feedforward inhibition is equal to $u=1$.

In order to analyze the dynamics of the feedforward inhibition model, we convert it to the rotated coordinates as we did the mutual inhibition model. Transforming Equation 2.15 via Equation 3.2 (as in Section 3.1) we obtain:

$$
\begin{align*}
\dot{x}_1 &= (1+u) \left( \frac{I_1-I_2}{\sqrt{2}} dt + cdW_t \right), \\
\dot{x}_2 &= (1-u) \left( \frac{I_1+I_2}{\sqrt{2}} dt + cdW_t' \right),
\end{align*}
$$

(3.15)

Thus, similar to the mutual inhibition model, the dynamics of the feedforward inhibition model in the rotated coordinates can be described by two uncoupled diffusion processes weighted by factors $1+u$ and $1-u$. When we defined the mutual and feedforward inhibition models we implicitly assumed that the weight of connections from input units to the integrating (decision) units is equal to 1. Note that as the weight of the inhibitory connection ($u$) approaches the weight of the excitatory connection (1), the weighting factor $1-u$ diminishes, and thus dynamics become slower along the direction of $x_2$ relative to the direction of $x_1$. However, unlike the case of the mutual inhibition model, in the feedforward inhibition model given by Equation (3.10) there is no attraction toward the decision line. Thus, approximation by the one dimensional DDM depends on the mean and variance in the $x_2$ direction remaining negligible for the duration of the decision task. If the excitatory and inhibitory weights are equal ($u=1$, as they are in the model described by Ditterich et al., 2003), then the value of coordinate $x_2$ remains equal to 0 throughout the duration of the decision process and the entire process proceeds along the decision line $x_1$. In other words, if $u=1$, the feedforward inhibition model is exactly equivalent to the pure DDM. In this case, diffusion
occurs along the line shown in Figure 7c and the relationships between the parameters of the
pure DDM and the feedforward inhibition model are the following:

\[ A \leftarrow \frac{2I_1 - I_2}{\sqrt{2}}, \ c \leftarrow 2c, \ z \leftarrow \sqrt{2Z}. \]  

(3.16)

where the left side terms are for the DDM, and the right side terms are for the feedforward
inhibition model. The relationships for the drift and noise terms come from the first of
Equations 3.15, and for the threshold from Figure 7c.

As in the previous section we can define an extended feedforward inhibition model, where we
also assume \( I_1 + I_2 = \text{constant} \), and \( y_1(0) + y_2(0) = 0 \). For \( u=1 \), the relationships between the
parameters of the extended versions of the feedforward inhibition model and DDM are the
following:

\[ m_A \leftarrow \frac{2m_{11} - m_{12}}{\sqrt{2}}, \ s_A \leftarrow \sqrt{2}s_j, \ s_A \leftarrow \sqrt{2}s_j, \ c \leftarrow 2c, \ z \leftarrow \sqrt{2Z}. \]  

(3.17)

Busemeyer & Diederich (2002) considered the following model combining mutual and
feedforward inhibition:

\[
\begin{bmatrix}
    dy_1 \\
    dy_2
\end{bmatrix} = \begin{bmatrix}
    -k & -w \\
    -w & -k
\end{bmatrix} \begin{bmatrix}
    y_1 \, dt \\
    y_2 \, dt
\end{bmatrix} + \begin{bmatrix}
    1 & -u \\
    -u & 1
\end{bmatrix} \begin{bmatrix}
    I_1 \, dt + cdW_1 \\
    I_2 \, dt + cdW_2
\end{bmatrix}
\]  

(3.18)

They have shown that as \( u=1 \), the above model can be related to the O-U process (Busemeyer
& Diederich, 2002; Diederich & Busemeyer, 2003). If such a combined model were balanced
\( k = w \), then as \( k \) and \( w \) increased and \( u \) approached 1, this combined model would be also
increasingly well approximated by the pure DDM.

---

\(^6\) The first square matrix in the Equation 3.18 was denoted by \( S \) and the second square matrix by \( C \) in Busemeyer
& Diederich (2002).
3.4. Mutual inhibition and pooled inhibition models

In this section we show that a connectionist model with pooled inhibitory connections can also be approximated by the pure DDM. Specifically, we consider a connectionist model with the architecture shown in Figure 3d, the dynamics of which are described by the following three (linearized) equations (we rewrite Equations 2.16):

\[
\begin{align*}
    dy_1 &= (-ky_1 - wy_3 + vy_1 + I_1)dt + cdW_1, \\
    dy_2 &= (-ky_2 - wy_3 + vy_2 + I_2)dt + cdW_2, \\
    dy_3 &= (-k_{inh}y_3 + w'(y_1 + y_2))dt.
\end{align*}
\]

In the above equation \(k_{inh}\) denotes the decay rate with which the pool of inhibitory neurons approaches firing rates determined by the excitatory populations. If we assume that this rate is fast relative to that of the excitatory populations\(^7\) (the effective decay rate of the excitatory population is decreased by the self-excitatoty connections, thus this assumption is equivalent to \(k_{inh} >> k - v\)), then inhibitory neurons may be said to have “little memory of their previous inputs,” closely tracking and thus approximating the value:

\[
y_3 = \frac{w'}{k_{inh}}(y_1 + y_2).
\]

\(^7\) The assumption that the decay rate \(k_{inh}\) is relatively fast is plausible, but will require validation in further physiological experiments. This rate depends on the dynamics of synaptic currents afferent to the inhibitory population. Although in the original model of Wang (2002), inhibitory neurons received a significant proportion of their input via NMDA receptors (described by relatively slow time constants of approximately 100ms (Wang, 2002)), the density of such receptors remains to be determined in area LIP. While current research indicates that inhibitory neurons in the hippocampus do possess NMDA receptors (McBain & Dugledine, 1993), Ling & Benardo (1995) have shown that the inhibition in the rat somatosensory cortex is not NMDA dependent. A parallel finding for the LIP inhibitory neurons, showing that they are excited via rapidly decaying AMPA receptors, would support the present assumption.
Substituting Equation 3.20 into 3.19 and regrouping terms, we obtain:

\[
\begin{align*}
    dy_1 &= \left( -\left( k + \frac{ww'}{k_{inh}} - v \right) \right) y_1 - \frac{ww'}{k_{inh}} y_2 + I_1 \right) dt + cdW_1, \\
    dy_2 &= \left( -\left( k + \frac{ww'}{k_{inh}} - v \right) \right) y_2 - \frac{ww'}{k_{inh}} y_1 + I_2 \right) dt + cdW_2.
\end{align*}
\]  

(3.21)

A comparison of Equations 3.21 and 2.14 reveals that the connectionist model with the architecture of Figure 3d is equivalent to the mutual inhibition model shown in Figure 3b with decay parameter equal to \((k + \frac{ww'}{k_{inh}} - v)\) and mutual inhibition \(\frac{ww'}{k_{inh}}\). Therefore, pooled and rapidly equilibrated inhibition provides not only competition but also contributes to the decay of each unit’s activity. Hence, given the considerations of Section 3.1, the connectionist model with the architecture shown in Figure 3d displays the same dynamics in the \(x_1\) direction as the pure DDM when the following balance among parameters holds:

\[ k + \frac{ww'}{k_{inh}} - v = \frac{ww'}{k_{inh}} \quad \text{or} \quad k = v. \]  

(3.22)

The dynamics in the \(x_2\) direction collapse to the decision line, exactly as for the mutual inhibition model, with a timescale proportional to \(\frac{k_{inh}}{ww'}\).

To summarize, the dynamics of the connectionist model with pooled inhibition can be approximated by the pure DDM when self-excitation exactly balances the decay of each decision unit’s activity (i.e., \(k=v\)), inhibitory strengths \((w, w')\) are sufficiently high, and the decay of inhibitory units \(k_{inh}\) is sufficiently large. Furthermore, paralleling the arguments of the previous section, an extended pooled inhibition model may be defined and related to the
We do not consider the pooled inhibition model further in view of its close relationship to the mutual inhibition model (for large $k_{\text{inh}}$, cf. Equations 3.21).

4. Optimal performance of decision networks

In this section we identify parameter values that allow decision networks to achieve optimal performance in the interrogation and free-response paradigms (Sections 4.1 and 4.2, respectively). Section 4.3 investigates whether the ERs in decision models can be made arbitrarily small if a sufficiently long time is allowed for the decision in the interrogation paradigm, or the threshold is made sufficiently high in the free-response paradigm. The results of this section for the mutual inhibition model are summarized in Table 1. In particular, we show that for the parameters for which decision networks reduce to DDM, they achieve optimal performance in both paradigms.

4.1. Highest accuracy for fixed DTs: The interrogation paradigm

As described in the Introduction, the pure DDM is also the optimal decision maker (i.e., is equivalent to the Neyman-Pearson test) in the interrogation paradigm; that is, it is the most accurate for a fixed DT. In this section we show that the mutual inhibition model achieves such optimality when it is balanced.

It is worth noting here that, in the interrogation paradigm, the ER of the mutual inhibition model does not depend on the magnitude of decay and inhibition, but only on their relative values. As we will see below, this contrasts with the free-response paradigm, in which

\[^8\] The more detailed firing rate model of Wong & Wang (2006), which includes specific synaptic time constants, indicates that Equation 3.22 may sometimes be violated. Depending on the overall stimulus strength, on which the analogs of $k$, $w$, etc., depend, both the stable and unstable O-U cases of Figure 6 can occur.
performance depends additionally on the absolute magnitude of decay and inhibition. The reason is that in the interrogation paradigm decisions are based solely on a comparison of activities of the decision units, and hence only on the value of the $x_1$ coordinate. In effect, the interrogation paradigm allows the two-dimensional probability density to evolve and then, at the time of interrogation, cuts it along the $x_2$ axis and sums it in the negative (error) and positive (correct) $x_1$ half-planes. Therefore, the density of solutions about the decision line (that tightens under increased decay and inhibition) is irrelevant. As a consequence, in contrast to free-response, in the interrogation paradigm the balanced mutual inhibition model achieves the same ER as the DDM irrespective of the magnitudes of decay and inhibition. This is true even in the extreme case of $k = w = 0$, corresponding to the race model in which the decision units integrate independently of one another. Similar conclusions hold for the feedforward inhibition model (which minimizes ER in the interrogation paradigm for any value of $u$) and the pooled inhibition model (if the dynamics in the $x_1$ direction are balanced by setting $k=v$, for any values of $w, w', k_{inh}$.) Hence, the mutual inhibition model can in general be approximated by an O-U process with parameter $\lambda$ equal to the difference between inhibition and decay (i.e., $\lambda = w - k$), and in the case of the interrogation paradigm these are identical. Therefore, we focus exclusively on the O-U model and ask what value of $\lambda$ optimizes performance.

First we analyze the pure O-U model (without variability of drift or staring point). In online Appendix A.3, we calculate explicit expressions for the ERs for different interrogation times $T$, and parameters $A, c$ and $\lambda$, as already given above in Equation 2.12. These verify that the O-U process achieves its lowest ER for a fixed DT when $\lambda = 0$ (i.e., when it simplifies to the DDM). This approximates the mutual inhibition model with decay equal to inhibition. Figure 9a shows ERs achieved by the O-U process with different values of $\lambda$ and different
interrogation times. Usher & McClelland (2001) noted that ERs depend only upon the absolute magnitude of $\lambda$ (e.g., they are the same for $\lambda = 0.2$ and $-0.2$, and for 0.4 and $-0.4$), and that the lowest ERs are achieved when $\lambda = 0$. These observations are consistent with the analyses of online Appendix A.3.1. In particular, the fact that the lowest ER in the interrogation paradigm is achieved by an O-U process with $\lambda = 0$ confirms the earlier observation that the mutual inhibition model achieves the lowest ER when it is balanced and approximates the DDM.

We now consider the extended O-U model, in which drift rate and/or starting point are allowed to vary from trial-to-trial. Online Appendix A.4.2 calculates the ER of the O-U model with variability of drift but not starting point, and shows that for any $m_A$, $s_A$, $c$, and interrogation time $T$, the ER is minimized for $\lambda = 0$. Thus, the mutual inhibition model with variability of inputs (but not starting points) achieves the lowest ER in the interrogation paradigm when it is balanced ($k=w$).

In contrast with variable drift rate, simulations of Figure 9c indicate that when starting point is allowed to vary, optimal performance of the O-U model (i.e., minimization of ER) may be achieved for values of $\lambda$ lower than zero. This may be intuitively understood as follows: values $\lambda < 0$ introduce a recency effect (with recently arrived inputs influencing accumulated evidence more than previous inputs or biases; see Section 2.4), which reduces the influence of the random starting point on the decision that is eventually made. This suggests that, when starting point varies, unbalanced versions of the mutual inhibition model may be optimal.

However, this observation needs to be considered with respect to the psychological interpretation of variable starting points. Typically, starting point is thought to reflect the influence of prior stimulus or response history on current processing, as a result of priming effects and/or expectations based on estimations of stimulus and response probabilities (e.g.,
Cho et al., 2002). In Section 6 we demonstrate that variations in the starting point are in fact optimal for all models when these variations appropriately track systematic changes in stimulus frequency. There, we will return to a discussion of how to interpret the effects of variable starting point in fits of the DDM and related models to empirical data.

The analyses and simulations described here provide insight into the parameters of the mutual inhibition model that produce optimal performance in the interrogation paradigm under various conditions. However, they do not explain how a biological system might actually make the decision at the required time in the interrogation paradigm; that is, how the system determines the sign of $x_1$ when it has not yet crossed a specified threshold. A number of mechanisms for doing so can be imagined. For example, this could be accomplished by setting a very high value of the decision threshold during the trial and then lowering it rapidly, or suddenly increasing $\lambda$ (perhaps via $w$), which would cause strong repulsion from the fixed point and thus rapid movement toward one of the thresholds used in the free-response paradigm. Such weight adjustments could be achieved by a transient increase in gain, which might in turn be mediated by release of neuromodulatory neurotransmitters. The locus coeruleus has been suggested as a possible element in such a ‘control circuit’ (Usher et al., 1999; Brown et al., 2004, 2005; Gilzenrat et al., 2004; Aston-Jones & Cohen, 2005), but further consideration of such mechanisms is beyond the scope of this article.

4.2. Fastest decisions for fixed accuracy: The free-response paradigm

As described above, the pure DDM achieves the fastest DTs for a given level of accuracy, and in this sense it is the optimal decision process for the free-response paradigm. In Section 3.1 we established that the mutual inhibition model’s behavior is similar to that of the DDM when decay equals inhibition and both assume a sufficiently large value. Usher & McClelland (1995, footnote on p. 22) noted that their model will also achieve its shortest DTs for fixed
accuracy when decay equals inhibition — that is, when it best approximates the DDM. Similar observations were made for the feedforward inhibition model. In this section we show that these relationships still hold when drift rate is allowed to vary across trials in these models.

In particular, we conjecture that when drift rate is allowed to vary but starting point is held constant, performance of the mutual inhibition model will be optimized when decay is equal to inhibition and both take large values (that is, when it approximates the DDM). To test this, we simulated the mutual inhibition model with different values of decay $k$, while all other parameters were held constant. Specifically, the weight of inhibition$^9$ was set to $w = 10$, the range of starting points was set to $s_y=0$, and the remaining parameters were set by fitting the extended DDM to data from a sample participant (see Figure 5) and then determining equivalent parameters for the extended mutual inhibition model as described in Section 3.2. For each value of decay, the threshold was set such that the ER of the network was 2%. Figure 10a shows the DTs produced by this model. As expected, the shortest decisions are achieved when decay equals inhibition (i.e., $k = 10$).

Even when the parameters are balanced in this manner, the mutual inhibition model only approximates the DDM, since its state merely approaches the decision line and remains with high probability in its proximity, rather than moving exactly along the line. As a result, sample paths cross thresholds $Z$ over an interval along which the coordinate $x_2$ takes different values. This is non-optimal: The SPRT requires a single pair of threshold points. The larger

$^9$ Note from Figure 9b that, for $w=10$, the model performs far from fully optimally. We choose this value to illustrate the robustness of the property that balanced decay and inhibition improve performance – this fact holds even if these quantities are not so large as to enforce collapse of the dynamics to a single dimension. That is, balancing these parameters improves performance, even in regimes in which performance is not fully optimal.
are decay and inhibition, the closer the mutual inhibition model is to the DDM (because attraction toward the line is stronger). Hence we expect that the balanced mutual inhibition model will achieve the shortest DTs for fixed ERs when decay and inhibition are larger. This prediction is verified in the simulations shown in Figure 10b. The mutual inhibition model was simulated with different values of decay $k$ and inhibition $w$, (maintaining equality of the two) while all other parameters were held constant across simulations. For each value of decay, the threshold was set such that the ER of the network was 10%. Figure 10b shows that as decay and inhibition increase, DT decreases. Furthermore, the mutual inhibition model's DT approaches that predicted by the optimal DDM for a 2% ER, shown by the dashed line in Figure 10b.

In summary, the mutual inhibition model approaches optimal performance in the free-response paradigm when decay equals inhibition and both increase in magnitude. Thus, as in the interrogation paradigm, optimality depends on the equivalence of decay and inhibition. However, in the free-response paradigm, optimality further depends on the absolute magnitude of these parameters. In the extreme, when both decay and inhibition are zero (left-most point in Figure 10b), the DT in the mutual inhibition model corresponds to that of the race model (since in this case the mutual inhibition model becomes precisely the race model, see Figure 4) which has a slower DT than the balanced mutual inhibition model (with any $k=w>0$).

Figure 10c shows similar results for the feedforward inhibition model: DT for fixed ER=2% and different weights ($u$) of the inhibitory connections. The DT for $u = 1$ again corresponds to the DT of the race model; both left-most points in Figures 10b and 10c correspond to the same value of DT (the small difference reflects a simulation artifact due to finite sample size). The race model has a slower DT than the feedforward inhibition model (for any $u > 0$), as it does in comparison with the mutual inhibition model. Note that when the inhibitory and
excitatory connections from the inputs are equal (at \( u = 1 \), as in the original model by Ditterich et al., 2003), then the feedforward inhibition model achieves the same DT predicted by the DDM (indicated by the dashed line). This is expected since, as described in Section 3.3, when \( u = 1 \) the feedforward inhibition model is exactly equivalent to the DDM and hence achieves optimal performance. We will not discuss the feedforward inhibition model further in this article, since it can be considered equivalent to the DDM when \( u = 1 \).

When starting point is allowed to vary, simulations indicate that, as in the case of the interrogation paradigm, balancing decay and inhibition may not be optimal (i.e., for fixed \( w + k \), a minimum DT for fixed ER is achieved for \( w \neq k \)). However, we note again that this finding should be considered within the context of a psychological interpretation of the starting point and its relationship to stimulus and response frequencies, which we address in Section 6.3.

### 4.3. Balanced systems and arbitrarily small errors

For the decision models considered in this article, increasing interrogation time in the interrogation paradigm or threshold in the free-response paradigm leads to increased accuracy. This suggests another sense in which these models may be optimal: we can ask whether ERs can be reduced to arbitrarily small values in the models of these two paradigms.\(^{10}\) We first consider the case of drift being constant across trials ("pure" models).

---

\(^{10}\) It must be noted here that humans rarely achieve perfect performance, even when a premium is placed on accuracy and participants are given unlimited time to respond (Reed, 1973; Ratcliff, 1978; McElree & Dosher, 1989; Usher & McClelland, 2001). This discrepancy with certain decision models has been cited as a challenge to their veracity (in particular, by Usher & McClelland (2001) with respect to the DDM). However, Ratcliff (1988) showed that the imperfect accuracy of human performance in long time limits may be explained by making a simple, biologically plausible modification to the DDM: adding boundaries limiting the maximum and
We show that in the interrogation paradigm, ERs can be reduced to arbitrarily small values for the DDM and the balanced mutual inhibition model (i.e., with decay equal to inhibition). In the free-response paradigm, this is true for both the balanced mutual inhibition model, as well as an unbalanced variant in which decay dominates inhibition. Then we show that if drift varies across trials, arbitrarily small error cannot be achieved by any of the models (results summarized in Table 1). The variability of starting point does not influence the ER as threshold or interrogation time goes to infinity.

Assuming collapse to a one-dimensional process for the long DTs in question, in what follows we consider only O-U processes that approximate the mutual inhibition model (although, as discussed above, we do not depend on this collapse for our O-U model to accurately describe the interrogation paradigm).

4.3.1. Interrogation paradigm

The minimum ER that can be achieved by an O-U model (i.e., after interrogation at infinite time) is given by:

\[
\text{ER} = \Phi \left( - \frac{2A^2}{c^2} \right)
\]

(from Equation 2.12 above via Equation A.94 of the online Appendix; Busemeyer & Townsend, 1992; cf. Equation 10 in Usher & McClelland, 2001). In Equation 4.1, \( \Phi \) denotes the normal standard cumulative distribution function, and \( A \) and \( c \) denote the drift and standard deviation of noise as before. From this formula it follows that the ER can be arbitrarily close to zero only when \( \lambda = 0 \); that is, when decay is equal to inhibition in the minimum values of the diffusion variable \( x \). We have confirmed this and provided analytical estimates, but we do not treat this issue here, leaving its further study for future work.

48
mutual inhibition model. Indeed, for $\lambda \neq 0$ there is a certain probability of error even for very long interrogation times, although this may be small for high drift rates (i.e., easy decisions). To illustrate this result, Figure 11 shows the evolution of states of the mutual inhibition model for three different ratios of decay and inhibition. These simulations represent moderately difficult decision tasks (mean input to the first unit $I_1 = 6.1 \, [s^{-1}]$, to the second unit $I_2 = 3 \, [s^{-1}]$, magnitude of noise $c = 1 \, [s^{-1}]$; values corresponding to middle difficulty condition of participant S1 in Experiment described by Usher & McClelland, 2001). Figure 11b illustrates the case for the balanced model (when decay equals inhibition). The top panel shows states of the model from independent trials after 0.5s: they have already reached the proximity of the decision line, but are still located close to their starting midpoint in the $x_1$ direction. The middle panel shows model states after 1.0s. States have begun to diffuse along the decision line, and the mean of the density has moved toward the correct decision (i.e., rightward). The bottom panel shows the state of the system after 1.5s: the density of points has a greater standard deviation in the $x_1$ direction, however the mean has now moved far to the right, so that a great majority of states are on the correct side of the line dividing decisions ($x_1=0$; marked as dashed line). A critical property of the pure DDM is that the mean of the probability density of its solutions eventually moves faster than the density spreads, so that the probability of being on the correct side of the line dividing decisions ($x_1=0$) converges to 1 with time. Indeed, it follows directly from Equation 2.7 that, as $T$ goes to infinity, ER approaches zero.

Figure 11a shows the evolution of the states of the system when decay is larger than inhibition (both parameters are modified by 15% from their values in Figure 11b). The evolution along the decision line is now described not by the diffusion process, but by a stable O-U process (i.e., $\lambda < 0$; also see Figure 6a). The middle and bottom panels show that the distribution
converges to a steady state, with little difference between the states at times 1s and 1.5s. The mean is influenced by the drift (note that the density is slightly shifted to the right, i.e., in the direction of the drift), but it no longer moves: it has converged to the attracting fixed point described in Section 3.1 (Figure 6a). Interrogation at future times will result in essentially the same proportion of solutions from the stationary distribution on each side of \( x_1=0 \), and therefore with an irreducibly finite number on the wrong side of the \( x_1=0 \) line contributing errors.

Figure 11c shows the evolution of model states for decay slightly smaller than inhibition. Evolution along the decision line is now described by an unstable O-U process (i.e., \( \lambda > 0 \); also see Figure 6c). The points which, due to noise, initially move to one or the other side of the fixed point tend to remain on that side of the fixed point and be pushed rapidly away from it. Hence, no matter how long the interrogation time is, there will again be a finite fraction of trials on which an error is made. Remarkably, Equation 4.1 shows that this fraction depends only on the magnitude of \( \lambda \) and not its sign.

Figure 11 also illustrates the causes of the recency and primacy effects in the O-U model discussed in Section 2.4. When \( \lambda < 0 \) as in Figure 11a, the solutions fluctuate around the fixed point and the effect of inputs decays over time, so that the most recent inputs have the biggest impact on the current state of the model. By contrast, when \( \lambda > 0 \) as in Figure 11c, the fluctuations in input at the beginning of the trial strongly influence the response in the interrogation paradigm, because if a state of the model departs from the fixed point sufficiently, the repulsion from the fixed point is so strong that it is likely to remain on the same side of the fixed point.

4.3.2. Free-response paradigm
In contrast to the above results for the interrogation paradigm, for free-response the minimum achievable ERs depend not only on the extent to which a model is unbalanced (i.e., on the absolute value of $\lambda$), but also on whether decay or inhibition dominates (i.e., on the sign of $\lambda$).

As we show in online Appendix A.2.2.2 (Equation A.59), if $\lambda \leq 0$ then, as the thresholds tend to infinity, ER tends to zero. However, for $\lambda < 0$, as the thresholds are increased to achieve low ERs, DTs become excessively long (Ratcliff & Smith, 2004): the greater distance of the threshold from origin is now compounded by the fact that when $\lambda < 0$, there is an interior attracting point (that is, one that is closer to origin than the threshold; see Figure 11a). This produces solutions that are confined, with exponentially decaying probability, near that attractor. (Recall from Section 4.2 that, for fixed ER, when $\lambda < 0$, DTs are at least expected to be longer than those of the DDM.) Thus, if the threshold is set above a certain level, many decisions will not be made within a given finite time. If the threshold is lowered to the level at which these decisions can be made, there will be always a non-zero fraction of errors.

Finally, for the case $\lambda > 0$ (see Figure 11c), minimum achievable ERs as thresholds tend to infinity are finite, and given by Equation 4.1. This is because the unstable process always forces some fraction of solutions across the threshold for the incorrect decision. This is shown in online Appendix A.2.2.2 (see Equation A.58 and recall the definition of $\tilde{a}$ in A.28).

It is interesting that this limit of ER as threshold goes to infinity in the free-response paradigm coincides with the limit of ER as DT goes to infinity in the interrogation paradigm when $\lambda \geq 0$, but that it does not in the case $\lambda < 0$.  

11 This observation could be used to adjudicate models that assume $\lambda \geq 0$ (such as the DDM) versus $\lambda < 0$, by comparing performance in the interrogation paradigm at long interrogation times to performance in the free-
Figure 9b summarizes these results by showing how the ER of the O-U process depends on the value of the threshold. Parameters for the simulation in Figure 9b match those for evolution along the decision line in Figure 11. For $\lambda = 0$, the ER converges to 0 for large thresholds. For $\lambda > 0$, the ER converges to a finite value after which increasing the threshold has negligible effect. For $\lambda < 0$, although the number of trials with errors decreases to 0 for larger thresholds, on a high proportion of trials a decision is not reached within a fixed allotted time (shown by thin line) since expected DTs diverge to infinity. Figure 9b is in agreement with the finding of Busemeyer & Townsend (1993) that lower (more negative) values of $\lambda$ lead to lower values of ER in the free-response paradigm (see left panel of Figure 9 in Busemeyer & Townsend, 1993).

4.3.3. Lower bounds on ER due to variability of drift

If the drift in the DDM is selected at the beginning of each trial from a normal distribution with mean $m_A$ and standard deviation $s_A$, then on a certain fraction of trials the drift will be negative (Ratcliff, 1978). Hence even as time goes to infinity in the interrogation paradigm or threshold is taken to infinity in the free-response paradigm, there is a non-zero probability of error equal to exactly the fraction of trials with negative drift, which is

$$ER = \Phi \left( -\frac{m_A}{s_A} \right).$$

(4.2)

Intuitively, this follows from the fact that, for long times, drift always dominates the effects of noise or initial conditions in the DDM. In fact, as long as $m_A > 0$, this value of ER is actually a lower bound on ERs for any value of interrogation time or decision threshold in both the response paradigm when strong emphasis has been placed on accuracy. $\lambda \geq 0$ predicts comparable ERs while $\lambda < 0$ predicts that they will differ.
extended DDM and extended O-U model, as shown in online Appendix A.4.3. For the O-U model, whether or not this bound is actually obtained depends on the task paradigm (interrogation versus free-response) and the value of \( \lambda \).

The main results of Section 4 are summarized in Table 1.

**5. A theory of optimal thresholds**

In any speeded decision task one has to choose whether to be faster or more accurate. In the DDM (and all of the related models considered above), this speed-accuracy tradeoff is determined by the value of the decision threshold. As we have seen, for any given threshold (that is, for any particular tradeoff between speed and accuracy), the DDM is the optimal decision making process. However, this begs the question: How should the threshold be set to achieve optimal performance? The answer to this question depends, of course, on the criterion that is used to evaluate performance, which in turn depends on the goals of the decision maker. There have been a number of efforts to explain human decision making performance in TAFC tasks in terms of the optimization of specific criteria (e.g., Edwards, 1965; Maddox & Bohil, 1998; Gold & Shadlen, 2002). Different criteria are optimized by different thresholds, which in turn have different implications for performance. In this section, we consider the various criteria that have been proposed, the derivation of their optimal thresholds, and their relationship to task parameters and performance.

We begin in Section 5.1 by reviewing the various criteria that have been proposed in the literature, including Bayes risk and reward rate. In Section 5.2, we find the thresholds that optimize these criteria for the pure DDM, show analytically that there is a unique optimal threshold for each of these criteria, and examine how this depends on task parameters (drift, noise, and the pace of the task). In Section 5.3 we show that, for each criterion, the DDM using the threshold optimal for that criterion outperforms all other decision processes. In
Section 5.4, we derive an equation that specifies the DT for a given ER (and vice versa) under optimal performance, thus defining a novel “optimal performance curve.” In Section 5.5 we apply our analysis to the extended DDM. Finally, in Section 5.6 we address the question of how the optimal threshold is discovered by the decision maker, and the implications that this may have for performance.

5.1. Criteria for optimality

Bayes Risk (BR). This criterion was first used by Wald & Wolfowitz (1948) in proving the optimality of the SPRT. It assumes that decision makers seek to minimize a cost function that is the weighted sum of DT and ER:\footnote{For consistency, note that since DT is in units of time and ER is dimensionless, the units of \( c_2 \) must also be time. This implies the use of a characteristic time to normalize the weights.}

\[
BR = c_1 DT + c_2 ER. \tag{5.1}
\]

Edwards (1965) generalized this criterion in considering how human participants choose decision thresholds. He considered an experimental paradigm in which (i) observing the stimulus has a cost \((c_1)\) that scales with the duration of the observation (participants lose \(c_1\) monetary units for each unit of observation time), and (ii) there is a payoff matrix for different types of errors \((c_2)\). For a payoff matrix with no reward for correct responses and a penalty \(c_2\) for errors, Edward’s cost function reduces to BR. Edwards (1965) derived the optimal threshold for the DDM minimizing this cost function (we consider these explicitly in Section 5.2 below). Rapoport & Burkheimer (1971) extended Edwards' theory, and Busemeyer & Rapoport (1988) designed an experiment to test it, finding that the participants chose decision thresholds close to those minimizing the BR.
The assumption that decision makers optimize BR requires further specification of the relative value that they place on speed versus accuracy – that is, the values of parameters $c_1$ and $c_2$ in Equation 5.1. These values can be inferred from fits to experimental data (e.g., Mozer et al., 2002), or explicitly manipulated (as in the case of Busemeyer & Rapoport, 1998). In contrast, assuming that decision makers optimize reward rate does not require any further specification; that is, it is a parameter-free criterion.

**Reward Rate (RR).** This can be defined as the proportion of correct trials divided by the average duration between decisions, namely (Gold & Shadlen, 2002):

$$RR = \frac{1 - ER}{RT + D + ER \cdot D_p}$$  \hspace{1cm} (5.2a)

$$= \frac{1 - ER}{DT + T_0 + D + ER \cdot D_p}$$  \hspace{1cm} (5.2b)

In Equation 5.2a, RT denotes total response time, $D$ the interval between a correct response and the next stimulus (the response-stimulus interval), and $D_p$ an additional “penalty” delay introduced following an error (which may, of course, be set to 0)\(^{13}\). Following the conventions used in previous sections, Equation 5.2b decomposes RT into its two components: DT, the mean decision time; and $T_0$, the time required for all non-decision related (e.g., sensory and motor) processing. This formulation assumes that, after each correct decision participants receive a reward, and after every response they have to wait a certain interval before the onset of the next stimulus (and hence the possibility of making the next choice).

\(^{13}\) Such delays are often used in studies of non-human animals, as penalties for incorrect responses, to encourage accurate performance while avoiding the use of aversive stimuli as punishment.
The BR and RR criteria differ in a number of important respects. First, long DTs are penalized differently: BR assumes that there is an explicit cost (e.g., financial) for each additional increment of stimulus observation time, while for RR long DTs simply slow down the rate at which rewards are received. The direct dependence of the RR criterion on time also predicts different behavior for different decision-stimulus intervals $D$ and $D_p$ (as will be shown later), parameters that do not enter into BR.

Most importantly however, BR and RR have different implications for the emphasis placed on reward versus accuracy. RR, by definition, seeks to maximize reward. As we show further on, optimizing RR may sometimes demand faster and less accurate responding. If the experiment is configured exactly as described in Section 5.1 (i.e., participants are financially penalized for errors and time spent on decision) minimization of BR is equivalent to maximization of reward. However, in general, BR allows that accuracy may be favored over speed, even if this means receiving less overall reward. A number of studies have suggested that decision makers do sometimes favor accuracy over maximizing reward.

Maddox & Bohil (1998) investigated participants’ choices in a version of the interrogation paradigm in which one stimulus is more frequent or more rewarded. They noticed that in this task participants may choose the more frequent or the more rewarded alternative even if the accumulated evidence at the interrogation time favors the other alternative. Let us define the decision boundary as the value of variable $x$ in the DDM above and below which different alternatives are chosen (so far we assumed that the decision boundary is at 0). Maddox & Bohil (1998) suggested that participants chose decision boundaries between the values maximizing reward and those maximizing accuracy, even when they were specifically instructed to maximize reward.
The investigators offered three reasons why participants may tend to maximize accuracy: (i) To choose the boundary maximizing reward, participants need to gain knowledge about the distributions of the stimuli of different classes (i.e., varying reward value). (ii) There is evidence that participants are more concerned about being accurate than about gaining modest rewards (Pitz & Reinhold, 1968; Stevenson et al., 1991). (iii) The trial-by-trial feedback about accuracy is often equally or more salient than the feedback about the reward (in psychological experiments, participants are often informed if they made an error and they may pay more attention to this information than to information about the reward). They also observed large individual differences in the emphases placed on reward versus accuracy among participants. On the basis of these findings, Maddox and Bohil proposed a model called Competition Between Reward and Accuracy (COBRA), which postulates that decision makers pursue a balance between reward and accuracy maximization.

**Reward Accuracy (RA) and Modified Reward Rate (RRm).** The COBRA theory was proposed in the context of the interrogation paradigm. Here, we extend it to the free-response paradigm. Furthermore, as Maddox & Bohil (1998) did not formulate an equation for the function being optimized, we consider two different ways of doing so, each of which penalizes errors in a different manner. The first is a reward/accuracy (RA) function that is a weighted difference of RR and accuracy:

$$\text{RA} = c_1 \text{RR} - \frac{c_2}{D_{total}} \text{ER},$$

(5.3)

where $D_{total} = D + T_0$ is the total delay between the end of one decision and the beginning of the next. We assume that penalty delay $D_p$ (from the expression for RR in Equations 5.2a,b) is set to zero, since errors are penalized explicitly in the second term of Equation 5.3. For $c_1$, $c_2 > 0$, maximization of this function emphasizes accuracy by negatively weighting errors (in
the second term) in addition to positively weighting correct responses (via RR in the 1st term). Note that inclusion of $D_{total}$ in the second term allows the units [1/time] to be consistent.

The second criterion, a modified reward rate ($RR_m$) describes the condition under which correct responses are rewarded by $c_1$ and errors are penalized by $c_2$:

$$RR_m = \frac{c_1(1-ER) - c_2ER}{DT + D_{total}} \quad (5.4)$$

Note that since errors are dealt with explicitly via the second term in the numerator, we again neglect the penalty delay, setting $D_p = 0$. Also note that Equation 5.4 is closely related to Equation 5.3, but that it now includes the full inter-stimulus interval $DT + D_{total}$ in the denominator of the second term. As expected, both RA and $RR_m$ reduce to RR when the accuracy weight $c_2=0$ (Equation 5.3 with $D_p=0$).

COBRA was introduced to explain the observation that decision makers sometimes seem to favor accuracy over a strict maximization RR. However, like the BR criterion, this requires the estimation of two parameters ($c_1$ and $c_2$). The question remains whether it is possible to explain decision making behavior using the parameter-free RR criterion. We will return to this question below, after we derive the thresholds for the DDM that optimize the various criteria discussed above.

5.2. Optimal thresholds for different criteria in the pure DDM

We begin by considering Edward’s (1965) calculations of thresholds minimizing BR, then derive the thresholds that optimize the three other criteria reviewed above, and examine the dependence of optimal thresholds on task parameters. For simplicity, here we consider decisions produced by the pure DDM; in Section 5.5 we explore the effects of introducing drift rate and starting point variability in the extended DDM.
5.2.1. Properties common to the criteria for optimality

First, we observe that thresholds optimizing the criteria of Equations 5.1, 5.3, and 5.4 do not depend on parameters $c_1$ and $c_2$ separately, but only on their ratio (the value of the functions themselves, of course, do depend on $c_1$ and $c_2$). Therefore, we define a parameter $q$ describing the weighting of accuracy relative to speed or reward:

$$ q = \frac{c_2}{c_1}. \quad (5.5) $$

Figure 12 shows thresholds that optimize the four criteria, computed by seeking a minimum of BR and maxima of RR, RA and RR_m respectively (as described further below and detailed in online Appendix A.2). Each criterion corresponds to a row of panels, and each column shows the dependence of the threshold on a particular task parameter while the others are kept fixed. Before focusing on individual criteria, we emphasize features common to all.

The left-most column of Figure 12 shows that, in contrast to the other parameters, the optimal threshold depends non-monotonically on the drift rate $A$. If the drift rate is equal to zero, the stimulus contains no information. In this case, the optimal strategy is to guess immediately, and so the optimal threshold is equal to 0. As the drift rate increases, and information becomes available, it becomes advantageous to integrate and so the optimal threshold increases. However, as the drift rate becomes sufficiently high, it begins to dominate the noise, so that threshold can again be lowered until decisions can be made almost immediately.

The second column from the left shows the dependence of the optimal threshold on noise ($c$). If $c$ is close to zero, then the drift is not obscured by noise (i.e., the signal is strong) and decisions can be made with little need to integrate information (i.e., average the signal) over time. Thus, the optimal threshold is close to zero. As noise increases, more integration is required in order to obtain an accurate estimate of the signal, and thus the optimal value of the
threshold increases. Interestingly, as noise goes to infinity, the optimal threshold appears to converge to a constant (we prove this for BR and RR below and we conjecture that it is also the case for the other criteria on the basis of simulations).

The third column from the left shows the dependence of the optimal threshold on the total delay $D_{\text{total}}$. As $D_{\text{total}}$ decreases towards zero (i.e., the next stimulus is presented immediately after the response, ignoring $T_0$), the optimal value of the threshold also decreases towards zero$^{14}$. That is, it is optimal to make decisions immediately, since there is no opportunity cost for errors. Although responding will be random (since no information has been accumulated), and therefore produce a reward on only 50% of trials, the arbitrarily high frequency of trials serves to maximize reward$^{15}$. As the delay $D_{\text{total}}$ increases, so too does the optimal value of the threshold: An increase in the opportunity cost for responding requires that each response be more accurate. There is no limit on the optimal value of the threshold.

The rightmost column of Figure 12 shows that for greater emphasis on accuracy (i.e., greater values of $q$), optimal thresholds rise, which can be expected because higher thresholds increase accuracy.

### 5.2.2. Derivation of optimal thresholds for individual criteria

---

$^{14}$ For BR, RR, and RR$_{\text{un}}$, the optimal threshold converges to 0 as $D_{\text{total}} \to 0$. However, for RA, the second term in Equation 5.3 can dominate as $D_{\text{total}}$ decreases to 0, leading to a nonzero limiting value for optimal threshold (not visible in Figure 14 as the minimum of $z_0(D_{\text{total}})$ occurs for $D_{\text{total}} \approx 0.01$ for the parameters used there).

$^{15}$ Of course, in practice this could not occur, since $D_{\text{total}}$ includes $T_0$ — the non-decision component of the response time, which includes the time for stimulus encoding and motor response, and therefore is necessarily non-zero. Consequently, the optimal threshold will always be non-zero (for $A>0$).
The features above are general to all criteria. Here we consider how optimal thresholds differ among criteria.

*Bayes Risk.* Edwards (1965) has shown that the BR is minimized by the threshold \( z \) satisfying the following condition (Equation 17 from Edwards, 1965 rewritten in the present notation; for completeness, we derive this in online Appendix A.2.5, Equation A.76):

\[
q \frac{2A^2}{c^2} - \frac{4Az}{c^2} + e^{\frac{-2Az}{c^2}} - e^{\frac{2Az}{c^2}} = 0
\]  

(5.6)

This transcendental equation does not have closed form solutions for \( z \), but can be solved numerically, as was done for the first row of Figure 12. Furthermore, the terms can be rearranged so that Equation 5.6 expresses equality between increasing and decreasing functions of \( z \), thus demonstrating the uniqueness of its solution. It is also useful to note that for low \( A \), high \( c \) or low \( q \), Equation 5.6 simplifies to:

\[
z = \frac{Aq}{4},
\]

(5.7)

which gives the slopes and the limit indicated in first row in Figure 12. We also note that when \( q = 0 \), the accuracy does not influence the BR and hence the optimal threshold is equal to 0.

*Reward Rate.* This criterion is defined in Equation 5.2 in terms of ER and DT. The optimal threshold can be derived by re-expressing these in terms of the parameters of the DDM (drift, noise and threshold) given by Equations 2.8 and 2.9, and then solving for the threshold that maximizes RR. Substitution of these equations into 5.2 followed by some algebraic manipulation yields the expression:

\[
\sqrt{\text{RR}} = \frac{z}{A} + D + T_0 + \left( D + T_0 + D_p - \frac{z}{A} \right) e^{\frac{2zA}{c^2}}
\]

(5.8)
Note that this expression depends only on performance variables and task parameters — that is, it does not rely on any criterion-specific parameters. Note also that the parameters of the pure DDM \((A, c, z)\) do not occur independently, and can be grouped into two new parameters. We define these two new parameters of normalized threshold \((\tilde{z})\) and signal to noise ratio \((\tilde{a})\) as follows:

\[
\tilde{z} = \frac{z}{A}, \quad \text{and} \quad \tilde{a} = \left(\frac{A}{c}\right)^2.
\] (5.9)

(The BR threshold condition in Equation 5.6 may also be written solely in terms of \(\tilde{z}, \tilde{a}\) and \(q\). Although \(\tilde{a}\) is the squared signal to noise ratio, we will refer to it simply as the signal to noise ratio in what follows. Note that the parameters \(\tilde{z}\) and \(\tilde{a}\) fully determine the ER and DT of the pure DDM (Equations 2.8 and 2.9)\(^{16}\). With these new parameters, Equation 5.8 becomes

\[
\frac{1}{\text{RR}} = \tilde{z} + D + T_0 + (D + T_0 + D_p - \tilde{z})e^{-2\tilde{a}}.
\] (5.10)

The above equation has a unique minimum (as shown in online Appendix A.2.2.3), at which the RR is maximized, given by the solution of:

\[
e^{2\tilde{a}} - 1 = 2\tilde{a}(D + D_p + T_0 - \tilde{z}).
\] (5.11)

While this (transcendental) equation does not admit an explicit solution in terms of elementary functions, it may easily be solved numerically for \(\tilde{z}\) to yield the optimal threshold (normalized by \(A\)), as a function \((f)\) of \(\tilde{a}\) and task delays as follows:

\[\text{\footnotesize{(Below, we take the unit of time to be seconds.)}}\]

\[16\text{ Although the new parameters are ratios, they have units: } \tilde{z} \text{ has units of time, and } \tilde{a} \text{ has units of } 1/\text{time.}\]
\[ z_o = f(\tilde{\alpha}, D + D_p + T_0). \] (5.12)

First, we note that the optimal threshold does not depend on \( D \) and \( D_p \) separately, but only on their sum \( D+D_p \) (or actually on \( D+D_p+T_0 \)). Hence, for example, the same value of the threshold is optimal for a long intertrial interval but no penalty delay (e.g., \( D=2s, D_p=0s \)) as for a much more rapid pace of presentation but with a corresponding increase in penalty delay (\( D=0.5s, D_p=1.5s \)). Therefore, provided that decision makers optimize their thresholds to maximize RR, they should have the same ER and mean RT in both of these conditions.

Let us denote the sum of the three delays influencing the optimal threshold together by \( D_{total} \):

\[ D_{total} = D + D_p + T_0. \] (5.13)

The dependence of the optimal threshold on \( D_{total} \) (rather than \( D \) and \( D_p \) separately) is not a property specific to the pure DDM; it is true for any decision making mechanism (e.g., an O-U model) seeking to optimize RR, as shown in online Appendix A.2.3. Multiplying Equation 5.12 by \( A \) gives the optimal absolute (non-normalized) threshold in terms of the original parameters:

\[ z_o = A\tilde{z}_o = Af \left( \frac{A^2}{c}, D_{total} \right) \] (5.14)

The second row of Figure 12 shows how \( z_o \) depends on the parameters \( A, c \) and \( D_{total} \), (see Equations A.38 and A.39 of online Appendix A.2.1.2). In addition to the general properties of optimal thresholds described above (and shown in Figure 12), we can define the following quantitative characteristics of the optimal threshold for the RR criterion: (i) it increases logithmically as \( D_{total} \) goes to infinity, (ii) it converges to \( AD_{total}/2 \) as \( c \) goes to infinity, (iii) it reaches a maximum of \( c^2/A \) (or \( 1/\tilde{\alpha} \)) as drift increases, and then decreases to 0 as drift is increased further (see Equations A.38-39 in online Appendix A.2.1.2). These relationships should be empirically testable.
Rewards/Accuracy and Modified Reward Rate. The expressions specifying the optimal threshold for these criteria are more complex and are given in the Equations A.78 and A.84 of the online Appendix A.2.5. We note here, however, that there is a unique optimal threshold for both of these criteria. This follows from the fact that they are both sums of a function with a single maximum in $z$ (the RR term), and a function that is decreasing in $z$ (the additional accuracy terms).

5.3. The DDM optimizes all criteria relative to other decision making processes

Above, we showed that for a given set of task parameters, there is a unique value of the threshold for the DDM that optimizes each of the criteria considered. We can also ask how, at optimal performance, the DDM compares to other models? Online Appendix A.1.1 shows that the optimality of the DDM (in the sense of Wald (1947), and discussed in Section 2.2) also implies that the DDM with optimal threshold for a given criterion gives the minimum BR or the maximum RR, RA, RR$_m$ achievable by any model. In fact the DDM optimizes any criterion involving accuracy and speed that satisfies the broad conditions specified in online Appendix A.1.1.

From these observations, we can further assert that, among mutual inhibition models with optimal thresholds, the overall maximum RR will be achieved when decay is equal to inhibition and both are of a sufficient magnitude (see Sections 2-3). Figure 10d demonstrates this point for a particular parameter set. By the same reasoning, the O-U model yields optimal RRs when $\lambda=0$; the calculations in online Appendix A.2.2.3 demonstrate this explicitly. Furthermore, the exact Equation A.57, and its approximate forms that follow, may be used to evaluate the extent to which RRs for O-U model depart from optimality.
5.4. The relationship between ER and DT at optimal performance

The analyses described above indicate that, for a given criterion and set of task parameters, there is a single threshold that will optimize performance of the pure DDM. This threshold defines a specific point in the speed-accuracy tradeoff; that is, a specific ER and DT that are optimal for a given criterion and set of task conditions. In this section, we derive an expression for each criterion that describes the optimal ER and DT as a function of task parameters (signal to noise ratio $\tilde{a}$, delay $D_{total}$, and, if appropriate, emphasis on accuracy $q$).

Each of these expressions describes a relationship between ER and DT at optimality, that we refer to as an optimal performance curve.

Similar curves can be derived for the extended DDM (as we will do below in Section 5.5). To the extent that the curves for different criteria and models can be distinguished from one another, they can be used to generate quantitative, empirically testable predictions concerning optimal performance under different decision making models and criteria.

The derivation of optimal performance curves for the DDM builds on the relationship between ER and DT for a given $D_{total}$, which can be deduced by rearranging Equations 2.8 and 2.9 to obtain the expressions for parameters $\tilde{z}$, $\tilde{a}$ as a function of ER and DT:

\[
\tilde{z} = \frac{DT}{1-2ER}, \quad (5.15)
\]

\[
\tilde{a} = \frac{1-2ER}{2DT} \log\left(\frac{1-ER}{ER}\right). \quad (5.16)
\]

Bayes Risk. Edwards (1965) showed that substituting the above equations into the optimality condition of Equation 5.6 and rearranging terms, yields a direct relation between ER, and DT normalized by the weight of accuracy $q$ under conditions of optimal performance defined as minimization of BR (Equation 20 in Edwards, 1965):
The left side of the above equation expresses the ratio of time in the trial used on decision processes to the maximum intertrial interval (total response-to-stimulus delay on error trials), while the right side is a function only of ER. Hence, the equation describes the relation between ER and normalized DT (as a fraction of maximum intertrial interval). This optimal performance curve is shown as the thick black curve in Figures 13b and c. It has a shape similar to that for BR, but the longest DTs are about 20% of the maximum inter-decision.
interval, and associated with an ER of about 18% (peak of the thick curves in Figures 13b, c)\textsuperscript{17}.

**Reward/Accuracy and Modified Reward Rate.** In online Appendix A.2.5 (Equations A.80-81) we derive analogous optimal performance curves for RA and RR\textsubscript{m} respectively (Equation 5.19-20):

\[
\frac{\text{DT}}{D_{\text{total}}} = \frac{E - 2q - \sqrt{E^2 - 4q(E + 1)}}{2q}, \text{ where } E = \frac{1}{\text{ER} \ln \frac{1 + \text{ER}}{\text{ER}}} + \frac{1}{1 - 2\text{ER}} \quad (5.19)
\]

\[
\frac{\text{DT}}{D_{\text{total}}} = (1 + q) \left( \frac{\frac{1}{\text{ER}} - q}{\ln \frac{1 + \text{ER}}{\text{ER}}} + \frac{1 - q}{1 - 2\text{ER}} \right)^{-1} \quad (5.20)
\]

The right hand sides of the above equations contain the additional parameter \( q \), hence they do not describe a single curve but rather a family of curves for different values of \( q \). As \( q \) approaches 0, both of these expressions converge to Equation 5.18, since in this case both criteria simplify to the RR. These families are shown in Figures 13b and 13c. Note that the RA and RR\textsubscript{m} criteria differ in their predictions: for RA the value of \( q \) does not influence the ER corresponding to maximum DT (i.e., the position of the peak), while for RR\textsubscript{m} increasing \( q \) moves the peak to the right. These observations provide a framework for constructing and testing the extent to which, and under what conditions, participants maximize these different criteria for optimality.

\textsuperscript{17} We recognize however, that this relationship may not be valid for intertrial intervals exceeding certain values (e.g., intertrial intervals of order of minutes), since for very large \( D_{\text{total}} \), the optimal threshold may be very large (see Figure 12a) and exceed the maximum biologically allowed firing rate of decision neurons.
5.5. Optimal performance curves for the extended DDM

In this section we compute the optimal threshold and optimal performance curve for the extended DDM. For simplicity, we consider only one criterion of optimality, RR: Optimal performance curves for the other criteria can be derived similarly, and we have also computed them for the O-U model with $\lambda \leq 0$ (results not shown here).

The variability in drift rate and starting point introduced in the extended DDM have a direct influence on the optimal threshold for the different criteria. Figure 14a shows that as the variability of drift increases, the optimal threshold first slightly increases, but then the information content of the drift becomes lower and the optimal threshold decreases. Numerical simulations show that as the variability of drift goes to infinity, the optimal threshold goes to zero, since the drift ceases to carry any useful information, and the optimal strategy is to guess. Figure 14b shows that as the variability of starting point increases, the optimal threshold also increases. Since the extended DDM assumes that the threshold in general must be higher than the range of starting points $s_x$, the optimal threshold is higher than $s_c$. It should be emphasized that in Figure 14b, starting point variability reflects noise in the initial conditions of the integrator, and not changing prior probabilities of alternatives (the threshold maximizing RR when the starting point correctly represents stimulus probabilities is discussed in Section 6).

Figures 14c and d show the optimal performance curves for the extended DDM for different values of $D_{total}$. These curves were determined as follows: parameters $s_A$, $s_x$, $D_{total}$, and $m_A$ were fixed, and $c$ was varied to obtain the different points on the curves (i.e., for each value of $c$, the optimal threshold was found numerically, and a point was added to the graph with corresponding values of ER and DT).
Four observations can be made about these curves. First, the shape of the curve depends on the value of $D_{total}$ (unlike for the pure DDM). Second, the variability of drift moves the optimal performance curve down, as can be expected since it predominantly decreases the optimal threshold (Figure 14a). Third, the variability of starting point moves the optimal curve up, as can be expected since it increases the threshold (Figure 14b). Finally, with variability of starting point, the optimal normalized DT for ER=0 is not equal to 0. This is because the extended DDM assumes that the threshold must be higher than the range of starting points, so it cannot be equal to 0, which is required for DT=0. Figure 14d shows that if $D_{total}$ is very low, the optimal performance curve with variability of starting point becomes monotonically decreasing.

5.6. Threshold setting algorithms

In Section 5.2 we derived expressions for the thresholds that optimize various criteria for a given set of task parameters. This raises the question: How do decision makers identify this optimal threshold? In this section we review published work addressing this question, and we show that any adaptive threshold setting model makes two predictions for a task in which participants maximize RR: (i) they will choose thresholds closer to optimal values for easier vs. harder tasks, and (ii) they will tend to choose thresholds at higher than optimal values.

One possibility is that the knowledge of optimal threshold is intrinsic; that is, the decision maker comes to the task with complete knowledge of the relationship between task parameters (signal, noise, delay, and their variability) and the optimal threshold. While this may be possible for highly practiced tasks within a narrow domain of parameters, it seems unlikely more generally and certainly for novel tasks. Here, we consider the more plausible (or at least more general) alternative that the decision maker has no intrinsic knowledge of the optimal threshold, but is able to determine (or approximate) it using an adaptive adjustment.
procedure: Different values of the threshold are sampled, the effects on a criterion are observed, and the threshold is adjusted to optimize the criterion\(^{18}\). Several studies have examined this possibility.

5.6.1. Review of threshold setting

Myung & Busemeyer (1989) performed an experiment using the free-response paradigm\(^{19}\) in which participants were explicitly required to minimize BR. They tested predictions of two algorithms for updating the decision threshold: (i) an error correction model, which assumed that participants increase the threshold after an error, and decrease it after a correct response; and (ii) a hill-climbing model which assumed that after receiving a reward participants make the same threshold modification as on the previous trial, while after an error they make a threshold modification opposite to that on the previous trial (this algorithm implements gradient-based reinforcement learning). Analysis of the experimental data suggested that both processes of threshold modification were in operation. During the course of the experiment, participants’ thresholds converged close to the vicinity of the optimal value, but convergence was slow and required hundreds of trials. Later work of Busemeyer & Myung (1992) showed that hill-climbing describes experimental data better than error-correction.

\(^{18}\) Note that such an adaptive mechanism is likely to be required even if the relationship of task parameters to optimal threshold is known. This is because, unless the task parameters are explicitly specified at the outset, they must be discovered. This, in turn, requires sampling at some threshold that is most likely not optimal (since the optimal one is not yet known), and therefore must be adapted as the task parameters are discovered.

\(^{19}\) In this experiment participants were presented with successive “results of medical tests” until they make a choice of “diagnosis”. Thus DT (in Equation for BR) for this experiment corresponds to the number of medical tests the participant observed before the diagnosis.
Erev (1998) proposed a different reinforcement learning algorithm for adjusting decision boundary in the interrogation paradigm. His algorithm did not involve gradients (i.e., differences between decision boundaries on sequential trials), but assumed that (i) there is a finite set of decision boundaries from which participants choose, (ii) participants store estimates of reward for each value of the decision boundary and update these estimates after each trial, (iii) the probability of choosing a certain decision boundary on a given trial is a function of this estimate. This algorithm is an extension of a reinforcement learning algorithm described in chapter 2 of Sutton & Barto (1998), and was shown to account for a large number of effects observed in the experimental data.

Maddox & Bohil (1998, 2001) and Bohil & Maddox (2003) observed that participants in their generalized interrogation paradigm task (see Section 5.1) chose decision boundaries closer to the value maximizing reward for stimuli that were easier to discriminate (d’=2.2) than for more difficult stimuli (d’=1). They explained this fact by noticing that the reward as a function of decision boundary is steeper for easier than for difficult stimuli, and hence that learning of the criterion maximizing reward is more rapid and robust for easier stimuli.

5.6.2. Threshold adaptation and the DDM

We now consider threshold adaptation in pure DDM in order to maximize RR. If a hill climbing procedure (Myung & Busemeyer, 1989) or Erev’s (1998) method is used, this process is guaranteed to converge on the threshold maximizing the RR, since we have shown that there is a single optimal threshold for a given set of task parameters, and RR drops monotonically as the threshold moves away from the optimal one (because the optimal threshold $\bar{z}_o$ defined by Equation 5.11 is the only threshold for which the derivative of RR with respect to $\bar{z}$ is zero, see online Appendix A.2.1.2, Equation A.33). Recently Simen et al. (2005, 2006) proposed a different threshold adaptation model which converges very rapidly to
the threshold maximizing the reward rate. Below, we focus on predictions common to algorithms that iteratively converge to optimal thresholds, rather than on their distinctions.

Figure 15 shows the RR as a function of the normalized decision threshold (defined in Section 5.2 as \( z/A \)) for three different levels of signal to noise ratio. Note that the RR is a steeper function of the threshold for higher signal to noise ratios. This provides an explanation of the observation made by Maddox & Bohil (1998, 2001) regarding RR maximization in the interrogation paradigm, and suggests that it should generalize to the free-response paradigm: the easier the task, the closer the decision threshold should be to its optimal value.

Furthermore, and perhaps most interestingly, considering the effects of adaptive threshold adjustment may explain the frequent observation that participants appear to favor accuracy over (or in addition to) reward maximization (Maddox & Bohil, 1998, 2001). From the perspective of pure RR maximization, this amounts to threshold overestimation. However, this may reflect an adaptive bias that is intrinsic to the threshold estimation procedure. Figure 15 plots RR as a function of threshold. Note that on the right side of the peak (denoting the optimal threshold, \( \tilde{z}_o \)) the curve falls with a shallower slope than on the left side. Thus, overestimation of the threshold by a small value \( \varepsilon \) gives higher RR than underestimation by the same amount; that is:

\[
RR(\tilde{z}_o - \varepsilon) < RR(\tilde{z}_o + \varepsilon).
\]  

(5.21)

Online Appendix A.2.4 shows that the above inequality is satisfied for all signal to noise ratios and experimental delays. That is, overestimation of the threshold is in general less costly than underestimation. Therefore, with any uncertainty of estimation, there should be a
bias toward threshold overestimation. Furthermore, if participants use a gradient-based learning algorithm to find the threshold, it is easier to establish the sign of the gradient of RR for $z < z_o$ (where RR($z$) is steeper) than for $z > z_o$ (where RR($z$) is shallower). Hence, it is easier to discover that the threshold should be increased when $z < z_o$, than to discover that it should be decreased for $z > z_o$. These factors suggest an account, strictly in terms of RR maximization, of the often observed bias toward threshold overestimation, without having to assume any additional criteria concerning accuracy such as RA and RR$_m$.

6. Biased decisions

When one of the alternatives of the decision process is correct more often than the other (e.g., one stimulus occurs more frequently in a TAFC task), then decision makers may use this information about the prior probability of the alternatives to optimize their responses. In this section we consider how this can be addressed formally by the DDM. We begin by reviewing existing theories about performance under conditions of biased alternatives, both when signal

These considerations suggest that we should in fact consider the distribution of thresholds (rather than the single deterministic value) that maximizes RR, with the prediction that the mean of this distribution will be greater than the single deterministic value. The characteristics of this distribution will be determined by factors such as RR sampling error and other sources of noise, which are parameters that control the degree of bias toward threshold overestimation.

We note that standard reinforcement learning algorithms for learning a single variable (chapter 2 of Sutton & Barto, 1998) also predict that decision makers will set higher than optimal thresholds. For example, the action-value method with soft-max selection (Sutton & Barto, 1998) would build estimates of the values of RR for different thresholds on the basis of the rewards experienced, and choose a new threshold on each trial with the probability being a monotonic function of the estimated RR for that threshold. Thus from Inequality 5.21, the probability of choosing threshold $\tilde{z}_o - \varepsilon$ would be lower than probability of choosing $\tilde{z}_o + \varepsilon$. 

---

20 These considerations suggest that we should in fact consider the distribution of thresholds (rather than the single deterministic value) that maximizes RR, with the prediction that the mean of this distribution will be greater than the single deterministic value. The characteristics of this distribution will be determined by factors such as RR sampling error and other sources of noise, which are parameters that control the degree of bias toward threshold overestimation.

21 We note that standard reinforcement learning algorithms for learning a single variable (chapter 2 of Sutton & Barto, 1998) also predict that decision makers will set higher than optimal thresholds. For example, the action-value method with soft-max selection (Sutton & Barto, 1998) would build estimates of the values of RR for different thresholds on the basis of the rewards experienced, and choose a new threshold on each trial with the probability being a monotonic function of the estimated RR for that threshold. Thus from Inequality 5.21, the probability of choosing threshold $\tilde{z}_o - \varepsilon$ would be lower than probability of choosing $\tilde{z}_o + \varepsilon$. 

---
strength is kept constant and when it is varied from trial-to-trial, and compare these with existing experimental data. We then consider how asymmetries in the probability as well as the reward value of each alternative influence the optimal threshold, find the relationship of ER to DT for this threshold, and relate these theoretical results to existing experimental data.

6.1. Biased probabilities and the optimal starting point

In this section we review three theories of how the starting point of integration in the pure DDM should depend on prior probabilities of the two alternatives (Edwards, 1965; Link, 1975; Carpenter & Williams, 1995).

Let us use \( \Pi \) to denote the probability of the alternative corresponding to the upper threshold \( +z \) being correct. Edwards (1965) showed that the optimal decision process (SPRT) in this case is the DDM with the starting point not at \( x_0=0 \), as for the unbiased case, but at a position \( x_0 \) reflecting the odds of the prior probabilities of alternatives (cf. online Appendix A.3.2):

\[
x_0 = \frac{c^2}{2A} \ln \frac{\Pi}{1-\Pi} \quad \text{or} \quad x_0 \frac{2A}{c^2} = \ln \frac{\Pi}{1-\Pi}.
\] (6.1)

Figure 16a confirms that \( x_0 \) given by Equation 6.1 is optimal, by demonstrating that, for a given threshold, it produces the minimum DT for a fixed ER. Note that according to Equation 6.1, we recover \( x_0=0 \) for \( \Pi = \frac{1}{2} \) (the unbiased case). More generally, however, the optimal starting point is proportional to the log of the ratio of prior likelihoods for the two alternatives. Hence, as the probability of a given response increases, the optimal starting point moves towards the threshold corresponding to that response. Furthermore, this effect is amplified as the decision becomes more difficult (i.e., the signal to noise ratio decreases). The dependence of the optimal starting point \( x_0 \) on the prior probability \( \Pi \) is shown by the solid line in Figure 16b. The pure DDM with starting point given by Equation 6.1 minimizes ER in the interrogation paradigm (as shown in online Appendix A.3.2) and implements the SPRT in the
free-response paradigm, hence it optimizes all four reward-based criteria analyzed in Section 5.

Link (1975) considered a different criterion of optimality: minimum ER for fixed decision threshold in the free-response paradigm. Although this may seem similar to the criterion of minimum DT for fixed ER that is optimized by Equation 6.1, Link showed that the optimal starting point in this case is halfway between zero and that of Equation 6.1, namely:

\[
x_0 = \frac{c^2}{4A} \ln \frac{\Pi}{1-\Pi} \quad \text{or} \quad x_0 = \frac{2A}{c^2} = \frac{1}{2} \ln \frac{\Pi}{1-\Pi}.
\]  

(6.2)

This dependence of starting point \(x_0\) on the prior probability \(\Pi\) is shown by the dashed line in Figure 16b. The differential predictions of Equations 6.1 and 6.2 are contrasted with experimental data in the next section.

Carpenter & Williams (1995) proposed a theory for the starting point of integration explaining the dependence of DT on prior probability \(\Pi\). However, their model assumes two separate neural integrators, each accumulating evidence for detection of one alternative (in a manner similar to the “race” model). As we showed in Section 2, this type of model cannot be reduced to the DDM, which is the optimal process for the discrimination tasks that are the focus of this article; therefore, we do not consider the Carpenter & Williams model further.

6.2. Experimental tests of the optimal starting point

Although the theories of Edwards (1965) and Link (1975) were proposed several decades ago, we are not aware of a systematic comparison of their predictions with experimental data. In this section, we show that predictions of Edwards’ theory matches better than those of Link’s theory with results of three experiments performed by Laming (1968), Link (1975), and Van Zandt et al. (2000) and further analyzed by Ratcliff & Smith (2004).
In order to perform such a comparison, it is necessary to estimate the parameters of the pure DDM (e.g., drift, starting point and threshold) from experimental data for tasks with biased alternatives. Ratcliff & Smith (2004) estimated parameters for the extended DDM (using methods similar to those described in Section 3.1) from the data of Van Zandt et al. (2000), and found that variability of drift $s_A$ and starting point $s_x$ were an order of magnitude smaller than other model parameters ($m_A$, $z$). Therefore, in what follows, we make the simplifying assumption that this task can be described by the pure DDM. Link (1975) made this assumption at the outset, and used it to develop the following direct method of parameter estimation from data. Denote the alternative corresponding to the negative threshold in DDM by ‘-’, and the other alternative, corresponding to the positive threshold, by ‘+’. Denote the probability of choosing the alternative ‘-’ if the stimulus ‘+’ is presented by $P_{-|+}$ (i.e., this is the probability of an error if stimulus ‘+’ is presented), and analogously $P_{+|+}$, $P_{-|+}$, $P_{+|\text{-}}$. All these probabilities can be directly estimated from the data (for example, $P_{-|+}$ can be estimated as the number of trials in which participant chose ‘-’ while ‘+’ was presented, divided by the total number of trials in which ‘+’ was presented). Link showed that the value of the starting point chosen by the participant (scaled by $2A/c^2$) may be estimated from data using the following equation:

$$x_0 \frac{2A}{c^2} = \frac{1}{2} \ln \left( \frac{P_{+|\text{-}}}{P_{-|+} P_{+|+}} \right).$$

(6.3)

and he computed this quantity for the experiments of Laming (1968) and Link (1975). Note that in deriving Equation 6.3, Link did not assume any theory of how participants choose the starting point, but only assumed that responses were generated according to the pure DDM; hence the estimates from Equation 6.3 may be used to contrast predictions of Equations 6.1 and 6.2.
In Figure 16b, we show that the starting points (scaled by $2A/c^2$) estimated by Ratcliff & Smith (2004) and Link (1975) are closer to the predictions of Equation 6.1 than 6.2 for all three experiments, and for the data of Laming (1968) this difference is significant (paired t-test, $p=0.04$). Such an outcome can be expected, since the starting point of Equation 6.1 optimizes criteria directly based on reward (i.e., RR or other measures discussed in Section 5), while the starting point of Equation 6.2 minimizes ER for fixed threshold, and does not optimize the RR. This result provides support for the hypothesis that decision makers adapt parameters to maximize the experimental reward. Below we do not discuss the criterion optimized by Equation 6.2 further, and for simplicity we refer to the staring point of Equation 6.1 as ‘optimal.’

6.3. Variable signal strength and adjustment of drift rate

In this section we review the argument that when drift varies across trials, and one of the alternatives is more probable, optimal performance is obtained by biasing both starting point and drift. We also verify the predictions of this theory with further experimental data from the literature.

Ratcliff et al. (1999) observed that decisions can be biased toward the more probable alternative within the DDM model in two ways (see Figure 32 of Ratliff et al., 1999): (i) by moving the starting point towards the threshold for the more probable alternative (as in Equation 6.1), or (ii) by increasing the drift in the direction of the more probable alternative and decreasing it for the other alternative. As shown by Edwards (1965; see Section 6.1), the optimal strategy for tasks with constant signal strength is to adjust the starting point and not the drift rate. However, if signal strength (e.g., stimulus difficulty) varies among trials within a block, then a correction of the drift rate, in addition to that of the starting point, is required to achieve optimal performance (Yang et al., 2005).
Analyses of experimental data indicate that these different strategies are indeed used, depending on the manipulations of signal strength across trials. In an experiment by Van Zandt et al. (2000) participants made same-different judgments about pairs of letters, and the probability $\Pi$ of one type of stimulus was varied between blocks. Recall from Section 6.2 that for this experiment the estimated variability of drift $s_A$ across trials was an order of magnitude lower than mean drift $m_A$. Ratcliff & Smith (2004) noticed that to accurately describe the data of Van Zandt et al. (2000), it sufficed to vary only the starting point $x_0$ to fit conditions of different prior probability $\Pi$, while mean drift $m_A$ remained constant (one value of parameter $m_A$ was used for different $\Pi$ conditions). By contrast, Ratcliff et al. (1999) reported that for their experiment, in which difficulty was varied across trials within the same block, both starting point $x_0$ and mean drift $m_A$ had to be allowed to vary significantly in order to fit performance in conditions of different prior probability. Furthermore, Ratcliff (1985) estimated parameters of pure DDM from data of Proctor & Rao (1983) and Proctor et al. (1984), and found that both starting point $x_0$ and drift $A$ varied between conditions of different prior probability. Both experiments of Proctor & Rao (1983) and Proctor et al. (1984) involved the same task: to discriminate whether two strings of four letters are the same or different, where the strings differed by at most one letter. Although the experiments did not involve explicit variability in difficulty across trials, it is reasonable to assume that, for example, noticing a disparity between strings differing in the first letter is easier than for strings differing in the third letter, making such variability implicit for this task design.

In summary, when asymmetries in the probability of alternatives are introduced, adjustments in the starting point of integration are sufficient to achieve optimal performance, and this appears to be what participants do. However, when variations in signal strength (such as stimulus difficulty) are also present, then an adjustment in drift rate is also required, and again there is evidence that participants respond accordingly.
These analyses have interesting implications for the neural network models described earlier, and in particular the optimal initial values of unit activities in the balanced mutual inhibition model. First, consider the case of constant signal strength across trials. According to the first of Equations 3.2 (describing the relationship between the activities of the decision units in the mutual inhibition model and the variable $x$ in the pure DDM), the difference between initial activities of the decision units should be equal to $\sqrt{2}x_0$ (where $x_0$ is given by Equation 6.1).

This can be achieved if the decision unit corresponding to the more probable alternative receives external input before the decision process. However, this input must cease when the signal appears and the decision process starts (implementing a bias in the prior expectation but not the processing of information supporting that alternative). Similar biases were used by Cho et al. (2002; although there they were allowed to persist during processing) in a mutual inhibition model to simulate the influence on RT and ER of changes in the local probabilities of stimuli in a sequence (experimentally manipulated by controlling the frequency of repetitions and alternations in the sequence of stimuli).

In the case of varying signal strength between trials, the analyses above indicate that drift rate should also be modified. According to Equation 3.12 (describing the relationship between mean drift in the extended DDM and mean inputs to the units of extended mutual inhibition model) the unit corresponding to the more probable alternative should, in addition to starting with an activity closer to its threshold, also receive external input during the decision process itself. However, the magnitudes of the input before and after onset of the stimulus need not be equal. In Sections 6.5-7, we focus on the simpler case of constant signal strength across trials.
6.4. Endogenous variations in drift rate and starting point (extended DDM)

In the preceding sections, we considered how parameters of the pure DDM (starting point and drift rate) should be adapted to optimize performance in response to manipulations of exogenous factors — namely stimulus probability and strength. Furthermore, we observed that there were specific optimal values of these parameters for a given set of task conditions — i.e., a fixed adjustment of starting point for a particular stimulus probability, and a fixed adjustment of drift rate for a particular distribution of stimulus strengths. These observations should be distinguished, however, from the fact that fits of the DDM to empirical performance are best when variability of drift and starting point are included in the model (Ratcliff et al., 1999). These concerns endogenous variability of parameters from trial-to-trial, and should therefore be distinguished from the fixed adjustments of parameters that are optimal in response to exogenous factors such as asymmetric stimulus frequency and variable stimulus strength. Indeed, the variability of drift and starting point in the extended DDM can improve fits to empirical data even when stimuli occur with equal frequency and fixed strength (Ratcliff & Rouder, 1998). In this case, such variability is clearly not optimal. In this section we discuss how such observations can be explained within the optimality framework.

One source of the variability of drift in the extended DDM may reflect variability of top-down attention. While this necessarily compromises optimality of task performance, elsewhere we have argued that such variability may serve an adaptive function on a broader scale, reflecting a tradeoff between the value of optimizing performance in the current task and exploring alternative opportunities for reward (Aston-Jones & Cohen, 2005).
Similarly, it is possible that the variability of the starting point reflects the function of other adaptive mechanisms operating on incomplete knowledge. For example, if the participant is not told the fixed probability with which each stimulus will occur, this must be inferred from observations about the frequency of their occurrence. Initial estimates will be subject to sampling error, and suggest starting points that are not appropriate for the true underlying probability of the stimuli. As estimates of stimulus frequency are revised, the starting point is adjusted. In addition to this source of variability, there is substantial evidence that the decision making apparatus is sensitive to the recent history of stimuli and responses, irrespective of knowledge about their overall probability. For example, several investigators have demonstrated priming effects based on local sequences of stimuli and responses (even when subjects are told that stimuli occur at random), that can be explained in terms of simple mechanisms sensitive to local repetitions and alternations that bias responding in a manner comparable to a shift in the starting point of the DDM (e.g., Cho et al, 2002). While such mechanisms are not optimal for processing truly random events, they may reflect adaptations that are well suited to the processing of more natural stimuli that exhibit local sequential structure. The operation of such mechanisms could contribute to variability of starting point captured by the extended DDM.

6.5. Biased probabilities and the optimal decision threshold

In this section we show that the decision threshold of the pure DDM maximizing RR is little influenced by the probability bias \( \Pi \), but above a certain \( \Pi \) the optimal starting point exceeds the optimal threshold, implying that the best strategy is to always choose the more probable alternative.

Edwards (1965) showed that there is a unique value of the decision threshold for the pure DDM that minimizes BR, and derived conditions that it must satisfy for different prior
probabilities of alternatives. Similar to the case of BR, there is a unique value of the decision threshold maximizing the RR that satisfies the following equation:

\[ e^{2\tilde{a}} - 1 = 2\tilde{a}(D_{total} - \overline{z}) + (1 - 2\Pi)\ln \frac{\Pi}{1 - \Pi}. \]  (6.4)

This expression is derived in online Appendix A.2.1.3 (Equation A.50). Note that when \( \Pi = \frac{1}{2} \), the last term of Equation 6.4 vanishes and it simplifies to the condition for the optimal threshold in the unbiased case of Equation 5.11. As the probability of the more frequent alternative \( \Pi \) deviates from \( \frac{1}{2} \), the optimal threshold decreases slightly, as shown for sample parameters in Figure 17a.

The modest effect of \( \Pi \) on the optimal threshold contrasts with its more profound effect on optimal starting point: As \( \Pi \) grows, the optimal starting point \( x_0 \) increases without bound toward the threshold (as discussed in the previous section, and illustrated in Figure 17a). Above a certain probability bias \( \Pi \), the optimal starting point exceeds the optimal threshold. In this case, the strategy maximizing RR is to forgo integration altogether, and always choose the more probable alternative immediately upon stimulus onset (i.e., \( DT=0 \)). That is, the task transitions from a discrimination to a signal detection task. The point at which this occurs is determined by setting \( x_0/A = z_0/A = \overline{z}_0 \) in Equation 6.4, and depends on \( \tilde{a} \) and \( D_{total} \).

The critical signal-to-noise ratio at which the optimal starting point reaches the optimal threshold is given by:

\[ \tilde{a}^c = \frac{2\Pi - 1}{1 - \Pi} + 2\Pi \ln \frac{\Pi}{1 - \Pi}. \]  (6.5)
Note that, as for the optimal threshold, this expression depends only on the total delay $D_{total}$ and not individually on $D$ or $D_p$. Similarly, the critical intertrial delay at which the starting point reaches the threshold is given by:

$$D_c^{total} = \frac{2\Pi - 1 + 2\Pi \ln \frac{\Pi}{1-\Pi}}{2\tilde{a}}. \quad (6.6)$$

In other words, for any signal-to-noise ratio that is below $\tilde{a}$ or intertrial delay shorter than $D_c^{total}$, integration does not improve RR, and information about the biased probabilities alone is sufficient for optimal decisions. Figure 18c plots a surface of critical delays $D_c^{total}$ as a function of $\tilde{a}$ ($=A^2/c^2$) and $\Pi$. For delays shorter than $D_c^{total}$, optimal decision makers should immediately choose the most frequent alternative without integrating. Note that $D_c^{total}$ decreases to 0 as $\Pi$ tends to 0.5 for all $\tilde{a}$ since, as we have seen, for unbiased alternatives (and all non-zero values of $D_{total}$, $A$ and $c$) the optimal strategy requires some integration. Also note that $D_c^{total}$ decreases as $\tilde{a}$ ($=(A/c)^2$) increases, reflecting the greater value of integration for more informative stimuli. Figure 17b shows ER as a function of $\Pi$, when $x_0$ and $z$ are set to their optimal values, and signal-to-noise ratio $\tilde{a}$ is fixed. As $\Pi$ increases, ER also increases (since thresholds come in, cf. Figure 17a) as long as decision making still relies on integration. When $x_0$ reaches the optimal threshold (for example, as $\Pi$ is increased), ER exactly matches that achieved by always selecting the more probable alternative, namely:

$$\lim_{x_0 \to z_0} ER = 1 - \Pi \cdot \frac{1}{\Pi}. \quad \text{Then, after the switch to prior-based decision making (i.e., signal detection), as $\Pi$ further increases, ER decreases. That is, the ER at the ‘transition point’ is, nonintuitively, the maximum ER achieved for any value of $\Pi$ (assuming optimal performance). These relationships make interesting predictions regarding empirical performance; for example, that ER should be jointly determined by the probability bias ($\Pi$)
and $D_{total}$ until it reaches the immediate response limit $1-\Pi$ (Equation 6.1). To our knowledge, no studies have yet been conducted that address these predictions.

### 6.6. Relationship of ER to DT for biased stimuli

Analogous to the unbiased case, we can also compute optimal performance curves for RR maximization in the biased case, by re-expressing $\tilde{z}$ and $\tilde{a}$ as functions of ER, DT and $\Pi$ (assuming that participants set their starting points optimally according to Equation 6.1). Doing so (using Equations A.46 and A.47 of the online Appendix A.2.1.3), we obtain:

\[
\tilde{z} = \frac{DT}{1-2\text{ER} + \frac{g(\Pi)}{\ln \frac{1-\text{ER}}{\text{ER}}}}, \quad (6.7)
\]

where

\[
g(\Pi) = (1-2\Pi)\ln \frac{\Pi}{1-\Pi}, \quad (6.8)
\]

and

\[
\tilde{a} = \frac{1-2\text{ER} + \frac{g(\Pi)}{\ln \frac{1-\text{ER}}{\text{ER}}}}{2\text{DT}} \ln \frac{1-\text{ER}}{\text{ER}}. \quad (6.9)
\]

Substituting the above equations into the optimal threshold condition of Equation 6.4 and rearranging terms, we obtain a generalized optimal performance curve for the biased case:

\[
\frac{DT}{D_{total}} = \frac{(1-2\text{ER})\ln \frac{1-\text{ER}}{\text{ER}} + g(\Pi)}{\frac{1-2\text{ER}}{\text{ER}} + \ln \frac{1-\text{ER}}{\text{ER}} - g(\Pi)}. \quad (6.10)
\]

Figure 17c shows the optimal performance curves given by Equation 6.10, for sample values of $\Pi$. 
6.7. Optimal decisions for biased rewards

In this section we describe the optimal starting point and critical delays for the case in which both alternatives are equally likely, but one alternative is associated with a greater reward value than the other.

For TAFC, this case can be described by allocating a fraction \( r \) of reward for each correct response to one alternative and a fraction \( (1-r) \) to the other, where \( r \) lies between \( \frac{1}{2} \) and 1. For example, when \( r=\frac{1}{2} \), the two alternatives are equally rewarded (i.e., unbiased); when \( r=\frac{3}{4} \), the reward for the correct response to the first alternative is three times higher than to the second alternative; and when \( r=1 \), only the correct response to the first alternative is rewarded.

This situation resembles that for biased probabilities, but simulations indicate that there is an important difference regarding the optimal starting point that yields maximal RRs. For biased rewards, the optimal starting point depends also on task delay \( D_{total} \), unlike the case of biased probabilities (Section 6.1) in which starting points depend only on \( c \), \( A \), and \( \Pi \). Based on numerical results, we find that the optimal starting point of the pure DDM for biased rewards appears to obey the following relationships:

\[
x_0 \to \frac{c^2}{2A} \ln \frac{r}{1-r}, \text{ for small values of } D_{total}, \tag{6.11}
\]

and

\[
x_0 \to \frac{c^2}{4A} \ln \frac{r}{1-r}, \text{ for large values of } D_{total}. \tag{6.12}
\]

Comparing the above equations with Equations 6.1 and 6.2, it is apparent that for small \( D_{total} \) the optimal starting point for biased rewards depends on \( r \), exactly as the optimal starting point for biased probabilities depends on \( \Pi \). For large \( D_{total} \), however, the optimal starting
point for biased rewards is halfway between 0 and the optimal starting point for biased probabilities (with equivalent values of \( \Pi \) and \( r \)). Interestingly, this is the same dependence as in Equation 6.2 (describing the starting point minimizing ER for fixed DT). In general, our simulations indicate that the optimal starting point for biased rewards moves monotonically as \( D_{total} \) increases from the value given in Equation 6.11 to that of Equation 6.12. Figure 18a illustrates how optimal starting points depend on \( D_{total} \) with fixed \( A, c, \Pi, \) and \( r \) for biased rewards vs. biased probabilities; Figure 18b illustrates how the starting point varies with \( \Pi \) vs. \( r \) for fixed \( a \) and \( D_{total} \).

As for biased probabilities, we may also derive a surface of critical delays \( D^c_{total} \) for biased rewards that is analogous to that of Figure 18c; for delays shorter than \( D^c_{total} \), optimal decision makers should immediately choose the most highly rewarded alternative without integrating. This surface, found via a numerical search, is shown in Figure 18d. While we are unable to derive an explicit expression analogous to Equation 6.6 for \( D^c_{total} \) in this biased reward case, we verified numerically that \( D^c_{total} \) has the same dependence on signal to noise ratio \( \tilde{a} \) for fixed reward bias \( r \neq \frac{1}{2} \) as it does for probability bias \( \Pi \neq \frac{1}{2} \). That is,

\[
D^c_{total} \sim \frac{1}{\tilde{a}} \quad (6.13)
\]

in both cases. Comparison of Figures 18c and 18d reveals that the critical delay \( D^c_{total} \) for a given signal to noise ratio and reward bias \( r \) is longer than the critical delay for the same signal to noise ratio and probability bias \( \Pi \) equal to \( r \) (the surface in Figure 18d lies above the surface in Figure 18c). We note that the starting point minimizing BR with biased rewards has been analysed by Edwards (1965).
6.8. Relationship to neurophysiological experiments and new predictions

In this section we compare the predictions of our theory with experiments by Platt & Glimcher (1999), who recorded neural activity during biased decisions, and we make new predictions for modified versions of this experiment.

The results of Section 6.1 suggest that when the two alternatives in a TAFC task occur with different probabilities and difficulty is constant across trials, the activities of decision neurons associated with each alternative should be modulated according to Equation 6.1 prior to stimulus presentation. That is, the pre-stimulus activity of neurons representing a particular alternative should depend on the probability of that alternative and its difficulty, as described by Equation 6.1 and illustrated in Figure 16b. In contrast, activity during the decision making period itself should follow the usual DDM, unaffected by the relative probability of the alternatives.

Platt & Glimcher (1999) trained monkeys to perform a TAFC involving visual stimuli in which they indicated responses by making a saccade to one of two locations. The prior probability Π favoring one of the two alternatives varied across blocks of trials from 20% to 80%. They observed that just before stimulus presentation, the activity of neurons corresponding to a given alternative grew linearly with the probability of the alternative being correct in that block of trials (Figure 2b in Platt & Glimcher, 1999). Note that in Figure 16b the dependence of the optimal starting point on outcome probability is close to linear for probabilities ranging from 20% to 80% (see also Figure 18b). Thus, Equation 6.1 provides a reasonable description of the empirical observations under conditions of biased probability.

Platt & Glimcher (1999) also varied the proportion of rewards allocated to one of the alternatives from 20% to 80% on separate blocks of trials (with the same intertrial intervals and task difficulty). They observed that the activity of neurons corresponding to a given
alternative, just before stimulus presentation, also grew linearly with the proportion of reward to the corresponding alternative (Figure 1b in Platt & Glimcher, 1999). Note that in Figure 18b the dependence of the optimal starting point on \( r \) is also close to linear for probabilities ranging from 20\% to 80\%, suggesting that these results can also be explained by Equations 6.11-6.12.

Moreover, Platt & Glimcher (1999) reported that the average slope of pre-stimulus firing rate as a function of probability bias \( \Pi \) was slightly larger than the slope of pre-stimulus firing rate as a function of reward bias \( r \) (11±4.5 (s.e.) vs. 10.5±4.5 (s.e.)). Although the difference is not significant, it is at least qualitatively consistent with the pattern described in Section 6.5. In particular, note that in Figure 18b the slope of the curve relating optimal \( x_0 \) to \( \Pi \) is slightly higher than that relating \( x_0 \) to \( r \) (the difference depends on \( D_{total} \)).

Taken together, the results of Platt & Glimcher’s (1999) manipulation of probability and reward suggest that the activities of neurons in area LIP just before response conform to predictions made by the DDM for RR optimization. To our knowledge, however, no experiments have yet been performed in which probability and difficulty are simultaneously varied across blocks but kept constant within blocks. The DDM predicts that in such an experiment, the modulation of pre-stimulus activity of neurons involved in evidence accumulation (e.g., LIP neurons) should be higher for difficult decisions than for easier (see Section 6.1). Similarly, we are not aware of any experiments that systematically manipulate intertrial delay. As described above, the DDM predicts that pre-stimulus neural activity should be influenced by \( D_{total} \) for biased rewards, but not for biased probabilities.
7. General discussion

In this section we summarize the theory developed in this article and its main predictions, we discuss further extensions of the theory, and consider the significance of our findings for more general issues concerning the mechanisms underlying decision making and cognitive control.

7.1. Summary

This article began with a review of existing models of TAFC performance, and demonstrated that under reasonable assumptions and for particular parameter values all of these models (except the race model) reduce to the simple DDM (Stone, 1960; Laming, 1968; Ratcliff, 1978). In the neural network models, inhibition plays a particularly important role, as this implements differencing of the competing sources of information that is central to DDM. We showed that suitable decay and inhibition values are consistent with data estimated from a typical subject in a TAFC task. A long tradition of work has shown that the DDM is the optimal decision strategy, insofar as it is the fastest to reach a fixed level of accuracy or, alternatively, the most accurate for a given speed of response. The specific tradeoff between speed and accuracy in the DDM is determined by the choice of threshold.

We reviewed two existing theories of how the threshold of the DDM should be chosen to optimize performance (Edwards, 1965; Gold & Shadlen, 2002) using different criteria for optimality (accuracy and RR). We also provided a formal interpretation of a theory that proposes human subjects use a weighted balance of both of criteria (Maddox & Bohil, 1998). We demonstrated that the DDM is the optimal TAFC decision making mechanism with respect to all of these criteria, and that for a given set of parameters (signal, noise, intertrial delay and, where applicable, the emphasis on accuracy), there is a single optimal threshold. We examined how this threshold varies as a function of task parameters, and used this to derive optimal performance curves relating normalized DT to ER for all the criteria.
considered, and generalized these to the extended DDM, which allows for variability of starting points and drift rates.

Finally, we extended our analyses to the case of biased probabilities and rewards. We showed that predicted effects conform to previously reported observations concerning behavioural performance (Link, 1975; Ratcliff & Smith, 2004) as well as the activity of neural accumulators under such conditions (Platt & Glimcher, 1999). We also computed optimal performance curves relating speed and accuracy for the biased probability case.

The main predictions of our theory consider the paradigm in which participants are required to maximize RR. They are: (i) The decision threshold should depend on task parameters in a specific way summarized in Figure 12. (ii) Decision threshold and behavioural statistics should depend on the sum $D_{\text{total}}$ of all task related delays, as opposed to, for example, being differentially influenced by error-related penalty delays. (iii) For a given ER, the DT achieved by participants should be proportional to $D_{\text{total}}$ and their ratio $\frac{\text{DT}}{D_{\text{total}}}$ should be equal to the value given by the optimal performance curve (Figure 13bc, heavy line; Figure 17c). (iv) In case of biased decisions, when the delay is below the critical value $D^c_{\text{total}}$, participants should stop integrating, and always immediately choose the most-probable or most-rewarded alternative. Furthermore, for a given probability and/or reward bias, the specific delay at which this occurs should vary with signal-to-noise ratio $\tilde{a}$ (Figures 18c and d).

### 7.2. Extensions

The theory of optimal performance in TAFC tasks developed in this article has several natural extensions that may contribute to a broader understanding of the mechanisms underlying human decision making.

#### 7.2.1. Multiple choice decisions
Perhaps the most important limitation of the current theory is that it pertains to decisions involving only two alternatives. Even simple decision processes often involve many more alternatives (e.g., retrieving an item from memory, searching a visual display for a target item, etc.). Numerous procedures have been devised for deciding among multiple alternatives (Eisenberg, 1991; Wetherill & Glazebrook, 1986), most involving combinations of SPRT’s between the various alternatives. Following the seminal work of Lorden (1977), Dragalin et al. (2000) have shown that, in the limiting case of arbitrarily low ER, optimality comparable to that of the SPRT (that is, the fastest response for fixed accuracy) is achieved by two procedures known as the multi-sequential probability ratio tests (MSPRT). The first MSPRT procedure calculates for each alternative the ratio of its likelihood and the sum of likelihoods of all alternatives considered; the decision is made whenever any of these ratios exceed a certain threshold. Recently Bogacz & Gurney (in press) demonstrated that that many aspects of the anatomy and physiology of a circuit involving the cortex and the basal ganglia are precisely those required to implement this version of the MSPRT. The other MSPRT, often referred to as the max-versus-next procedure, compares the alternative receiving the most evidence with the next-most supported alternative, and terminates when the difference between these exceeds a fixed value (corresponding to the threshold in the SPRT). The max-versus-next procedure has been implemented in the word-identification model by Ratcliff & McKoon (1997) and provided a good fit to experimental data.

In contrast to the MSPRT, Usher & McClelland (2001) proposed that a multiple decision process can be modeled by a direct extension of the mutual inhibition model, in which each alternative inhibits and receives inhibition from all others (equivalent to a multiple choice extension of the pooled inhibition model). This implements a max-vs-average procedure, in which the evidence favoring the most supported alternative is compared to the average of the evidence in support of all other alternatives. Usher & McClelland (2001) showed that this
performed best among several alternative models. Recent analytical and simulation work by McMillen & Holmes (2006) suggests that under certain conditions (ER ~ 10%) the max-vs-average procedure closely approximates the performance of the MSPRT, although as ER approaches zero it is significantly worse. Moreover, multi-unit mutual inhibition models approximate such a procedure. Additional work is needed to fully characterize the dynamics of this procedure. As this work proceeds, it will be interesting to consider how this compares to models that have been developed for multiple choice decision making in domains such as memory retrieval (e.g., Diller et al., 2001; Anderson, 1983) and lexical decisions (e.g., Wagenmakers et al., 2004), and to the work of Roe et al. (2001). The latter generalized the model combining mutual and feedforward inhibition (described in Equation 3.18) to three alternatives and used it to explain a number of fundamental empirical results described in multi-alternative preference literature. An important avenue for future inquiry will be to extend the methods and analyses described in this article to apply to the domain of multi-choice decision making.

7.2.2. Optimization and cognitive control

The evidence reviewed in this article suggests that, at least under some conditions, human decision makers can adapt their behavior to approximate optimal performance. This ability can be considered as a simple, but well defined example of cognitive control. Broadly speaking, cognitive control is the ability to guide and adapt behavior to meet desired goals. The challenge, in psychology and neuroscience, has been to operationalize these constructs in terms that lead to a precise understanding of the mechanisms involved, and can generate quantitatively detailed predictions regarding task performance. The DDM provides a framework for doing so, by defining a specific decision making mechanism, with critical parameters that have a direct impact on performance: namely, the drift rate (and its variability), the starting point of integration (and its variability), and the decision threshold.
These parameters can be interpreted in terms of psychological processes: signal strength, stimulus difficulty, and the allocation of attention (drift); expectancy biases (starting point); and the speed-accuracy tradeoff (threshold). Each of these psychological constructs has been closely associated with control. Within the framework of the DDM, we can think of control as the adjustment of these parameters to optimize performance. This formulation calls into focus important questions: To what end are these parameters adjusted? That is, what are the objectives of performance, and how do we know when these are being met? The framework provided by the DDM allows us to address these questions in a formally explicit manner, by defining an objective, or “utility” function that the individual seeks to maximize. We can then more precisely define control as the mechanisms that monitor performance by evaluating this utility function, and adjusting parameters of the decision process to maximize its value. While performance may never actually achieve the theoretical optimum, the ability to define this provides valuable structure for constructing theories about the mechanisms involved in control, and generating detailed quantitative predictions that can be tested in empirical studies.

In particular, this framework allows us to formally identify and characterize the factors involved in control — that is, the factors that influence an organism’s ability to optimally meet its goals. These factors fall roughly into three broad categories: 1) parameters of the decision process that can be adjusted to optimize performance; 2) criteria used to guide such adaptations (i.e., the cost functions that the organism uses to determine how effectively it is meeting its goals); and 3) algorithms used to carry out these adaptations. This framework suggests specific directions for extending the DDM to address factors in each of these categories.

**Parameters.** We have focused on TAFC tasks in which signal strength (drift rate) remains constant over the course of a single trial. However, if signal-to-noise ratio varies within a
trial, optimal performance requires adjusting the “gain” of the integrator (i.e., its sensitivity to inputs), as suggested by classical signal processing theory (Papoulis, 1977). We have begun to characterize the trajectories for dynamic gain adjustment that optimize performance in the DDM and various related neural network models (Brown et al., 2005).

Work on dynamic gain adjustment has also identified, and begun to address an interesting tension between the optimality of a single layered neural network that implements the DDM (such as the models considered in this article) and the reality that such a network is almost certainly embedded in a multilayered system within the brain (required, among other things, to translate the outcome of the decision process into overt behavior). Numerical optimization studies suggest that dynamic gain adjustment, mediated by brainstem neuromodulatory systems that are triggered by the outcome of the relevant DDM implemented in a particular layer, can recover a substantial fraction of the performance of a single-layered system (Brown et al., 2004, Gilzenrat et al., 2004).

In the present article, we assumed that the experimental delay $D$ does not directly affect the parameters of the DDM, beyond determining what their optimal values would be. However, it is possible that $D$ interacts with attentional effects, and hence the drift $A$. Delay $D$ may also influence biases or variability in the starting point (cf. “preparatory cycle” in Cho et al. 2002, Brown & Holmes, 2001), as well as trial-to-trial threshold adaptation, as both of these may plausibly require a certain intertrial duration to take effect.

Another critical factor is attention. As noted earlier, assumptions of linearity in processing can be justified by assuming that attention acts to place non-linear integrators in the most sensitive, and approximately linear range of their response functions (e.g., Cohen et al., 1990). This suggests that fluctuations in attention, both between and within trials, could have a significant impact on performance. Indeed, such effects have been exploited to model the
effects of attention on the dynamics of performance in a variety of paradigms (e.g., Grossberg & Stone, 1986; Cohen et al., 1992; Servan-Schreiber et al., 1998; Sperling et al., 2001). The present work provides a framework for analyzing the influence of such factors, identifying optimal protocols for attentional adjustment, and using these as a guide for the design of further empirical studies. For example, there are at least two ways in which attentional mechanisms could modulate processing: one is by gain adjustment (along the lines suggested above); the other is by additive bias adjustment that, interacting with the non-linearity of the response function, can change the dynamic range of the integrator. There has been considerable debate in the behavioral, neurophysiological, and modeling literatures as to which of these best describe attentional influences. However, to our knowledge, there is no analytic treatment of this issue within the optimality framework. The present framework allows for such a treatment, in terms of the influence on drift rate, that can bridge between abstract characterization of behavior and neural implementation. Finally, as we noted earlier, periodic shifts in attention may explain the variability in drift rate that has been observed empirically under a number of conditions. Such shifts of attention may reflect the effects of exogenous factors, or a fundamental tension between the exploitation of opportunities for reward in the current task and the value of periodically exploring the possibility of other opportunities for reward (e.g., Aston-Jones & Cohen, 2005).

Criteria for optimization. Optimality theory requires that we specify an objective or cost function that defines what is being optimized. This has the virtue of forcing us to be precise, but the peril of being too narrow in our assumptions. In this article, we considered four criteria involving reward and accuracy in various combinations. The importance of reward is plausible given that rewards are, by definition, a form of utility valued and actively sought by the organism. The value of accuracy also seems self-evident, given that it is often a proxy for reward (especially when time-pressure is not a critical factor). However, these criteria have
not taken account of other relevant factors, such as the potential costs associated with behavior (e.g., uncertainty of outcome, energy expenditure, or opportunity costs associated with the pursuit of other types of reward). In particular, behavior that may seem suboptimal under the criteria and tasks considered here may in fact be optimal when richer criteria or task environments are considered. For example, the tendency to explore previously unrewarded alternatives is not optimal in a static environment, but could yield higher rewards in a changing environment (Sutton & Barto, 1998). Such factors can readily be explored within the context of the optimization framework, and may help guide further empirical research (e.g., Aston-Jones & Cohen, 2005; Yu & Dayan, 2005).

Another important consideration is the source of information available to the organism about the outcome of performance, and used to adapt behavior. In the simplest case, earned rewards or information about accuracy are immediately available, and can be used directly by the mechanisms of adaptation to optimize the parameters of the decision process. But what about situations in which reward delivery is delayed, or information about accuracy is not immediately available? For example, in most human studies payments based on performance are not made until the end of the experiment. Other forms of feedback can serve as proxies for reward (e.g., a visual display of the dollar amount won). However, in many circumstances people seem able to adapt their behavior even when no explicit feedback is provided (e.g., Laming, 1979; Gratton et al., 1988; Botvinick et al., 1999). Measurements of brain activity (using both scalp electrical recordings and functional MRI) have produced strong evidence of neural signals that are responsive to both endogenous and exogenous sources of information about performance (e.g., Falkenstein et al., 1991; Gehring et al., 1993; Carter et al., 1998; Holroyd et al., 2004; Nieuwenhuis et al., 2004). In modeling work, we have begun to explore neurally plausible mechanisms for performance monitoring (e.g., Botvinick et al., 2001; Yeung et al., 2004; Holroyd et al., 2005), that can provide internal indices of performance and
use these to adaptively adjust behavior without access to explicit information about reward or accuracy. However, to date, such work remains disconnected from more principled, and formally rigorous considerations of optimal performance. By relating models of performance monitoring to the ones examined in this article, it may be possible to derive formal principles that better justify existing models, and help guide their further development.

*Algorithms for adaptation.* The foregoing considerations assume mechanisms for translating information about the outcome of performance into adjustments of parameters in the decision making process that produce improvements. In Section 5 we suggested the possibility that reinforcement learning algorithms are a natural candidate, given that they are designed to maximize reward (e.g., Simen et al., 2005, 2006). It remains to be determined, however, to what extent such algorithms can reproduce detailed features of sequential adjustment effects that have been observed in the empirical literature (e.g., Laming, 1979; Gratton et al., 1988; Botvinick et al., 1999; Kerns et al., 2004). It is also possible other mechanisms may be involved in behavioral adaptation. For example, we have shown that simple alternation and repetition detectors, using exponentially-decaying averages of prior response history, can produce adjustments in starting points in a mutual inhibition model that accurately simulate human RT and ER effects (Cho et al., 2002). Similarly, Botvinick et al. (2001) used a simple exponentially-decaying average of response conflict to adjust attentional influences and thresholds, simulating sequential adjustment effects such as post-error slowing. An important goal for further work will be to determine whether these mechanisms converge on parameters that produce optimal performance and, if not, how mechanisms that do so compare in fitting empirical data.
7.3. Conclusions

We hope in this article to have contributed to the development of a framework for the formal analysis of human decision making behavior. We have used this framework to examine how existing models relate to the optimal TAFC decision making process – the DDM – and have extended the analysis of optimality to address maximization of reward and accuracy. This has generated insights into issues relevant to both psychology, neuroscience, and their shared interests. For example, theories of optimal performance make strong predictions about how human decision makers seeking to maximize reward should trade speed against accuracy – a phenomenon that pervades experimental research. At the neural level, our analyses make predictions about underlying mechanisms and how they may implement optimal performance. Confirmation of these predictions would at once strengthen confidence in our framework, and help adjudicate between competing neural network models of behavior. Most importantly, we believe that our approach offers hope that complex neural systems responsible for behavior can, under reasonable assumptions, yield to simpler, more abstract descriptions that are tractable to formal analysis, and yet remain faithful to and can be used to make predictions about behavior. Success in this effort promises a principled understanding of the neural mechanisms underlying psychological function and overt behavior, akin to the one that has emerged from physics relating atomic structure to the macroscopic properties of matter.
**Acknowledgements**

This work was supported by the following grants: NIH P50 MH62196, DOE DE-FG02-95ER25238 (P.H.), EPSRC EP/C514416/1, NSF Mathematical Sciences Postdoctoral Research Fellowship (held by J.M. and E.B.), and the Burroughs-Wellcome Program in Biological Dynamics and Princeton Graduate School (E.B.). We thank Peter Hu for collecting data in the experiment described in Section 3.1. We thank Josh Gold for providing us with the program to generate the “moving dots” stimuli, and for discussion. We thank Tyler McMillen for contribution to the section on multiple decisions, and Mark Gilzenrat, and Angela Yu, for discussions and valuable insights and ideas that have driven the work reported in this article. We also thank the reviewers for their very detailed reading of the manuscript and constructive comments.
References


Bogacz, R., & Gurney, K. (in press). The basal ganglia and cortex implement optimal decision making between alternative actions. Neural Computation.


Appendix B. Experimental methods

Participants. These were 20 adults (9 males and 11 females) recruited via announcements posted around Princeton University campus. They were predominantly Princeton undergraduate and graduate students (average age: 20 years). Participants were paid 1 cent for each correct choice. To further increase motivation, participants were informed that the one who earned the most would receive an additional prize of $100, that was awarded at the end of the experiment. The experiment was approved by the Institutional Review Panel for Human Subjects of Princeton University and all participants expressed written consent for participation.

Stimuli and apparatus. We used the same stimuli that have been used in other studies of decision making (e.g., Gold & Shadlen, 2003; Palmer et al., 2005). The display was a field of randomly moving dots all of which appeared within a 5° circular aperture in the center of the screen. Dots were white squares 2 by 2 pixels (0.7° square) displayed against a black background, with a density of 16.7 dots/degree²/s (6 dots per frame). On each trial, a fraction of the dots moved in a single direction over time, corresponding to that trial’s correct direction, while the remaining dots were randomly repositioned over time. On each frame, 11% of the dots were independently selected as the coherently-moving dots, and were shifted 0.2 deg from their position for each 40 ms (3 video frames) elapsed, corresponding to a speed of 5 deg/sec (either leftward or rightward). The remaining dots were re-plotted in random positions on each frame. The display was generated in MATLAB on a Macintosh computer, using the Psychophysics Toolbox extension (Brainard, 1997; Pelli, 1997) and software written by Josh Gold.

Procedure. Participants were instructed to gain as many points as possible by deciding, for each trial, whether the prevailing motion of the dots was left or right, and to indicate their
responses by pressing the “M” key (rightward motion) or the “Z” key (leftward motion) on a standard keyboard (the mapping of keys to the right and left responses was not counterbalanced across participants). After each correct response participants were informed by a short beep that the response was correct and they had scored a point. There was no feedback after incorrect responses. After each response participants were required to release the key, and there was a delay $D$ before presentation of the next stimulus ($D$ was kept constant within each block, but varied across blocks). On some blocks (see below) an additional delay $D_p$ was imposed after error responses. During these delay intervals the current score was displayed in the center of the screen.

**Design.** Trials were blocked by delay condition. This was manipulated to test predictions made by the DDM (discussed below in Section 5.2). There were four delay conditions: (1) $D=0.5s$; (2) $D=1s$; (3) $D=2s$; and (4) $D=0.5s$ and $D_p=1.5s$ (in the first three conditions $D_p=0$). Each block lasted 7 minutes (block length was limited by fixing its overall duration rather than by the number of trials completed within it). One block of trials was run for each delay condition, except for condition 3 ($D=2s$) for which two blocks were run in order to permit a sufficient number of trials for analysis (because in this delay condition trials lasted longer). Hence, in total the experiment consisted of five blocks of trials randomly ordered across subjects. Before the start of the experiment participants had three blocks of practice, in which they did not receive money for correct choices. Participants were informed about the number of blocks, their duration, and that blocks differed in delays, but were not told the exact durations of delays.
Figure 1. a) Sample reaction time distribution of in two alternative force choice task; based on data from a sample participant of Experiment 1 described in Section 3.1 (trials in condition $D=1$). b,c) Cartoon of typical peri-stimulus time histograms of neuronal activity during ‘the moving dots task’. The figure does not show the actual data, but it is a sketch based on data described by Britten et al. (1993), Shadlen & Newsome (2001), and Schall (2001). Horizontal axes show time from stimulus onset. Vertical axes indicate firing rate. Representative firing rates are shown for stimulus with coherent leftward motion. b) Firing rate of neurons in the area MT: gray line represents a typical neuron selective to leftward motion, and black line for rightward motion. c) Firing rate of neurons in the area LIP: gray line represents a typical neuron selective for leftward saccades, and black line for rightward saccades.
Figure 2. Examples of the time evolution of variables in decision models. a) The pure drift diffusion model (DDM). Horizontal axis denotes time; vertical axis denotes the accumulated difference $x$ between the evidence supporting the correct and incorrect alternatives. The model was simulated for 100,000 trials using the Euler method with timestep $\Delta t = 0.01$ and the following parameters: drift $A = 1$, noise $c = 1$, threshold $z = 1$. Each path corresponds to one sample simulated decision process. The histograms outside the thresholds show proportions of trials reaching the threshold after different intervals. 
b) An example of the evolution of the mutual inhibition model, showing $y_1$ and $y_2$ as functions of time. 
c) The phase- or state space of the mutual inhibition model. Horizontal axis denotes the activation of the first decision unit; vertical axis denotes the activation of the second decision unit. The path shows the decision process from stimulus onset (where $y_1 = y_2 = 0$) to reaching a decision threshold (decision thresholds are shown by dashed lines). The mutual inhibition model was simulated for the following parameters: $I_1 = 4.41$, $I_2 = 3$, $c = 0.33$ (parameters of the inputs correspond to those estimated from the participant shown in Figure 5, via Equation 3.8), $w = k = 10$, $Z = 0.4$. 

[Diagram of drift diffusion model, mutual inhibition model, and phase space diagram]
Figure 3. Architectures of decision models: a) race model (continuous version of Vickers, 1970), b) mutual inhibition model (simplified from Usher & McClelland, 2001), c) feedforward inhibition model (simplified from Ditterich et al., 2003), d) pooled inhibition model (simplified from Wang, 2002). Arrows denote excitatory connections, line with filled circles denotes inhibitory connections. In panel d) discs containing small open circles denote populations of excitatory neurons, and with filled circles – populations of inhibitory neurons.
Figure 4. Relationships among decision models. An arrow between two models indicates that the computations of the first model simplify to computations of the second under conditions in the arrow’s label. The horizontal dotted lines separate classes of models with different numbers of dimensions (1D, 2D and 3D), so that arrows crossing the dotted lines indicate simplifications involving dimension reduction. A more detailed taxonomy of diffusion and race models (distinguishing between different variants of these models) is given in Figure 1 of Ratcliff & Smith (2004).
Figure 5. Fit of extended drift diffusion model (DDM) to a sample participant of the experiment. Each panel corresponds to one experimental delay condition labelled on the top of the panel. Circles indicate experimental data: in each panel, the left column of circles corresponds to error trials, and the right to correct trials. The horizontal axes show the probability of response (i.e., error rate for the left column and probability of correct response for the right column). The vertical axes show the values of five quantiles of reaction time: 0.1, 0.3, 0.5, 0.7, 0.9. Error bars indicate standard error. The approximately horizontal lines connect predictions of fitted DDM for each of five quantiles. The following parameters of the extended DDM were estimated: $s_x=0.14$, $m_A=1$, $s_A=0.31$, $c=0.33$, $z_1=0.16$, $z_2=0.19$, $z_3=0.22$, $z_4=0.26$, $T_0=0.37$ s.
a) Decay > Inhibition $(\lambda < 0)$
b) Decay = Inhibition $(\lambda = 0)$
c) Decay < Inhibition $(\lambda > 0)$

Figure 6. Vector fields for the mutual inhibition model. In all plots $I_1 = 2$, $I_2 = 1$. Inhibition ($w$) and decay ($k$) have different values in different panels: a) $w = 0.5$, $k = 1.5$; b) $w = 1$, $k = 1$; c) $w = 1.5$, $k = 0.5$. See text for discussion.
Figure 7. Geometry of mutual and feedforward inhibition models. a) Transformation from $y$ to $x$ coordinates. Small circle denotes a sample state of the network. Labels $y_1$ and $y_2$ indicate the activation levels of decision units. Labels $x_1$ and $x_2$ indicate the values of $x$ (difference and sum) coordinates for this state of the network. b) Thresholds for the balanced mutual inhibition model. c) Thresholds for the feedforward inhibition model with $u = 1$. In panels b and c, the grey arrow denotes the decision line, or the line along which the diffusion occurs. Lines marked by $Z$ denote the thresholds of the mutual and feedforward inhibition models. Labels A, B, C, D, O are used in the main text.
Figure 8. Similarity in ER and DTs between extended diffusion model and balanced mutual inhibition model. Top panels as explained in legend of Figure 5. Approximately horizontal solid lines show the ER and DTs of diffusion model fitted to condition $D=0.5$ $D_p=1.5$ of the participant from Figure 5 (these lines in all top panels are the same as in Figure 5d). Dashed lines correspond to values obtained from simulations of the extended mutual inhibition model (10,000 iterations, for different values of decay and inhibition shown above panels), and all other parameters defined as in Section 3.2 to give equivalence with the extended diffusion model. Specifically: $m_{I1} = 1.42$, $m_{I2} = 0$, $c = 0.33$, $s_I = 0.22$, $s_y=0.096$. Note that for higher values of decay and inhibition, the match between the models improves. Note difference in scales in top panel c. Bottom panels show the states from 1000 trials, at time $t=T_0+200ms$. Long-dashed lines show decision threshold. Short-dash lines show the decision line.
Figure 9. Error rates (ER) for the O-U model.  a) ER in the interrogation protocol for decisions made at different times, and for different values of $\lambda$, according to Equations 2.8 and 2.12, with $A = 2.19$, and $c = 1$ (parameters estimated from data from middle-difficulty condition of participant S1 of experiment described in Usher & McClelland, 2001).  b) ER in the free-response protocol.  O-U models were simulated for values of parameters corresponding to the evolution along the decision line in Figure 11; that is, for $A = 2.19$, $c = 1$.  O-U models with 3 different values of $\lambda$ were simulated: –6, 0, 6 shown by the three thick curves in different pattern (see legend insert).  The value of the decision threshold is shown on the horizontal axis.  For each set of parameters, the model was simulated 1000 times, and the proportion of incorrect decisions is shown by solid lines.  The thin line denotes the proportion of trials for which the decision has not been reached by time $t = 100$ in the case $\lambda = -6$.  The ER for $\lambda = 0$ and $\lambda = 6$ very closely match the values predicted by Equations 2.8 and A.55 in the Appendix A.2.2.1, respectively (although theoretical predictions are not shown for clarity of the figure).  c) ER of the extended O-U model with the variability of starting point in the interrogation paradigm.  The model was simulated for the following parameters: $m_A = 2.19$, $s_A = 0$, $c = 1$, (parameters estimated from data from middle-difficulty condition of participant S1 of experiment described in Usher & McClelland, 2001), $s_x = 0.5$, at interrogation time $t = 0.1s$.  For each value of $\lambda$ the O-U model was simulated 20,000 times; the average ER is shown by the solid line.  Error bars shows the standard error of the mean.
Figure 10. Performance of mutual and feedforward inhibition models. During all simulations the following parameters were kept fixed: $m_{I1} = 1.42$, $m_{I2} = 0$, $c = 0.33$, $s_I = 0.22$ (the same values as in Figure 8), $s_y = 0$. Panels a, b and c show decision time (DT) for the threshold set such that error rate (ER) = 2%. a, b) DT for the mutual inhibition model for different values of inhibition. In Panel (a) inhibition is kept constant at $w = 10$ and decay ($k$) is varied (shown on horizontal axis). In Panel (b) inhibition is equal to decay ($w = k$), and they are varied together. c) DT for the feedforward inhibition model as the weight of the inhibitory connections ($u$) is varied (shown on horizontal axis). For each set of parameter values, the threshold was increased from zero in steps of 0.01 until the model reached an ER less than or equal to 2%. For each value of the threshold 10,000 trials were simulated. The error bars indicate standard error of mean DT estimation. d) Reward rate achieved by the mutual inhibition model with optimal thresholds at different levels of decay. During simulations the following parameters were kept constant: $w = 10$ and the delays between trials were equal to: $D = 2$, $D_p = 0$. The decay rate $k$ is shown on the horizontal axis. For each set of parameter values, the threshold was chosen that maximized the reward rate (by simulation), and the model was run for 100 simulated blocks of 10,000 trials each. The lines on the charts connect the mean reward rates achieved by the model and the error bars show the standard error across the blocks.
Figure 11. Distributions of solutions to the mutual inhibition model with different values of inhibition (w) and decay (k). Three mutual inhibition models were simulated. In all the models the signal and noise from the environment were the same, with $I_1 = 3 + \sqrt{2} \times 2.19$, $I_2 = 3$, and $c = 1$ (parameters corresponding to the parameters of pure drift diffusion model $A = 2.19$; $c = 1$ estimated from data from middle-difficulty condition of participant S1 of experiment described in Usher & McClelland, 2001). Dashed lines indicate $x_1 = 0$ axis, dividing correct and incorrect decisions in the interrogation paradigm, and decision lines are shown solid. The models shown in the three panels differed in the ratio of inhibition to decay: a) $k = 11.5$, $w = 8.5$ ($\lambda < 0$); b) $k = 10$, $w = 10$ ($\lambda = 0$); and c) $k = 8.5$, $w = 11.5$ ($\lambda > 0$). Each model was simulated 1000 times, and the plots show the states (i.e., the activities of decision units) of the simulated models, with each dot corresponding to one simulation. The horizontal axes denote the activations of the first (correct) unit, and the vertical axes denote the activations of the second (incorrect) unit. The states of the models are shown at three moments in time: top panels for $t = 0.5s$, middle panels for $t = 1s$, and bottom panels for $t = 1.5s$. Many states have left the region plotted in the bottom right ($\lambda > 0$) panel.
Figure 12. The optimal thresholds for the pure drift-diffusion model (DDM). Each row of panels shows thresholds optimizing a different criterion (indicated on the left side of the figure). Each column of panels shows the dependence of the optimal threshold on one task parameter (indicated on top and the bottom of the column), with other task parameters held fixed at values chosen from the following: $A=1$, $c=0.33$ (corresponding to values estimated from a sample participant; see Figure 5), $D_{\text{total}}=2$ (corresponding to one of the delay conditions in the task performed by the participant), $q=0.5$ (chosen arbitrarily). The values of the optimal thresholds were found numerically. Values at the angles are equal to the rate of increase (i.e., the slope) of the optimal threshold function for the varied parameter equal to zero and are derived analytically. The arrows indicate the values to which the optimal threshold converges as the parameter on the horizontal axis goes to infinity. Panels are omitted where criteria do not depend on those parameters.
Figure 13. The optimal performance curves for the drift diffusion model. Horizontal axes show the error rate (ER) and vertical axes show the normalized decision time (DT), i.e. DT divided by accuracy weight $q$ in panel a, and by total delay $D_{total}$ in panels b and c. a) Optimal performance curves for Bayes risk. b, c) Optimal performance curves for reward accuracy and modified reward rate criteria. Each curve corresponds to different value of $q$ ranging from -0.2 (bottom curves) to 0.5 (top curves) in steps of 0.1. The thick line (identical in panels b and c) corresponds to $q = 0$; that is, the thick line is the unique optimal performance curve for the reward rate RR.
Figure 14. Optimal thresholds and optimal performance curves for the extended drift diffusion model (DDM). a, b) Dependence of the RR-maximizing threshold for the extended DDM on drift variability (panel a) and starting point variability (panel b), with other parameters held fixed. The values of the varied parameters are shown on horizontal axes, while the values of the fixed parameters were equal to: $s_x=0.14$, $m_d=1$, $s_A=0.31$, $c=0.33$, (corresponding to values estimated from a sample participant; see Figure 5), $D_{total}=2$ (one of the delay conditions performed by the participant). c, d) Optimal performance curves for extended DDM, obtained numerically by varying noise parameter $c$, while other parameters were kept fixed: $m_d=1$, $s_x$ and $s_A$ as shown in figure key, $D_{total}=2$ in panel c, and $D_{total}=0.5$ in panel d.
Figure 15. The reward rate for the drift diffusion model as a function of the normalized threshold $\tilde{z}$ (shown on horizontal axis), for $D=1s$, $D_p=0s$, $T_0=0s$, and the values of $\alpha$ corresponding to those estimated from 3 sample participants of the experiment described in Section 3.1 (the values are shown in the figure key; $\alpha = 9.1$ corresponds to the participant whose performance was illustrated in Figure 5). $\tilde{z}_o$ indicates the optimal normalized threshold for the case $\alpha = 14.1$, and $\varepsilon$ denotes a small deviation. Note that underestimation of the threshold by $\varepsilon$ causes greater decrease in the reward rate than overestimation.
Figure 16. Optimal starting point for biased decisions. a) Mean decision time (DT) (y-axis) for the biased-probability drift diffusion model (DDM) at fixed accuracy (error rate (ER)=2%) and different starting points (x-axis). DTs were obtained numerically using equations A.44 and A.45 from online Appendix A for drift $A = 1$, noise $c = 0.33$ (estimates from the participant shown in Figure 5), and probability of one of the alternatives $\Pi = 0.7$. Dashed line indicates position of the optimal starting point predicted by Equation 6.1. b) Dependence of the starting point of the pure DDM (scaled, see vertical axis) on the prior probability $\Pi$ of the alternative with positive drift $A$ (horizontal axis). Solid line shows the starting point minimizing DT for fixed ER, obtained from Equation 6.1. Dashed line shows the starting point minimizing ER for fixed threshold, obtained from Equation 6.2. Circles indicate starting points estimated from experiment of Laming (1968) by Link (1975; Table 2), squares indicate starting points estimated by Link (1975) from his experiment (Link, 1975; Table 6), and stars indicate starting points estimated from the experiment of Van Zandt (2000) by Ratcliff & Smith (2004; Table 8; the value of the drift was taken as the average of mean drifts for the two alternatives: $(0.2161+0.3659)/2)$. 
Figure 17. Maximization of reward rate for biased decisions a) The optimal threshold $z_o$ (solid line) and the optimal starting point $x_0$ (dotted line) as functions of prior probability $\Pi$ (x-axis), for fixed parameters $A=1$, $c=0.33$ (values corresponding to the data of Figure 5), $D_{total}=1$. There are no data points for $\Pi > 0.93$, because for these probability biases, the optimal strategy does not involve integration of input. b) Error rate (ER) (y-axis) as a function of probability bias (x-axis) for fixed parameters $A=1$, $c=0.33$, $D_{total}=1$, and for optimal values of $z$ and $x_0$ shown in panel a. c) The optimal performance curves generalized for biased decisions. The five curves correspond to different prior probabilities $\Pi$. Horizontal axis shows ER, vertical axis shows decision time normalized by $D_{total}$. The right endpoint of each curve corresponds to immediate responding with ER=1-$\Pi$. 
Figure 18. Comparison of optimal decision strategies for biases in the probability ($\Pi$) and reward proportion ($r$) allocated to the two alternatives. 
a) The optimal starting points $x_0$ as a function of $D_{total}$ for fixed $A=1$, $c=0.33$ (values corresponding to the data of Figure 5), and with $\Pi=0.7$, $r=0.5$ for the solid line and $r=0.7$, $\Pi=0.5$ for the dotted line. 
b) The optimal starting point as function of $\Pi$ (solid line) and $r$ (dotted line) for fixed $A=1$, $c=0.33$, and $D_{total}=2$. 
c) Surface of critical total delays ($D_{total}$), as a function of signal-to-noise ratio $\tilde{a}$ and biased prior probability $\Pi$ for one of the two alternatives. 
d) Surface of critical total delays ($D_{total}$), as function of $\tilde{a}$ and biased reward proportion $r$ allocated to one of the two alternatives.
<table>
<thead>
<tr>
<th>Property</th>
<th>Variability</th>
<th>None</th>
<th>Variability of drift</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Optimal performance</strong></td>
<td>Interrogation</td>
<td>$k=w$</td>
<td>$k=w$</td>
</tr>
<tr>
<td></td>
<td>Free-response</td>
<td>$k=w$</td>
<td>$k=w$ (suggested by simulations)</td>
</tr>
<tr>
<td></td>
<td>Free-response</td>
<td>$k,w \to \infty$</td>
<td>$k,w \to \infty$</td>
</tr>
<tr>
<td><strong>ER→0</strong></td>
<td>Interrogation</td>
<td>$k=w$</td>
<td>Not attainable</td>
</tr>
<tr>
<td></td>
<td>Free-response</td>
<td>$k \geq w$</td>
<td>Not attainable</td>
</tr>
</tbody>
</table>

Table 1. Conditions on parameters $k$ (decay) and $w$ (inhibition) of the mutual inhibition model for optimal performance (i.e., minimal error rate (ER) in the interrogation protocol or maximal reward rate for free response), and for vanishing ER in the limit of long decision time (interrogation protocol) or large thresholds (free response).