Stochastic Choice and Optimal Sequential Sampling[*]

Drew Fudenberg[†]   Philipp Strack[‡]   Tomasz Strzalecki[§]

May 20, 2015

Abstract

We model the joint distribution of choice probabilities and decision times in binary choice tasks as the solution to a problem of optimal sequential sampling, where the agent is uncertain of the utility of each action and pays a constant cost per unit time for gathering information. In the resulting optimal policy, the agent's choices are more likely to be correct when the agent chooses to decide quickly, provided that the agent's prior beliefs are correct. For this reason it better matches the observed correlation between decision time and choice probability than does the classical drift-diffusion model, where the agent is uncertain which of two actions is best but knows the utility difference between them.

[†]Department of Economics, Harvard University
[‡]Department of Economics, University of California, Berkeley
[§]Department of Economics, Harvard University

1 Introduction

In laboratory experiments where individuals are repeatedly faced with the same choice set, the observed choices are stochastic—individuals don't always choose the same item from a given choice set, even when the choices are made only a few minutes apart.[1] In addition, individuals don't always take the same amount of time to make a given decision—response times are stochastic as well. Our goal here is to model the joint distribution of choice probabilities and decision times in choice tasks, which we call a *choice process.*

We restrict attention to the binary choice tasks that have been used in most neuroscience choice experiments, and suppose that the agent is choosing between two items that we call left (l) and right (r). In this setting we can ask how the probability of the more frequent (i.e., modal) choice varies with the time taken to make the decision. If the agent is learning during the decision process, and is stopped by the experimenter at an exogenous time, we would expect the data to display a *speed-accuracy tradeoff,* in the sense that the agent makes more accurate decisions (chooses the modal object more often) when given more time to decide. However, in many choice experiments there is instead the opposite correlation: slower decisions are less likely to be modal (Swensson, 1972; Luce, 1986; Ratcliff and McKoon, 2008).

To explain this, we develop a new variant of the drift diffusion model (DDM); other versions of the DDM have been extensively applied to choice processes in the neuroscience and psychology literatures.[2] The specification of a DDM begins with a diffusion process Z_t that represents information the subjects is receiving over time, and two disjoint stopping regions S_t^l and S_t^r. The agent stops if at some point in time t it happens that $Z_t \in S_t^l$ (in which case he chooses l) or $Z_t \in S_t^r$ (in which case he chooses r); otherwise the agent continues. Because the evolution of the diffusion depends on which choice is better, the model predicts a joint probability distribution on choices and response times conditional on the true state of the world, which is unknown to the agent.

The oldest and most commonly used version of the DDM (which we will refer to as *simple* DDM) specifies that the stopping regions are constant in time, i.e., $S_t^l = S^l$ and $S_t^r = S^r$,

---

[1] See Hey (1995, 2001), Ballinger and Wilcox (1997), Cheremukhin, Popova, and Tutino (2011).

[2] The DDM was first proposed as a model of choice processes in perception tasks, where the subjects are asked to correctly identify visual or auditory stimuli. (For recent reviews see Ratcliff and McKoon (2008) and Shadlen, Hanks, Churchland, Kiani, and Yang (2006).) More recently, DDM-style models have recently been applied to choice tasks, where subjects are choosing from a set of consumption goods presented to them. Clithero and Rangel (2013); Krajbich, Armel, and Rangel (2010); Krajbich and Rangel (2011); Krajbich, Lu, Camerer, and Rangel (2012); Milosavljevic, Malmaud, Huth, Koch, and Rangel (2010a); Reutskaja, Nagel, Camerer, and Rangel (2011)

and that Z_t is a Brownian motion with drift equal to the difference in utilities of the items. This specification corresponds to the optimal decision rule for a Bayesian agent who believes that there are only two states of the world corresponding to whether action l or action r is optimal, pays a constant flow cost per unit of time, and at each point in time decides whether to continue gathering the information or to stop and take an action.[3] The constant stopping regions of the simple DDM imply that the expected amount of time that an agent will gather information depends only on the current value of Z_t, and not on how much time the agent has already spent observing the signal process, and that the probability of the modal choice is independent of the distribution of stopping times.[4] In contrast, in many psychological tasks (Churchland, Kiani, and Shadlen, 2008; Ditterich, 2006) reaction times tend to be higher in incorrect than correct trials. For this reason, when the simple DDM is applied to choice data, it predicts response times that are too long for choices in which the stimulus is weak, or the utility difference between them is small. Ad-hoc extensions of DDM have been developed to better match the data, by allowing more general processes Z_t or stopping regions S_t^i, see e.g., Laming (1968); Link and Heath (1975); Ratcliff (1978). However, past work has left open the question of whether these generalizations correspond to any particular learning problem, and if so, what form those problems take.

Our main focus in this paper is to provide learning-theoretic foundations for an alternative form of DDM, where the agent's behavior is the solution to a sequential sampling problem with a constant cost per unit time as in the simple DDM but with a prior that allows her to learn not only which alternative is better, but also by how much. In this *uncertain-difference DDM*, the state of the world is the vector $\theta = (\theta^l, \theta^r)$ of the utilities of the two choice. In this model an agent with a large sample and Z_t close to zero will decide the utility difference is small, and so be more eager to stop than an agent with the same Z_t but a small sample.

Our main insight is that the nature of the learning problem matters for the optimal stopping strategy and thus for the distribution of choices and response times. In particular, we show that in the uncertain-difference DDM it is optimal to have the range of Z_t for which the agent continues to sample collapse to 0 as time goes to infinity, and moreover that is does so asymptotically at rate $1/t$. The intuition for the fact that the boundary should converge to 0 is not itself new, and has been put forward both as a heuristic in various

---

[3] Wald (1947) stated and solved this as a hypothesis testing problem; Arrow, Blackwell, and Girshick (1949) solved the corresponding Bayesian version. These models were brought to the psychology literature by Stone (1960) and Edwards (1965).

[4] Stone (1960) proved this independence directly for the simple DDM in discrete time. Our Theorem 1 shows that the independence is a consequence of the stopping boundaries being constant.

related models and as a way to better fit the data (see, e.g., Shadlen and Kiani, 2013) but we provide the first precise statement and solution of an optimization problem that generates decreasing boundaries, thus providing a foundation for their use in empirical work, such as the exogenous exponentially-declining boundaries in Milosavljevic, Malmaud, Huth, Koch, and Rangel (2010b).[5] We then use approximation results and numerical methods to determine the functional form of the boundary, thus providing guidance about what the rate of collapse might be expected to be.

Finally, we investigate the consequences of allowing the flow cost to vary arbitrarily with time. Intuitively, if the cost decreases quickly enough, this might outweigh the diminishing effectiveness of learning and lead to an increasing boundary. We show that this intuition is correct, and more strongly that any stopping region at all can be rationalized by a suitable choice of a cost function. Thus optimal stopping on its own imposes essentially no restrictions on the observed choice process, and so it is compatible with the boundaries assumed in Ratcliff and McKoon (2008) and Shadlen and Kiani (2013). However, the force of the model derives from its joint assumptions about the evolution of beliefs and the cost function, and the cost functions needed to rationalize various forms of boundary may or may not seem plausible in the relevant applications.

One motivation for modeling the joint distribution of decision times and choices is that the additional information provided by decision times can lead to models that are closer to the underlying neural mechanisms and may therefore be more robust.[6] In addition, as shown by Clithero and Rangel (2013), including decision times leads to better out-of-sample predictions of choice probabilities. In other settings than the simple choice tasks we study here, decision times can been used to classify decisions as "automatic/instinctive/heuristic" or "cognitive/considered/reflective," as in Rubinstein (2007), Rand, Greene, and Nowak (2012), and Caplin and Martin (2014).

In addition to the papers cited above, our theoretical approach is closely related to the recent work of Woodford (2014). In his model the agent can optimize the dependence of the process Z_t on θ subject to a Shannon capacity constraint, but the stopping rule is constrained to have time-invariant boundaries. In our model the process Z_t is exogenous but the stopping

---

[5]Drugowitsch, Moreno-Bote, Churchland, Shadlen, and Pouget (2012) state a decision problem with two states (e.g. known payoff difference), a fixed terminal time, and time-dependent cost functions, and discuss how to use dynamic programming to numerically compute the solution. Note that even with constant costs the boundary is decreasing when the horizon is finite.

[6]See Shadlen and Kiani (2013) and Bogacz, Brown, Moehlis, Holmes, and Cohen (2006) for discussions of how DDM-type models help explain the correlation between decision times and neurophysiological data such as neuronal firing rates.

rule is optimal subject to a cost, so the two approaches are complementary: both models feature an optimization problem, but over different spaces.

Gabaix and Laibson (2005) and Ke, Shen, and Villas-Boas (2013) look at decisions derived from optimal stopping rules where the gains from sampling are exogenously specified as opposed to being derived from Bayesian updating, as they are here; neither paper examines the correlation between decision time and accuracy. Vul, Goodman, Griffiths, and Tenenbaum (2014) studies the optimal predetermined sample size for an agent whose cost of time arises from the opportunity to make future decisions; they find that for a range of parameters the optimal sample size is one.

Natenzon (2013) and Lu (2013) study models with an exogenous stopping rule. They treat time as a parameter of the choice function, and not as an observable in its own right. The same is true of decision field theory (Busemeyer and Townsend, 1992; Busemeyer and Johnson, 2004), which discusses the effect of time pressure but does not treat the observed choice times as data, and of Caplin and Dean's (2011) model of sequential search. Accumulator models such as (Vickers, 1970) specify an exogenous stopping rule; Webb (2013) shows that the distribution of choices induced by these models is consistent with random utility. Our model makes joint predictions about decisions and response times because the stopping time is chosen optimally. These additional predictions provide more structure on stochastic choice and can help us develop more accurate models.

2   Choice Processes and DDMs

2.1   Observables

Let $A = \{l, r\}$ be the set of alternatives, which we will call left (l) and right (r). Let $T = [0, +\infty)$ be the set of decision times—the times at which the agent is observed to state a choice. The analyst observes a joint probability distribution $P \in \Delta(A \times T)$; we call this a *choice process*. We will decompose P as

$$p^i(t) \text{ and } F(t)$$

where $p^i(t)$ is probability of choosing $i \in A$ conditional on stopping at time t and $F(t) = P(A \times [0, t])$ is the cdf of the marginal distribution of decision times.[7]

---

[7]Formally, we assume that P is a Borel probability measure on $A \times T$. The conditional probabilities $p^i(t)$ exist by Theorems 10.2.1 and 10.2.2 of Dudley (2002).

It will also be useful to decompose P the other way

$$P^i \text{ and } F^i(t)$$

where $P^i = P(\{i\} \times T)$ is the overall probability of choosing $i \in A$ at any time, and $F^i(t) = \frac{P(\{i\} \times [0,t])}{P^i}$ is the cdf of time conditional on choosing $i \in A$.

2.2   Speed and Accuracy

It is easy to define accuracy in perceptual decision tasks, since in such settings the analyst knows which option is 'correct.' However, in choice tasks the agents' preferences are subjective and may be unknown to the researcher.[8] One way of defining accuracy is with respect to the modal choice, as we expect that the objects the agent chooses more often are in some sense "better;" we denote the modal choice by m, the other one by o.

The simplest possible relationship between choices and times is no relationship at all, that is when the distribution of stopping times is independent of the distribution of choices. We will equivalently define this property as follows.

Definition 1. P displays *speed-accuracy independence* iff $p^m(t)$ is a constant function of t

Speed-accuracy independence is a necessary implication of the simple DDM, which we introduce formally in the next section. The case of independence is by nature very knife-edge, and is natural to relax it. In this paper, we focus on qualitatively capturing a positive or a negative correlation between choices and time. To do this, we introduce the following definition.[9]

Definition 2. P displays a *speed-accuracy tradeoff* iff $p^m(t)$ is an increasing function of t.

Note that this definition requires that the tradeoff holds for all times t. We expect there to be a speed-accuracy tradeoff if the agent is learning about the options before her and is stopped by the experimenter at an exogenous stochastic time, as by waiting longer he obtains more information and can make more informed choices. But even if the agent is learning, the observed choice process P need not display a choice-accuracy tradeoff if the stopping

---

[8]In some cases the analyst has a proxy of the preference in form of a separately elicited ranking, see, e.g., Krajbich, Armel, and Rangel (2010), Milosavljevic, Malmaud, Huth, Koch, and Rangel (2010b), Krajbich, Lu, Camerer, and Rangel (2012).

[9]In the literature on perceptual tasks this concept is referred to as "fast errors;" we do not use this terminology here.

time is not exogenous but chosen by the agent as a function of what has been learned so far. In this case, the agent might stop sooner when she thinks his choice is likely to be accurate, so the choice of a stopping time may push towards the opposite side of the speed-accuracy tradeoff.

Definition 3. P displays *speed-accuracy complementarity* iff $p^m(t)$ is a decreasing function of t.

A priori we could observe both kinds of P, perhaps depending on the circumstances. This is indeed the case; for example, Swensson (1972) and Luce (1986) report that speed-accuracy complementarity is observed under normal conditions, but speed-accuracy tradeoff is observed when subjects are incentivized on speed; see also, Shadlen and Kiani (2013).[10]

The speed-accuracy tradeoff can be equivalently expressed in terms of the monotone likelihood ratio property. Let P be a choice process and let f^i be the density of F^i with respect to F. Suppose that F^m is absolutely continuous w.r.t. F^o; we say that F^m and F^o have the *monotone likelihood ratio property*, denoted $F^m \succsim_{\text{MLRP}} F^o$, if the likelihood $f^m(t)/f^o(t)$ is an increasing function.

Though the above concepts will be useful for theoretical analysis, in empirical work time periods will need to be binned to get useful test statistics. For this reason we introduce two weaker concepts that are less sensitive to finite samples, as their oscillation is mitigated by conditioning on larger sets of the form $[0, t]$. First, let $Q^i(t) := \frac{P(\{i\} \times [0,t])}{F(t)}$ be the probability of choosing i conditional on stopping in the interval $[0, t]$. Second, we say that F^m first order stochastically dominates F^o, denoted $F^m \succsim_{\text{FOSD}} F^o$ if $F^m(t) \leq F^o(t)$ for all $t \in T$. Below, we summarize the relationships between these concepts.

Fact 1.

1. *Let P be a choice process and suppose that F^m is absolutely continuous w.r.t. F^o. Then P displays the speed-accuracy tradeoff (complementarity/independence) if and only if $F^m \succsim_{MLRP} F^o$ ($F^m \precsim_{MLRP} F^o$ / $F^m = F^o$).*

2. *If P displays a speed-accuracy tradeoff (complementarity, independence), then $Q^m(t)$ is an increasing (decreasing, constant) function*

3. *If $Q^m(t)$ is an increasing (decreasing, constant) function, then $F^m \succsim_{FOSD} F^o$ ($F^m \precsim_{FOSD} F^o$, $F^m = F^o$).*
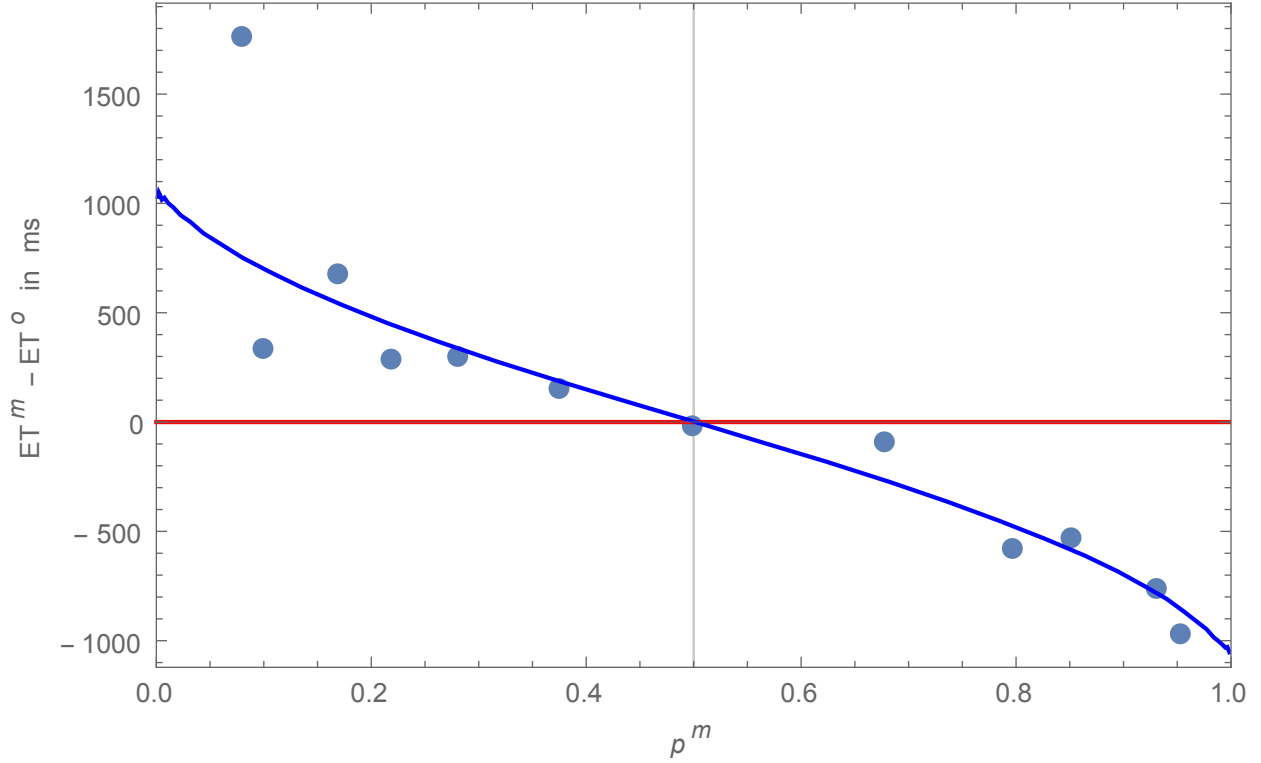
Figure 1: The dots display the data from an experiment by Krajbich, Armel, and Rangel (2010). (Since each alternative pair is sampled only once, we used self-reported rankings collected by Krajbich, Armel, and Rangel (2010) to cluster the budget sets and compute choice probabilities.) The blue line displays simulations of the approximately optimal barrier $\bar{b}$ for the uncertain-difference DDM (derived in Section 3.3) with parameters $c = 0.05, \sigma_0 = \alpha = 1$; each simulation uses $5 \cdot 10^6$ draws and time is discretized with $dt = 0.01$. The expected time to choose l is lower than the expected time to choose r for pairs of alternatives where l is the modal choice; it is higher for pairs where r is modal. The red line displays predictions of the simple DDM for any combination of parameter values.

In particular, under speed-accuracy complementarity, the expected time to choose the modal alternative is smaller than the expected time to choose the other alternative, i.e., $Et^m := \int_0^\infty t dF^m(t) \leq \int_0^\infty t dF^o(t) =: Et^o$. Figure 1 shows that this fits the experimental data of Krajbich, Armel, and Rangel (2010).

The proof of Fact 1 and all other results is in the Appendix. In the next section of the paper we study a particular class of choice processes P, called *drift-diffusion models* (DDM). Such models are defined by a signal process Z_t and a stopping boundary b_t. We characterize the three subclasses of DDM: speed-accuracy tradeoff, complementarity, and independence in terms the function b_t.

2.3   DDM representations

DDM representations have been widely used in the psychology and neuroscience literatures (Ratcliff and McKoon, 2008; Shadlen, Hanks, Churchland, Kiani, and Yang, 2006; Fehr and Rangel, 2011). The two main ingredients of a DDM are the stimulus process Z_t and a boundary b_t.

The stimulus process Z_t is assumed to be a Brownian motion with drift δ and variance α^2:

$$Z_t = \delta t + \alpha B_t, \tag{1}$$

where B_t is a standard Brownian motion. In early applications of DDM such as Ratcliff (1978) Z_t was is not observed by the experimenter. In some recent applications of DDM to neuroscience, the analyst may observe signals that are correlated with Z_t; for example the neural firing rates of both single neurons (Hanes and Schall, 1996) and populations of them (e.g., Ratcliff, Cherian, and Segraves, 2003). In the later sections we interpret the process Z_t as a signal about the utility difference between the two alternatives.

Suppose that the agent stops at a fixed time s and choose l if $Z_s > 0$ and r if $Z_s < 0$ (and flip a coin if there is a tie). Let $p^m(s)$ be the frequency of the modal choice. It is easy to see that if $\delta > 0$ then the modal choice is l, and its probability $p^m(s)$ is an increasing function of the exogenous stopping time s: the process starts at $Z_0 = 0$, so if $s = 0$ each choice is equally likely. At each subsequent point in time $s > 0$ the distribution of Z_s is $\mathcal{N}(\delta s, \alpha^2 s)$; thus, the probability of $Z_s > 0$ increases. The same happens if the agent stops at a stochastic time τ that is independent of the process Z_t. Thus, an *exogenously* stopped process Z_t leads to

---

[10]This could be explained if agents are stopping at random under time pressure, but are using some other rule under normal circumstances; for example they are following the uncertain-difference DDM, to be described in Section 3.3.

a speed-accuracy tradeoff. We will now see that if the process Z_t is stopped *endogenously*, i.e., depending on its value, then this effect can be reversed.

The canonical example of a stopping time that depends on Z_t is the hitting time of a boundary. A *boundary* is a function $b : \mathbb{R}_+ \to \mathbb{R}$. Define the hitting time τ

$$\tau = \inf\{t \geq 0 : |Z_t| \geq b_t\}, \tag{2}$$

i.e., the first time the absolute value of the process Z_t hits the boundary. Let $P(\delta, \alpha, b) \in \Delta(A \times T)$ be the choice process induced by τ and a decision rule that chooses l if $Z_t = b_t$ and r if $Z_t = -b_t$.[11]

Definition 4. A choice process P has a *DDM representation* (δ, α, b) if $P = P(\delta, \alpha, b)$. A choice process P has an *average DDM representation* (μ, α, b) if $P = \int P(\delta, \alpha, b)d\mu(\delta)$, where $\mu \in \Delta(\mathbb{R})$.[12]

We note that the assumption that the process Z_t is Brownian is an important one, as without it the model is vacuous.

Fact 2. *Any choice process P has a "DDM-like" representation where the stochastic process Z_t is arbitrary.*

In particular an arbitrary P may correspond to a process Z_t with jumps (or a filtration that is not right-continuous). However, the standard assumption in the literature is that Z_t is Brownian.[13] The next result characterizes the relationship between speed and accuracy in the class of choice processes that admit a DDM representation.

Theorem 1. *Suppose that P has a DDM representation (α, δ, b). Then P displays a speed-accuracy tradeoff (complementarity, independence) if and only if b_t is increasing (decreasing, constant).*

---

[11]There are boundaries for which there is a positive probability that $\tau = \infty$. This cannot happen for the primitive objects that we consider here, which are choice processes. Thus, we only focus on those boundaries that lead the agent to stop in finite time with probability 1. This property will be satisfied in any model where the stopping time comes from a statistical decision problem in which never stopping incurs an infinite cost and the value of full information is finite.

[12]Note that the parameter α can be removed here by setting $\alpha' = 1, \delta' = \delta/\alpha$, and $b' = b/\alpha$. By a similar argument, δ can be assumed to be $-1, 0$, or 1. We nonetheless retain α and δ here as we will use them in the next section to distinguish between utility and signal strength.

[13]For the Brownian motion Smith (2000) and Peskir et al. (2002) show that the distribution of hitting times satisfies a system of integral equations depending on the boundary. The inverse problem of finding a boundary such that the first hitting time has a given distribution was studied in Iscoe, Kreinin, and Rosen (1999) in the case Brownian Random walks and they propose a Monte Carlo algorithm for solving it numerically.

The intuition behind the proof of this theorem is as follows: Suppose that $\delta > 0$ (so the modal action is l), that the barrier is decreasing, and that the process stopped at time t. The odds that a modal decision is made in this situation are

$$\frac{p^l(t)}{p^r(t)} = \frac{\mathbb{P}[Z_t = b_t | \{\tau = t\} \cap \{|Z_t| = b_t\}]}{\mathbb{P}[Z_t = -b_t | \{\tau = t\} \cap \{|Z_t| = b_t\}]},$$

where $\{\tau = t\}$ is the event that the process Z has not crossed the barrier before time t. From Bayes rule and the formula for the density of the normal distribution

$$\frac{\mathbb{P}[Z_t = b_t | \{|Z_t| = b_t\}]}{\mathbb{P}[Z_t = -b_t | \{|Z_t| = b_t\}]} = \exp\left(\frac{4\delta b_t}{\alpha^2}\right),$$

which is a decreasing function of t whenever b is. Moreover, a symmetry argument using the Brownian bridge shows that the conditioning event $\{\tau = t\}$ does not matter.

Theorem 1 says that the speed-accuracy tradeoff generated by exogenous stopping can be reversed if stopping is endogenous, i.e., the stopping time depends on the process Z_t.

The special case of the constant boundary DDM is well known to arise as the solution to an optimal sampling problem, for an agent who thinks there are only two possible states of the world. The next section presents that model in detail, and then focuses on a related model in which the agent is also learning about the intensity of her preference.

3 Optimal Stopping

3.1 Statement of the model

Both the simple DDM used to explain data from perception tasks and our uncertain-difference DDM are based on the idea of sequential learning and optimal stopping. As we will see, the models differ only in their prior, but this difference leads to substantially different predictions. In the learning model, the agent doesn't know the true utilities, $\theta = (\theta^l, \theta^r) \in \mathbb{R}^2$, but has a prior belief about them $\mu_0 \in \Delta(\mathbb{R}^2)$. The agent observes a signal $(Z_t^i)_{t \in \mathbb{R}_+}$ which as in the DDM has the form of a drift plus a Brownian motion; in the learning model we assume that the drift of each Z^i is equal to the corresponding state, so that

$$dZ_t^i = \theta^i dt + \alpha dB_t^i. \tag{3}$$

where α is the noisiness of the signal and the processes $\{B_t^i\}$ are independent.[14] The signals and prior lie in a probability space $(\Omega, \mathbb{P}, \{\mathcal{F}_t\}_{t\in\mathbb{R}_+})$, where the information $\mathcal{F}_t$ that the agent observed up to time t is simply the paths $\{Z_s^i\}_{0\le s<t}$. We denote the agent's posterior belief about θ given this information by μ_t. Let $X_t^i = \mathbb{E}_{\mu_t}\theta^i = \mathbb{E}\left[\theta^i \mid \mathcal{F}_t\right]$ be the posterior mean for each $i = l, r$. As long as the agent delays the decision she has to pay flow cost, which for now we assume to be constant $c > 0$. (Section 3.4 explores the implications of time varying cost.) The agent's problem is to decide which option to take and at which time. Waiting longer will lead to more informed and thus better decisions, but also entails higher costs. What matters for this decision is the difference between the two utilities, so a sufficient statistic for the agent is

$$Z_t := Z_t^l - Z_t^r = (\theta^l - \theta^r)t + \alpha\sqrt{2}B_t,$$

where $B_t = \frac{1}{\sqrt{2}}(B_t^1 - B_t^2)$ is a Brownian Motion.

When the agent stops it is optimal to choose the option with the highest posterior expected value; thus, the value of stopping at time t is $\max_{i=l,r} X_t^i$. The agent decides optimally when to stop: he chooses a *stopping time* τ, i.e., a function $\tau : \Omega \to [0, +\infty]$ such that $\{\tau \le t\} \in \mathcal{F}_t$ for all t; let $\mathcal{T}$ be the set of all stopping times. Hence, the problem of the agent at $t = 0$ can be stated as

$$\max_{\tau\in\mathcal{T}} \mathbb{E}\left[\max_{i=l,r} X_\tau^i - c\tau\right]^{15}. \tag{4}$$

3.2 Certain Difference

In the simple DDM the agent's prior is concentrated on two points: $\theta_l = (\theta'', \theta')$ and $\theta_r = (\theta', \theta'')$, where $\theta'' > \theta'$. The agent receives payoff θ'' for choosing l in state θ_l or r in state θ_r, and $\theta' < \theta''$ for choosing r in state θ_l or r in state θ_l, so she knows that the magnitude of the utility difference between the two choices is $|\theta'' - \theta'|$, but doesn't know which action is better. We let μ_0 denote the agent's prior probability of θ_l.

This model was first studied in discrete time by Wald (1947) (with a trade-off between type I and type II errors taking the place of utility maximization) and by Arrow, Blackwell, and Girshick (1949) in a standard dynamic programming setting. A version of the result for the continuous-time, Brownian-signal case can be found for example in Shiryaev (1969, 2007).

---

[14]This process was also studied by Natenzon (2013) to study stochastic choice with exogenously forced stopping times; he allows utilities to be correlated, which can explain context effects.

[15]Following the literature, in cases where the optimum is not unique, we focus on the minimal optimal stopping time.

Theorem 2. *With a binomial prior, there is $k > 0$ such that the minimal optimal stopping time is $\hat{\tau} = \inf\{t \geq 0 : |l_t| = k\}$, where $l_t = \log\left(\frac{\mathbb{P}[\theta = \theta_l \,|\, \mathcal{F}_t]}{\mathbb{P}[\theta = \theta_r \,|\, \mathcal{F}_t]}\right)$. Moreover, when $\mu_0 = 0.5$, the optimal stopping time has a DDM representation with a constant boundary b:*

$$\hat{\tau} = \inf\{t \geq 0 \colon |Z_t| \geq b\}.^{16}$$

Theorems 1 and 2 imply that the simple DDM satisfies speed-accuracy independence. From the point of view of most economic applications, the simple DDM misses an important feature, as the assumption that the agent knows the magnitude of the payoff difference rules out cases in which the agents is learning the intensity of his preference. At the technical level, the assumption that the utility difference is known, so there are only two states, implies that the current value of the process Z_t is a sufficient statistic, regardless of the elapsed time; this is why the stopping boundary b in this model is constant. Intuitively, one might expect that if Z_t is close to zero and t is large, the agent would infer that he is close to being indifferent between l and r and so stops, even though for the same value of Z_t the agent would choose to continue when t is small. This inference is ruled out by the binomial prior, which says that the agent is sure that he is not indifferent. We now turn to a model with a Gaussian prior which makes such inferences possible.

3.3 Uncertain-difference DDM

In the uncertain-difference DDM, the agent's prior μ_0 is independent for each action and $\mathcal{N}(X_0, \sigma_0^2)$. Given the specification of the signal process (1), the posterior μ_t is $\mathcal{N}(X_t, \sigma_t^2)$, where

$$X_t^i = \frac{X_0^i \sigma_0^{-2} + Z_t^i \alpha^{-2}}{\sigma_0^{-2} + t\alpha^{-2}} \quad \text{and} \quad \sigma_t^2 = \frac{1}{\sigma_0^{-2} + t\alpha^{-2}}. \tag{5}$$

To gain intuition, consider the agent at time t contemplating a strategy of waiting dt more seconds and stopping then. The agent's utility of stopping now is $\max_{i=l,r} X_t^i$. If the agent waits, he will have a more accurate belief and hence he will be able to make a more informed decision, but he will pay an additional cost, leading to an expected change in utility of $\left(\mathbb{E}_t \max_{i=l,r} X_{t+dt}^i - \max_{i=l,r} X_t^i\right) - c\,dt$. The main intuition for our results is that the value of the additional information gained per unit time is decreasing in t, which leads to stopping regions being time-dependent.

---

[16]This is essentially Theorem 5, p. 185 of Shiryaev (2007), but our formulation is superficially more general, as in his model the drift only depends on the sign of the utility difference, but not on its magnitude.

The following theorem states this intuition formally.[17]

Theorem 3. *Let τ^* be the minimal optimal strategy in* (4). *Then*

1. *There is a strictly decreasing, strictly positive function $k: \mathbb{R}_+ \to \mathbb{R}_+$ such that*

$$\tau^* = \inf\{t \geq 0 \colon |X_t^l - X_t^r| \geq k_t\}.$$

Moreover $\lim_{t\to\infty} k_t = 0$.

2. *If $X_0^l = X_0^r$, then for $b_t = \alpha^2 \sigma_t^{-2} k_t$ we have*

$$\tau^* = \inf\{t \geq 0 \colon |Z_t^l - Z_t^r| \geq b_t\}.$$

Part (1) of the theorem describes the optimal strategy τ^* in terms of stopping regions for posterior means $X_t^l - X_t^r$: It is optimal for the agent to stop once the expected utility difference exceeds a decreasing threshold k_t. This follows from the principle of optimality for continuous time processes and the shift invariance property of the value function, which is due to the normality of the posterior. Intuitively, if the expected utility difference is small for large t, the agent concludes that the two items are most likely indifferent and since σ_t^2 is small for large t, the expected gain from learning more is low, so the agent stops. On the other hand, at $t = 0$ even if the expected utility difference is small, σ_0^2 is large so the expected value of learning is high and thus the agent continues.

Part (2) of the theorem describes the optimal strategy τ^* in terms of stopping regions for the signal process $Z_t := Z_t^l - Z_t^r$.[18] This facilitates comparisons with the simple DDM, where the process of beliefs lives in a different space and is not directly comparable.

One way to understand the difference between this model and the one from the previous section is to consider the agent's posterior beliefs when $Z_t \approx 0$ for large t. In the certain difference model, the agent interprets the signal as noise, since according to his prior the utilities of the the two alternatives are a fixed distance apart, so the agent disregards the signal and essentially starts from anew. This is why the optimal boundaries are constant in this model. On the other hand, in the uncertain difference model the agent's interpretation

---

[17]Time-dependent stopping thresholds also arise if the cost or utility functions are time-dependent or if there is a fixed terminal date, see e.g. Rapoport and Burkheimer (1971) and Drugowitsch, Moreno-Bote, Churchland, Shadlen, and Pouget (2012).

[18]When $X_0^l \neq X_0^r$, the optimal strategy can be described in terms of asymmetric boundaries for the signal process: $\underline{b}_t = \alpha^2 \big[-k_t \sigma_t^{-2} - (X_0^l - X_0^r)\sigma_0^{-2} \big]$ and $\bar{b}_t = \alpha^2 \big[k_t \sigma_t^{-2} - (X_0^l - X_0^r)\sigma_0^{-2} \big]$.

of $Z_t \approx 0$ for large t is that the two alternatives are nearly indifferent, which prompts the agent to stop the costly information gathering process and make a decision right away. This is why the optimal boundaries are decreasing in this model.

A similar observation applies when the agent knows the utility difference but is uncertain about the signal intensity, as in Example (i) of Bather (1962) and Drugowitsch, Moreno-Bote, Churchland, Shadlen, and Pouget (2012). In both of these cases the boundaries collapse to zero because when $Z_t \approx 0$ for large t the agent thinks he is unlikely to learn more in the future.[19]

We do not know of a closed-form solution for the functions k and b; however, we can show that, as functions of the initial variance σ_0^2, c, and noisiness α, they have to satisfy the following conditions. The conditions provide useful information about the identification of the parameters of the model, and about how the predictions of the model change as the parameters are varied in experiments. They are also used to show that k is Lipschitz continuous, which simplifies the analysis of the boundary value problem, and that it declines with time at rate at least $1/\sqrt{t}$, which is at the heart of the proof of Theorem 5 below.

Theorem 4. *The optimal solution $k(t, c, \sigma_0, \alpha)$ to problem* (4) *is Lipschitz continuous in t and satisfies:*

$$k(t, c, \sigma_0, \alpha) = k(0, c, \sigma_t, \alpha) \text{ for all } t \geq 0 \tag{6}$$

$$k(0, c, \lambda\sigma_0, \alpha) = \lambda k(0, c\lambda^{-3}, \sigma_0, \alpha) \text{ for all } \lambda > 0 \tag{7}$$

$$k(t, c, \sigma_0, \lambda\alpha) = \lambda k(t, \lambda^{-1}c, \lambda^{-1}\sigma_0, \alpha) \text{ for all } t, \lambda > 0 \tag{8}$$

$$k(0, \lambda c, \sigma_0, \alpha) \geq \lambda^{-1} k(0, c, \sigma_0, \alpha) \text{ for all } \lambda > 0. \tag{9}$$

The first equality follows from the fact that an agent who starts at time 0 with prior $\mathcal{N}(X_0, \sigma_0^2)$, has the same beliefs at each time $t' > t$ as an agent who started at time t with prior $\mathcal{N}(X_t, \sigma_t^2)$. This equality is used at various steps in the proofs, including showing that k is Lipschitz continuous in t, which is convenient for technical reasons; it allows us to ignore the complications of viscosity solutions and have an exact solution to the PDE that characterizes the optimal stopping rule. The proofs of equalities (7) and (8) use a space-time change. Inequality (9) follows from a space-time change argument and the fact that more information is always better for the agent.

---

[19]In Bather (1962) the time horizon is infinite; his results imply that in this model the boundary b_t decreases at the rate $1/\sqrt{t}$. The finite horizon in Drugowitsch, Moreno-Bote, Churchland, Shadlen, and Pouget (2012) provides an additional reason for the boundaries to decrease.

Theorem 4 implies that if P has an average DDM representation with respect to the agent's prior, then it displays speed-accuracy complementarity.

Theorem 5. *If P has an average DDM representation $P(\mu_0, \alpha, b)$, then P displays speed-accuracy complementarity.*[20]

This is true because the boundary k is decreasing at the rate at least $1/\sqrt{t}$, which, as we show, follows from Theorem 4. This theorem implies that the analyst will observe speed-accuracy complementarity in an experiment in which in the agent faces a series of decisions with states (θ^r, θ^l) and the values of θ are drawn according to the agent's prior. In particular, as long as the prior is correct, speed-accuracy complementarity will hold for the average P in a given experiment. In addition, we expect that the speed-accuracy complementarity should hold at least approximately if the agent's beliefs are approximately correct but we have not shown this formally. Moreover, the complementarity can hold even across experiments as long as the distributions of the states are close enough. That is, while we expect choice-accuracy complementarity to hold within a given class of decision problems, it need not hold across classes: if l and r are two apartments with a given utility difference $\delta = \theta^l - \theta^r$, we expect the agent to spend on average more time here than on a problem where l and r are two lunch items with the same utility difference δ. This is because we expect the prior belief of the agent to be domain specific and in particular, the variance of the prior, σ_0^2, to be higher for houses than for lunch items. Similarly, the complementarity can hold across subjects (indeed, the data in Figure 1 is a cross-section of subjects), as long as their boundaries are not too different. However, in the extreme case when one subject has a cost much lower than another subject, the first one will make choices which are longer and more accurate than the choices of the second subject.

So far we have assumed that the agent receives signals about both items simultaneously. Note that the same conclusions apply if instead the agent has to allocate a fixed amount of attention between the two signals, and this allocation can be altered at will from instant to instant. Because beliefs are normally distributed, the agent will optimally allocate attention in a way that maximizes the reduction of variance in the difference in utilities. The variance in the difference is equal to the sum of the two variances, so the agent will allocate allocation to equate these variances and thus give equal attention to both items at all times.[21]

---

[20]Here we understand μ_0 to be a probability distribution on $\delta = \theta^l - \theta^r$. Note that if b is decreasing, then by Theorem 1, any for each realization of δ the induced choice probabilities $P(\delta, \alpha, b)$ display speed-accuracy complementarity.

[21]In practice people may be only able to attend to one signal at a time, and be physically limited in how

16

To gain more insight into the form of the optimal policy, we characterize the functional form for $\bar{k}$ that satisfies conditions (6–9) with equality, and the $\bar{b}$ that corresponds to it. Using the results of Bather (1962) on the Chernoff (1961) model (which we show to be equivalent to ours in Fact 4) we then show that $\bar{b}$ approximates the solution well for large t.

Fact 3. *Let*

$$\bar{k}(t, c, \sigma_0, \alpha) = \frac{1}{2c\alpha^2(\sigma_0^{-2} + \alpha^{-2}t)^2} \tag{10}$$

and

$$\bar{b}(t, c, \sigma_0, \alpha) = \frac{1}{2c(\sigma_0^{-2} + \alpha^{-2}t)}. \tag{11}$$

Then $\bar{k}$ is the only function that satisfies (6)–(9) with equality, and $\bar{b}$ is the associated boundary in the signal space. Moreover, there are constants $\beta, T > 0$ such that for all $t > T$

$$\left|\bar{b}(t, c, \sigma_0) - b(t, c, \sigma_0)\right| \leq \frac{\beta}{(\sigma_0^{-2} + \alpha^{-2}t)^{5/2}}.$$

Fact 3 implies that b asymptotically declines to zero at rate $1/t$. Using numerical methods, we have found that $\bar{k}$ and $\bar{b}$ are good numerical approximations to the optimal boundaries k and b.[22]

Finally, we note that the uncertain difference model is equivalent to the Chernoff (1961) ex post regret model, where for any stopping time τ the objective function is

$$\text{Ch}(\tau) := \mathbb{E}\left[-\mathbf{1}_{\{x_\tau^l \geq x_\tau^r\}}(\theta^r - \theta^l)^+ - \mathbf{1}_{\{x_\tau^r > x_\tau^l\}}(\theta^l - \theta^r)^+ - c\tau\right];$$

that is, the agent gets zero for making the correct choice and is penalized the foregone utility for making the wrong choice.

Fact 4. *For any stopping time τ*

$$Ch(\tau) = \mathbb{E}\left[\max\{X_\tau^l, X_\tau^r\} - c\tau\right] + \kappa,$$

where κ is a constant independent of τ; therefore, these two objective functions induce the

---

fequently they can switch between them. See Krajbich, Armel, and Rangel (2010); Krajbich, Lu, Camerer, and Rangel (2012) for experimental evidence on eye tracking in these choice tasks.

[22]The approximations for $t \to 0$ obtained by Bather (1962) show that $b \neq \bar{b}$; however our simulations and the bounds from Fact 3 indicate that $\bar{b}$ approximates the solution quite accurately for moderate t, so it may be useful for estimation purposes.

same choice process.

Chernoff and following authors have focused on the behavior of the optimal boundary for very small and very large values of t. We have not found any relevant monotonicity results in this literature.

3.4 Non-Linear Cost

In deriving the DDM representation from optimal stopping, we have so far assumed that the cost of continuing per unit time is constant. We have seen that in the uncertain-difference model, the optimal boundary decreases due to the fact there is less to learn as time goes on. One would expect that the boundary could increase if costs decrease sufficiently quickly. This raises the question of which DDM representations can be derived as a solution to an optimal stopping problem when the cost is allowed to vary arbitrarily over time. The next result shows that for any boundary there exists a cost function such that the boundary is optimal in the learning problem with normal or binomial priors. Thus optimal stopping on its own imposes essentially no restrictions on the observed choice process; the force of the model derives from its joint assumptions about the evolution of beliefs and the cost function.

Theorem 6. *Consider either the Certain or the Uncertain-Difference DDM. For any finite boundary b and any finite set $G \subseteq \mathbb{R}_+$ there exists a cost function $d : \mathbb{R}_+ \to \mathbb{R}$ such that b is optimal in the set of stopping times T that stop in G with probability one*

$$\inf\{t \in G \colon |X_t| \geq b_t\} \in \arg\max_{\tau \in \mathcal{T}} \mathbb{E}\left[\max\{X_\tau^1, X_\tau^2\} - d(\tau)\right] .$$

In particular, there is a cost function such that the exponentially decreasing boundaries in Milosavljevic, Malmaud, Huth, Koch, and Rangel (2010b) are optimal, and a cost function such that there is speed-accuracy independence.

Intuitively, the reason this result obtains is that the optimal stopping rule always takes the form of a cut-off: If the agent stops at time t when $X_t = x$, she stops at time t whenever $|X_t| > x$. This allows us to recursively construct a cost function that rationalizes the given boundary by setting the cost at time t equal the expected future gains from learning. To avoid some technical issues having to do with the solutions to PDE's, we consider a discrete-time finite-horizon formulation of the problem, where the agent is only allowed to stop at times in a finite set G. This lets us construct the associated cost function period by period

instead of using smoothness conditions and stochastic calculus.[23]

4   Conclusion

The recent literature in economics and cognitive science uses drift-diffusion models with time-dependent boundaries. This is helpful in matching observed properties of reaction times, notably their correlation with chosen actions, and in particular a phenomenon that we call speed-accuracy complementarity, where earlier decisions are better than later ones. In Section 2 we showed that the monotonicity properties of the boundary characterize whether the observed choice process displays speed-accuracy complementarity, or the opposite pattern of a speed-accuracy tradeoff. This ties an observable property of behavior (the correlation between reaction times and decisions) to an unobservable construct of the model (the boundary). This connection is helpful for understanding the qualitative properties of DDMs; it may also serve as a useful point of departure for future quantitative exploration of the connection between the rate of decline of the boundary and the strength of correlation between reaction times and actions.

In Section 3 we investigated the DDM as a solution to the optimal sequential sampling problem, where the agent is unsure about the utility of each action and is learning about it as the time passes, optimally deciding when to stop. We studied the dependence of the solution on the nature of the learning problem and also on the cost structure. In particular, we proposed a model in which the agent is learning not only about which option is better, but also by how much it is better. We showed that the boundary in this model asymptotically declines to zero at the rate $1/t$. We also showed that any boundary could be optimal if the agent is facing a nonlinear cost of time.

The analysis of our paper provides a precise foundation for DDMs with time-varying boundaries and establishes a set of useful connections between various parameters of the model and predicted behavior, thus enhancing the theoretical understanding of the model as well as making precise its empirical content. We expect the forces identified in this paper to be present in other decisions involving uncertainty: not just in tasks used in controlled laboratory experiments, but also in decisions involving longer time scales, such as choosing an apartment rental, or deciding which papers to publish (as a journal editor). We hope

---

[23]The proof of the theorem relies on a result on implementable stopping times from Kruse and Strack (2015). In another paper Kruse and Strack (2014) generalize this result to continuous time, but as the absolute value is not covered by their result we can not use it here. Nevertheless, we conjecture that the methods used in that paper can be extended to prove the result in continuous time directly.

these results will be a helpful stepping stone for further work.

Appendix: Proofs

A General Results

A.1 Proof of Fact 1

Proof of Fact 1: To prove part (1) note that by the definition of a conditional distribution (property (c) p. 343 of Dudley, 2002) we have $F^i(t) = \int_{[0,t]} \frac{p^i(s)}{P^i} dF(s)$, so the density of F^i is $f^i(t) = \frac{p^i(t)}{P^i}$. Since F^m is absolutely continuous w.r.t. F^o, the ratio $\frac{f^m(t)}{f^o(t)}$ is well defined F-almost everywhere and equals $\frac{p^m(t)}{p^o(t)} \frac{P^m}{P^o}$. This expression is increasing (decreasing, constant) if and only if $p^m(t)$ is increasing (decreasing, constant).

To prove part (2), note that by the definition of a conditional distribution we have

$$Q^i(t) = \frac{P^i F^i(t)}{F(t)} = \frac{\int_{[0,t]} p^i(s) dF(s)}{F(t)}. \tag{12}$$

Thus, for $t < t'$ we have

$$
\begin{aligned}
Q^l(t) > Q^l(t') \ \ \text{iff} \ \ & \frac{\int_{[0,t]} p^l(s) dF(s)}{F(t)} \geq \frac{\int_{[0,t]} p^l(s) dF(s) + \int_{(t,t']} p^l(s) dF(s)}{F(t) + [F(t') - F(t)]} \\
\text{iff} \ \ & \frac{\int_{[0,t]} p^l(s) dF(s)}{F(t)} \geq \frac{\int_{(t,t']} p^l(s) dF(s)}{F(t') - F(t)},
\end{aligned}
$$

which is true if $p^l(\cdot)$ is a decreasing function since the LHS is the average of p^l on $[0,t]$ and the RHS is the average on $(t, t']$. However, the opposite implication may not hold, for example, consider $p^l(t) := (t - 2/3)^2$ and $F(t) = t$ for $t \in [0,1]$. Then $p^l(t)$ is not decreasing, but $Q^l(t)$ is.

To prove part (3), note that by (12) we have

$$F^l(t) > F^r(t) \ \ \text{iff} \ \ \frac{Q^l(t)}{P^l} \geq \frac{Q^r(t)}{P^r} \ \ \text{iff} \ \ Q^l(t) \geq P^l = \lim_{s \to \infty} Q^l(s),$$

where we used the fact that $Q^l(t) + Q^r(t) = 1$ and $P^l + P^r = 1$. Thus, if Q^l is a decreasing function, the RHS will hold. However, the opposite obviously doesn't have to hold. $\square$

A.2 Proof of Fact 2

Let P be a Borel probability measure on $\Omega := A \times T$. Define the process Z_t as follows. For any $\omega = (a^*, t^*) \in \Omega$ let

$$
Z_t(\omega) := \begin{cases} 0 & \text{if } t < t^* \\ +1 & \text{if } t \geq t^* \text{ and } a^* = l \\ -1 & \text{if } t \geq t^* \text{ and } a^* = r. \end{cases}
$$

Note that Z_t is Borel measurable for each t, so Z_t is a stochastic process on the probability space $(\Omega, \mathcal{B}_\Omega, P)$. Set $b := 1$ and observe that (Z_t, b) is a DDM-like representation of P.

Note that a representation with continuous paths is possible, where

$$Z_t(\omega) := \begin{cases} \frac{t}{t^*} & \text{if } a^* = l \\ -\frac{t}{t^*} & \text{if } a^* = r; \end{cases}$$

however, in this representation the filtration generated by Z_t is not right-continuous at zero. □

A.3 Proof of Theorem 1

Let $f : R_+ \to R_+$ be the density of the distribution of stopping times, and $g : R_+ \times R \to R_+$ be the density of Z_t i.e

$$g(t, y) = \frac{\partial}{\partial y} \mathbb{P}\left[Z_t \le y | \delta, \alpha\right] = \frac{\partial}{\partial y} \mathbb{P}\left[\frac{B_t}{\sqrt{t}} \le \frac{y - \delta t}{\alpha \sqrt{t}}\right] = \phi\left(\frac{y - \delta t}{\alpha \sqrt{t}}\right),$$

where $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^r/2}$ is the density of the standard normal. By Bayes rule:

$$p^l(t) = \mathbb{P}[Z_t = b_t | \tau = t, \delta, \alpha] = \frac{g(t, b_t)\mathbb{P}[\tau = t | Z_t = b_t, \delta, \alpha]}{f(t)}$$

$$p^r(t) = \mathbb{P}[Z_t = -b_t | \tau = t, \delta, \alpha] = \frac{g(t, -b_t)\mathbb{P}[\tau = t | Z_t = -b_t, \delta, \alpha]}{f(t)}$$

It follows from $Z_0 = 0$ and the symmetry of the upper and the lower barrier that

$$\mathbb{P}[\tau = t | Z_t = b_t, \delta, \alpha] = \mathbb{P}[\tau = t | Z_t = -b_t, -\delta, \alpha], \tag{13}$$

because for any path of Z that ends at b_t and crosses any boundary before t, the reflection of this path ends at $-b_t$ and crosses some boundary at the same time.

The induced probability measure over paths conditional on $Z_t = b_t$ is the same as the probability of the Brownian Bridge.[24] The Brownian Bridge is the solution to the SDE $dZ_s = -\frac{b_t - Z}{t - s} ds + \alpha dB_s$ and notably does not depend on the drift δ, which implies that

$$\mathbb{P}[\tau = t | Z_t = -b_t, -\delta, \alpha] = \mathbb{P}[\tau = t | Z_t = -b_t, \delta, \alpha] \tag{14}$$

Thus, by (13) and (14) we have that

$$\frac{p^l(t)}{p^r(t)} = \frac{g(t, b_t)}{g(t, -b_t)} = \exp\left(\frac{4\delta b_t}{\alpha^2}\right).$$

Wlog $m = l$ and $\delta > 0$; the above expression is decreasing over time if and only if b_t is decreasing over time. □

---

[24] See, e.g., Proposition 12.3.2 of Dudley (2002) or Exercise 3.16, p. 41 of Revuz and Yor (1999).

B The Uncertain-Difference Model

B.1 The Value Function

Our results use on the following representation of the posterior process in the uncertain-difference model.

Lemma 1. *For any $t > 0$*

$$X_t^i = X_0^i + \int_0^t \frac{\alpha^{-1}}{\sigma_0^{-2} + s\alpha^{-2}} dW_s^i$$

where W_s^i is a Brownian motion with respect to the filtration information of the agent.

Proof: This follows from Theorem 10.1 and equation 10.52 of Liptser and Shiryaev (2001) by setting $a = b = 0$ and $A = 1, B = \alpha$. □

Define the continuation value as the expected value an agent can achieve by using the optimal continuation strategy if he beliefs the posterior means to be (x^l, x^r) at time t and the variance of his prior equaled σ_0^2 at time 0 and the noisiness of the signal is α.

$$V(t, x^l, x^r, c, \sigma_0, \alpha) := \sup_{\tau \geq t} \mathbb{E}_{(t,x^l,x^r,\sigma_0,\alpha)} \left[\max\{X_\tau^l, X_\tau^r\} - c\,(\tau - t) \right].$$

Lemma 2. *The continuation value has the following properties:*

1. *$\mathbb{E}_{(t,x^l,x^r,\sigma_0,\alpha)} \max\{\theta^l, \theta^r\} \geq V(t, x^l, x^r, c, \sigma_0, \alpha) \geq \max\{x^l, x^r\}$.*

2. *$V(t, x^l, x^r, c, \sigma_0, \alpha) - k = V(t, x^l - k, x^r - k, c, \sigma_0, \alpha)$ for every $k \in \mathbb{R}$.*

3. *The option value $V(t, x^l, x^r, c, \sigma_0, \alpha) - x^i$ is decreasing in x^i for $i \in \{l, r\}$.*

4. *$V(t, x^l, x^r, c, \sigma_0, \alpha)$ is increasing in x^l and x^r.*

5. *The continuation value is Lipschitz continuous in x^l and x^r.*

Proof of Lemma 2

In this proof we equivalently represent a continuation strategy by a pair of stopping times (τ^l, τ^r), one for each alternative.

Proof of 1: For the lower bound, the agent can always stop immediately and get x^l or x^r. For the upper bound, the agent can't do better than receiving a fully informative signal right away and pick the better item immediately.

Proof of 2: Fix a continuation strategy (τ^l, τ^r); the expected payoff equals

$$\mathbb{E}\left[\mathbf{1}_{\{\tau^l \leq \tau^r\}} X_{\tau^l}^l + \mathbf{1}_{\{\tau^l > \tau^r\}} X_{\tau^r}^r - c\left(\min\{\tau^l, \tau^r\} - t\right) \mid X_t^l = x^l - k, X_t^r = x^r - k\right]$$

$$= \mathbb{E}\left[\mathbf{1}_{\{\tau^l \leq \tau^r\}}\left(\int_t^{\tau^l} \frac{\alpha^{-1}}{\sigma_0^{-2} + s\alpha^{-2}} \mathrm{d}W_s^l + x^l - k\right) + \mathbf{1}_{\{\tau^l > \tau^r\}}\left(\int_t^{\tau^r} \frac{\alpha^{-1}}{\sigma_0^{-2} + s\alpha^{-2}} \mathrm{d}W_s^r + x^r - k\right)\right.$$

$$\left. - c\left(\min\{\tau^l, \tau^r\} - t\right) \mid X_t^l = x^l, X_t^r = x^r\right]$$

$$= \mathbb{E}\left[\mathbf{1}_{\{\tau^l \leq \tau^r\}} X_{\tau^l}^l + \mathbf{1}_{\{\tau^l > \tau^r\}} X_{\tau^r}^r - c\left(\min\{\tau^l, \tau^r\} - t\right) \mid X_t^l = x^l, X_t^r = x^r\right] - k.$$

Intuitively, this comes from the translation invariance of the Brownian motion, i.e., the distribution of X_s^i conditional on $X_t^i = x^i - k$ is the same as the distribution of $X_s^i - k$ conditional on $X_t^i = x^i$. As V is defined as the supremum over all continuation strategies (τ^l, τ^r) the result follows.

Proof of 3: The expected difference between stopping at time t with option l and using the continuation strategy (τ^l, τ^r) is

$$\mathbb{E}\left[\mathbf{1}_{\{\tau^l \leq \tau^r\}} X_{\tau^l}^l + \mathbf{1}_{\{\tau^l > \tau^r\}} X_{\tau^r}^r - c\left(\min\{\tau^l, \tau^r\} - t\right) \mid X_t^l = x^l, X_t^r = x^r\right] - x^l$$

$$= \mathbb{E}\left[\mathbf{1}_{\{\tau^l \leq \tau^r\}}(X_{\tau^l}^l - x^l) + \mathbf{1}_{\{\tau^l > \tau^r\}}(X_{\tau^r}^r - x^l) - c\left(\min\{\tau^l, \tau^r\} - t\right) \mid X_t^l = x^l, X_t^r = x^r\right]$$

$$= \mathbb{E}\left[\mathbf{1}_{\{\tau^l \leq \tau^r\}} \int_t^{\tau^l} \frac{\alpha^{-1}}{\sigma_0^{-2} + s\alpha^{-2}} \mathrm{d}W_s^l + \mathbf{1}_{\{\tau^l > \tau^r\}}(X_{\tau^r}^r - x^l) - c\left(\min\{\tau^l, \tau^r\} - t\right) \mid X_t^l = x^l, X_t^r = x^r\right]$$

Note that the first part is independent of x^l, and $(X_{\tau^r}^r - x^l)$ is weakly decreasing in x^l. As for every fixed strategy (τ^l, τ^r) the value of waiting is decreasing the supremum over all continuation strategies is also weakly decreasing in x^l. Thus it follows that the difference between continuation value $V(t, x^l, x^r, c, \sigma_0, \alpha)$ and value of stopping immediately on the first arm x^l is decreasing in x^l for every t and every x^r.

Proof of 4: The expected value of using the continuation strategy (τ^l, τ^r) equals

$$\mathbb{E}\left[\mathbf{1}_{\{\tau^l \leq \tau^r\}} X_{\tau^l}^l + \mathbf{1}_{\{\tau^l > \tau^r\}} X_{\tau^r}^r - c\left(\min\{\tau^l, \tau^r\} - t\right) \mid X_t^l = x^l, X_t^r = x^r\right]$$

$$= \mathbb{E}\left[\mathbf{1}_{\{\tau^l \leq \tau^r\}}\left(\int_t^{\tau^l} \frac{\alpha^{-1}}{\sigma_0^{-2} + s\alpha^{-2}} \mathrm{d}W_s^l + x^l\right) + \mathbf{1}_{\{\tau^l > \tau^r\}} X_{\tau^r}^r - c\left(\min\{\tau^l, \tau^r\} - t\right) \mid X_t^l = x^l, X_t^r = x^r\right]$$

$$= \mathbb{E}\left[\mathbf{1}_{\{\tau^l \leq \tau^r\}}\int_t^{\tau^l} \frac{\alpha^{-1}}{\sigma_0^{-2} + s\alpha^{-2}} \mathrm{d}W_s^l + \mathbf{1}_{\{\tau^l > \tau^r\}} X_{\tau^r}^r - c\left(\min\{\tau^l, \tau^r\} - t\right) \mid X_t^l = x^l, X_t^r = x^r\right]$$

$$+ x^l \mathbb{E}\left[\mathbf{1}_{\{\tau^l \leq \tau^r\}} \mid X_t^l = x^l, X_t^r = x^r\right],$$

which is weakly increasing in x^l. Consequently, the supremum over all continuation strategies (τ^l, τ^r) is weakly increasing in x^l. By the same argument it follows that $V(t, x^l, x^r, c, \sigma_0, \alpha)$ is increasing in x^r.

Proof of 5: The following argument establishes that the value function is Lipschitz continuous with constant one. The initial belief about the mean of either option is additively separable from the change in belief caused by the information the agent observed after time zero. Thus changing the initial beliefs moves the posterior beliefs linearly. Hence, for any fixed stopping time the change

in initial belief can at most linearly move the posterior beliefs about the mean. Furthermore, the expected cost are unaffected by a change in prior beliefs. Thus, the supremum over all stopping times can at most be linearly affected by a change in initial belief. To see this explicitly, observe that

$$
|V(0, x^l, x^r, c, \sigma_0, \alpha) - V(0, y^l, x^r, c, \sigma_0, \alpha)|
$$
$$
= \left| \sup_\tau \mathbb{E}\left[\max\{x^l + \int_0^\tau \frac{\alpha^{-1}}{\sigma_0^{-2} + s\alpha^{-2}} \mathrm{d}W_s^1, x^r + \int_0^\tau \frac{\alpha^{-1}}{\sigma_0^{-2} + s\alpha^{-2}} \mathrm{d}W_s^2\} - c(\tau) \right] \right.
$$
$$
\left. - \sup_\tau \mathbb{E}\left[\max\{y^l + \int_0^\tau \frac{\alpha^{-1}}{\sigma_0^{-2} + s\alpha^{-2}} \mathrm{d}W_s^1, x^r + \int_0^\tau \frac{\alpha^{-1}}{\sigma_0^{-2} + s\alpha^{-2}} \mathrm{d}W_s^2\} - c(\tau) \right] \right|
$$
$$
\leq \left| \sup_\tau \mathbb{E}\left[\max\{x^l + \int_0^\tau \frac{\alpha^{-1}}{\sigma_0^{-2} + s\alpha^{-2}} \mathrm{d}W_s^1, x^r + \int_0^\tau \frac{\alpha^{-1}}{\sigma_0^{-2} + s\alpha^{-2}} \mathrm{d}W_s^2\} \right. \right.
$$
$$
\left. \left. - \max\{y^l + \int_0^\tau \frac{\alpha^{-1}}{\sigma_0^{-2} + s\alpha^{-2}} \mathrm{d}W_s^1, x^r + \int_0^\tau \frac{\alpha^{-1}}{\sigma_0^{-2} + s\alpha^{-2}} \mathrm{d}W_s^2\} \right] \right| \leq |y^l - x^l|. \quad \square
$$

B.2   Proof of Theorem 3

B.2.1   Characterization

Note that due to the symmetry of the problem $V(t, x^l, x^r, c, \sigma_0, \alpha) = V(t, x^r, x^l, c, \sigma_0, \alpha)$. Without loss of generality suppose $x^l \leq x^r$ otherwise swap x^l and x^r. As X_t is a Markov process, the principle of optimality[25] implies that the agent's problem admits a solution of the form $\tau = \inf\{t \geq 0 : \max_{i=l,r} X_t^i \geq V(t, X_t^l, X_t^r, c, \sigma_0, \alpha)\}$. Thus, it is optimal to stop if and only if

$$
0 = V(t, x^l, x^r, c, \sigma_0, \alpha) - \max\{x^l, x^r\} = V(t, x^l, x^r, c, \sigma_0, \alpha) - x^r = V(t, x^l - x^r, 0, c, \sigma_0, \alpha).
$$

An envelope argument yields that V is continuous. Using the continuity and the monotonicity of V we can define the function k implicitly by

$$
k(t) := \inf\{x \in \mathbb{R} : 0 = V(t, -x, 0, c, \sigma_0, \alpha)\}.
$$

As $x^l - x^r \leq 0$, V is monotone increasing in the second argument and $V(t, x^l - x^r, 0, c, \sigma_0, \alpha) \geq 0$ by Lemma 2 we have

$$
\{0 = V(t, x^l - x^r, 0, c, \sigma_0)\} = \{x^l - x^r \leq -k(t)\} = \{|x^l - x^r| \geq k(t)\}.
$$

Hence the optimal strategy equals $\tau = \inf\{t \geq 0 : |X_t^l - X_t^r| \geq k(t)\}$.

---

[25]Our model does not satisfy condition (2.1.1) of Peskir and Shiryaev (2006) because for some stopping times the expected payoff is minus infinity, but as they indicate on p. 2 the proof can be extended to our case.

B.2.2 Monotonicity

First, we show that $V(t, x^l, x^r, c, \sigma_0, \alpha)$ is decreasing in t. Note that by Doob's optional sampling theorem for every fixed stopping strategy τ

$$
\begin{aligned}
\mathbb{E}\left[\max\{X_\tau^l, X_\tau^r\} - c\tau \mid X_t = (x^l, x^r)\right] &= \mathbb{E}\left[\max\{X_\tau^l - X_\tau^r, 0\} + X_\tau^r - c\tau \mid X_t = (x^l, x^r)\right] \\
&= \mathbb{E}\left[\max\{X_\tau^l - X_\tau^r, 0\} - c\tau \mid X_t = (x^l, x^r)\right] + x_t^r.
\end{aligned}
$$

Define the process $X_t := X_t^l - X_t^r$, and note that

$$
X_t = X_t^l - X_t^r = X_0^l - X_0^r + \int_0^t \frac{\alpha^{-1}}{\sigma_0^{-2} + s\alpha^{-2}} \left(dW_s^l - dW_s^r\right) = X_0^l - X_0^r + \int_0^t \frac{\sqrt{2}\alpha^{-1}}{\sigma_0^{-2} + s\alpha^{-2}} d\tilde{W}_s, \tag{15}
$$

where $\tilde{W}$ is a Brownian motion. Define a time change as follows: Let $q(k)$ solve $k = \int_0^{q(k)} \left(\frac{\sqrt{2}\alpha^{-1}}{\sigma_0^{-2} + s\alpha^{-2}}\right)^2 ds$. This implies that $q(k) = \frac{k\alpha^2\sigma_0^{-2}}{2\sigma_0^2 - k}$. Define $\psi(t) = \frac{2\sigma_0^2 t}{\alpha^2\sigma_0^{-2} + t}$. By the Dambis, Dubins–Schwarz theorem (see, e.g., Theorem 1.6, chapter V of Revuz and Yor, 1999) $W_s := (X_{q(s)})_{s \in [0, 2\sigma_0^{-2}]}$ is a Brownian motion and thus we can rewrite the problem as

$$
\begin{aligned}
V(t, x^l, x^r, c, \sigma_0, \alpha) &= \sup_{\tau \geq \psi(t)} \mathbb{E}\left[\max\{W_\tau, 0\} - c\left(q(\tau) - q(\psi(t))\right) \mid W_{\psi(t)} = x^l - x^r\right] + x^r \\
&= \sup_{\tau \geq \psi(t)} \mathbb{E}\left[\max\{W_\tau, 0\} - c\left(\int_{\psi(t)}^\tau q'(s)ds\right) \mid W_{\psi(t)} = x^l - x^r\right] + x^r, \\
&= \sup_{\tau \geq \psi(t)} \mathbb{E}\left[\max\{W_\tau, 0\} - c\left(\int_{\psi(t)}^\tau \frac{2\alpha^2}{(2\sigma_0^2 - s)^2}ds\right) \mid W_{\psi(t)} = x^l - x^r\right] + x^r.
\end{aligned}
$$

Next, we remove the conditional expectation in the Brownian motion by adding the initial value

$$
V(t, x^l, x^r, c, \sigma_0, \alpha) = \sup_{\tau \geq \psi(t)} \mathbb{E}\left[\max\{W_\tau + \left(x^l - x^r\right), 0\} - c\int_{\psi(t)}^\tau \frac{2\alpha^2}{(2\sigma_0^2 - s)^2}ds\right] + x^r.
$$

Define $\hat{\tau} = \tau - \psi$ and let wlog $x^l < x^r$, then

$$
V(t, x^l, x^r, c, \sigma_0, \alpha) = \sup_{\hat{\tau} \geq 0} \mathbb{E}\left[\max\{W_\tau - |x^l - x^r|, 0\} - c\int_{\psi(t)}^{\psi(t) + \hat{\tau}} \frac{2\alpha^2}{(2\sigma_0^2 - s)^2}ds\right] + \max\{x^l, x^r\};
$$

because the current state is a sufficient statistic for Brownian motion we have

$$
V(t, x^l, x^r, c, \sigma_0, \alpha) = \sup_{\hat{\tau} \geq 0} \mathbb{E}\left[\max\{W_\tau - |x^l - x^r|, 0\} - c\int_0^{\hat{\tau}} \frac{2\alpha^2}{(2\sigma_0^2 - s - \psi(t))^2}ds\right] + \max\{x^l, x^r\}.
$$

Note that for every fixed strategy τ the cost term is increasing in t and $\psi(t)$ and thus $V(t, x^l, x^r, c, \sigma_0, \alpha) - \max\{x^l, x^r\}$ is non-increasing.

Finally, suppose that $t < t'$; then

$$0 = V(t, -k(t, c, \sigma_0, \alpha), 0, c, \sigma_0, \alpha) \geq V(t', -k(t, c, \sigma_0, \alpha), 0, c, \sigma_0, \alpha).$$

By Lemma 2, $V(t', -k(t, c, \sigma_0, \alpha), 0, c, \sigma_0, \alpha) \geq 0$ and hence $0 = V(t', -k(t, c, \sigma_0, \alpha), 0, c, \sigma_0, \alpha)$. Hence

$$k(t, c, \sigma_0, \alpha) \geq \inf\{x \in \mathbb{R} \colon 0 = V(t', -x, 0, c, \sigma_0, \alpha)\} = k(t', c, \sigma_0, \alpha).$$

B.2.3 Positivity

The payoff of the optimal decision rule is at least as high as the payoff from using the strategy that stops at time Δ for sure. Because the information gained over a short time period Δ is of order $\epsilon^{\frac{1}{2}}$ and the cost is linear, we expect that it is always worth buying some information when the expected utility of both options is the same. To see this formally, note that

$$
\begin{aligned}
V(t, x, x, c, \sigma_0, \alpha) - x &= \sup_\tau \mathbb{E}\left[\max\{W_\tau, 0\} - \int_0^\tau \frac{2c\alpha^2}{(2\sigma_0^2 - s - \psi(t))^2}\mathrm{d}s\right] \\
&\geq \mathbb{E}\left[\max\{W_\epsilon, 0\} - \int_0^\epsilon \frac{2c\alpha^2}{(2\sigma_0^2 - s - \psi(t))^2}\mathrm{d}s\right] \\
&= \int_0^\infty z\frac{1}{\sqrt{2\pi\epsilon}}e^{-\frac{z^2}{2\epsilon}}\mathrm{d}z - \int_0^\epsilon \frac{2c\alpha^2}{(2\sigma_0^2 - s - \psi(t))^2}\mathrm{d}s \\
&\geq \sqrt{\frac{\epsilon}{2\pi}} - \frac{2c\alpha^2\epsilon}{(2\sigma_0^2 - \psi(t) - \epsilon)^2} \\
&\geq \sqrt{\frac{\epsilon}{2\pi}} - \frac{2c\alpha^2\epsilon}{(2\sigma_0^2 - \psi(t) - \tilde{\epsilon})^2} \quad \text{for all fixed } \tilde{\epsilon} \in [\epsilon, 2\sigma_0^2 - \psi(t)).
\end{aligned}
$$

As the first term goes to zero with the speed of square root while the second term shrinks linearly we get that $V(t, x, x, c, \sigma_0, \alpha) - \max\{x, x\} > 0$ for some small $\epsilon > 0$ and thus the agent does not stop when his posterior mean is the same on both options.

B.2.4 Zero limit

Let $k(s, c, \sigma_0, \alpha) \geq K^* > 0$ for all $s \geq t$. Consider the time t history where $X_t^l = X_t^r$. The probability that the agent never stops (and thus pays infinity costs) is bounded from below by the probability that the process $X^l - X^r$ stays in the interval $[-K^*, K^*]$,

$$\mathbb{P}\left[\sup_{s \in [t, \infty)}|X_s^l - X_s^r| \leq K(s)\big| X_t^l = X_t^r\right] \geq \mathbb{P}\left[\sup_{s \in [t, \infty)}|X_s^l - X_s^r| \leq K^*\big| X_t^l = X_t^r\right]$$

By the above time change argument this equals the probability that a Brownian motion leaves the interval $[-K, K]$ in the time from $\psi(t)$ to $2\sigma_0^2$,

$$\mathbb{P}\left[\sup_{s \in [t, \infty)}|X_s^l - X_s^r| \leq K^*\big| X_t^l = X_t^r\right] = \mathbb{P}\left[\sup_{s \in [\psi(t), 2\sigma_0^2]}|W_s| \leq K^*(s)\right]$$

27

This probability is non-zero. Thus, there is a positive probability the agent incurs infinite cost. Because the expected gain is bounded by the full information payoff, this is a contradiction. □

B.3 Proof of Theorem 4

Lemma 3.

$$V(0, x^l, x^r, c\lambda, \sigma_0, \alpha) =$$

$$= \lambda^{-1} \sup_{\tau'} \mathbb{E}\left[\max\left\{\lambda x^l + \int_0^{\tau'} \frac{\alpha^{-1}}{\sigma_0^{-2} + s\lambda^{-2}\alpha^{-2}} \mathrm{d}M_s^l, \lambda x^r + \int_0^{\tau'} \frac{\alpha^{-1}}{\sigma_0^{-2} + s\lambda^{-2}\alpha^{-2}} \mathrm{d}M_s^r\right\} - c\tau\right]$$

Proof: We have that $V(0, x^l, x^r, c\lambda, \sigma_0, \alpha)$ equals

$$\sup_\tau \mathbb{E}\left[\max\left\{x^l + \int_0^\tau \frac{\alpha^{-1}}{\sigma_0^{-2} + s\alpha^{-2}} \mathrm{d}W_s^l, x^r + \int_0^\tau \frac{\alpha^{-1}}{\sigma_0^{-2} + s\alpha^{-2}} \mathrm{d}W_s^r\right\} - c\lambda\tau\right]$$

$$= \lambda^{-1} \sup_\tau \mathbb{E}\left[\max\left\{\lambda x^l + \int_0^\tau \frac{\lambda\alpha^{-1}}{\sigma_0^{-2} + s\alpha^{-2}} \mathrm{d}W_{(s\lambda^2)\lambda^{-2}}^l, \lambda x^r + \int_0^\tau \frac{\lambda\alpha^{-1}}{\sigma_0^{-2} + s\alpha^{-2}} \mathrm{d}W_{(s\lambda^2)\lambda^{-2}}^r\right\} - c\lambda^2\tau\right]$$

$$= \lambda^{-1} \sup_\tau \mathbb{E}\left[\max\left\{\lambda x^l + \int_0^{\tau\lambda^2} \frac{\lambda\alpha^{-1}}{\sigma_0^{-2} + s\lambda^{-2}\alpha^{-2}} \mathrm{d}W_{s\lambda^{-2}}^l, \lambda x^r + \int_0^{\tau\lambda^2} \frac{\lambda\alpha^{-1}}{\sigma_0^{-2} + s\lambda^{-2}\alpha^{-2}} \mathrm{d}W_{s\lambda^{-2}}^r\right\} - c\lambda^2\tau\right]$$

$$= \lambda^{-1} \sup_\tau \mathbb{E}\left[\max\left\{\lambda x^l + \int_0^{\tau\lambda^2} \frac{\alpha^{-1}}{\sigma_0^{-2} + s\lambda^{-2}\alpha^{-2}} \mathrm{d}M_s^l, \lambda x^r + \int_0^{\tau\lambda^2} \frac{\alpha^{-1}}{\sigma_0^{-2} + s\lambda^{-2}\alpha^{-2}} \mathrm{d}M_s^r\right\} - c\lambda^2\tau\right]$$

For the step from the second to third line apply Proposition 1.4 of Chapter V of Revuz and Yor (1999) with $C_s := s\lambda^{-2}$ and $H_s := \frac{\alpha^{-1}\lambda}{\sigma_0^{-2} + \alpha^{-2}\lambda^{-2}s}$ (pathwise to the integrals with limits τ and $\tau\lambda^2$). In the next step we apply a time-change, where $M_r^i := \lambda W_{r\lambda^{-2}}^i$ is a Brownian motion and τ' is a stopping time measurable in the natural filtration generated by M. □

B.3.1 Proof of (6)

Note that Lemma 1 implies that for any $t < t'$

$$X_{t'}^i = X_t^i + \int_0^{t'-t} \frac{\alpha^{-1}}{(\sigma_0^{-2} + \alpha^{-2}t) + \alpha^{-2}s} dW_{t+s}^i$$

where W_s^i is a Brownian motion with respect to the filtration information of the agent. Thus, if the agent starts with a prior at time 0 equal to $\mathcal{N}(X_0, \sigma_0^2)$, then his belief at time t' is exactly the same as if he started with a prior at t equal to $\mathcal{N}(X_t, \sigma_t^2)$ where $\sigma_t^{-2} = \sigma_0^{-2} + \alpha^{-2}t$. Thus, $V(t, x^l, x^r, c, \sigma_0, \alpha) = V(0, x^l, x^r, c, \sigma_t, \alpha)$ so

$$k(t, c, \sigma_0, \alpha) = \inf\{x > 0 : 0 = V(t, 0, -x, c, \sigma_0, \alpha)\}$$
$$= \inf\{x > 0 : 0 = V(0, 0, -x, c, \sigma_t, \alpha)\} = k(0, c, \sigma_t, \alpha). \quad \square$$

B.3.2   Proof of (7)

By Lemma 3, $V(0, x^l, x^r, c\lambda, \sigma_0, \alpha)$ equals

$$= \lambda^{-1} \sup_{\tau'} \mathbb{E}\left[\max\left\{\lambda x^l + \int_0^{\tau'} \frac{\alpha^{-1}}{\sigma_0^{-2} + s\lambda^{-2}\alpha^{-2}} \mathrm{d}M_s^l, \lambda x^r + \int_0^{\tau'} \frac{\alpha^{-1}}{\sigma_0^{-2} + s\lambda^{-2}\alpha^{-2}} \mathrm{d}M_s^r \right\} - c\tau\right]$$

$$= \lambda \sup_{\tau'} \mathbb{E}\left[\max\left\{\lambda^{-1} x^l + \int_0^{\tau'} \frac{\alpha^{-1}}{\sigma_0^{-2}\lambda^2 + s\alpha^{-2}} \mathrm{d}M_s^l, \lambda^{-1} x^r + \int_0^{\tau'} \frac{\alpha^{-1}}{\sigma_0^{-2}\lambda^2 + s\alpha^{-2}} \mathrm{d}M_s^r \right\} - c\lambda^{-3}\tau\right]$$

$$= \lambda V(0, x_1 \lambda^{-1}, x_2 \lambda^{-1}, c\lambda^{-3}, \sigma_0/\lambda, \alpha).$$

Thus,

$$
\begin{aligned}
k(0, c, \sigma_0, \alpha) &= \inf\{x > 0 \colon 0 = V(0, 0, -x, c, \sigma_0, \alpha)\} \\
&= \inf\{x > 0 \colon 0 = V(0, 0, -x\,\lambda^{-1}, c\,\lambda^{-3}, \sigma_0/\lambda, \alpha)\} \\
&= \lambda \inf\{y > 0 \colon 0 = V(0, 0, -y, c\,\lambda^{-3}, \sigma_0/\lambda, \alpha)\} = \lambda\, k(0, c\,\lambda^{-3}, \sigma_0/\lambda, \alpha).
\end{aligned}
$$

Setting $\tilde\sigma_0 = \sigma_0/\lambda$ gives the result. $\qquad\square$

B.3.3   Proof of (8)

First, observe that $V(t, x_1, x_2, c, \sigma_0, \lambda\alpha)$ equals

$$= \sup_{\tau \geq t} \mathbb{E}\left[\max\left\{x_1 + \int_t^{\tau} \frac{\lambda^{-1}\alpha^{-1}}{\sigma_0^{-2} + \lambda^{-2}\alpha^{-2}s} \mathrm{d}W_s^1, x_2 + \int_t^{\tau} \frac{\lambda^{-1}\alpha^{-1}}{\sigma_0^{-2} + \lambda^{-2}\alpha^{-2}s} \mathrm{d}W_s^2 \right\} - c(\tau - t)\right]$$

$$= \lambda \sup_{\tau \geq t} \mathbb{E}\left[\max\left\{\lambda^{-1} x_1 + \int_t^{\tau} \frac{\alpha^{-1}}{\lambda^2\sigma_0^{-2} + \alpha^{-2}s} \mathrm{d}W_s^1, \lambda^{-1} x_2 + \int_t^{\tau} \frac{\alpha^{-1}}{\lambda^2\sigma_0^{-2} + \alpha^{-2}s} \mathrm{d}W_s^2 \right\} - (c\lambda^{-1})(\tau - t)\right]$$

$$= \lambda V(t, \lambda^{-1} x_1, \lambda^{-1} x_2, \lambda^{-1} c, \lambda^{-1}\sigma_0, \alpha).$$

Thus,

$$
\begin{aligned}
k(t, c, \sigma_0, \lambda\alpha) &= \inf\{x > 0 \colon 0 = V(t, 0, -x, c, \sigma_0, \lambda\alpha)\} \\
&= \inf\{x > 0 \colon 0 = V(t, 0, -\lambda^{-1}x, \lambda^{-1}c, \lambda^{-1}\sigma_0, \alpha)\} \\
&= \lambda \inf\{y > 0 \colon 0 = V(t, 0, -y, \lambda^{-1}c, \lambda^{-1}\sigma_0, \alpha)\} = \lambda\, k(t, \lambda^{-1}c, \lambda^{-1}\sigma_0, \alpha). \quad \square
\end{aligned}
$$

B.3.4   Proof of (9)

By Lemma 3, $V(0, x^l, x^r, c\lambda, \sigma_0, \alpha)$ equals

$$\lambda^{-1} \sup_{\tau'} \mathbb{E}\left[\max\left\{\lambda x^l + \int_0^{\tau'} \frac{\alpha^{-1}}{\sigma_0^{-2} + s\lambda^{-2}\alpha^{-2}} \mathrm{d}M_s^l, \lambda x^r + \int_0^{\tau'} \frac{\alpha^{-1}}{\sigma_0^{-2} + s\lambda^{-2}\alpha^{-2}} \mathrm{d}M_s^r \right\} - c\tau\right]$$

As receiving more information is always better we have that for all $\lambda > 1$

$$V(0, x^l, x^r, c\lambda, \sigma_0, \alpha) \geq \lambda^{-1} \sup_\tau \mathbb{E}\left[\max\left\{\lambda x^l + \int_0^\tau \frac{\alpha^{-1}}{\sigma_0^{-2} + s\alpha^{-2}} dM_s^l, \lambda x^r + \int_0^\tau \frac{\alpha^{-1}}{\sigma_0^{-2} + s\alpha^{-2}} dM_s^r\right\} - c\tau\right]$$

$$= \lambda^{-1} V(0, \lambda x^l, \lambda x^r, c, \sigma_0, \alpha)$$

This implies that $k(t, \lambda c, \sigma_0, \alpha) \geq \lambda^{-1} k(t, c, \sigma_0, \alpha)$ for all $\lambda > 1$

$$k(t, \lambda c, \sigma_0, \alpha) = \inf\{x > 0: 0 = V(t, 0, -x, \lambda c, \sigma_0, \alpha)\}$$
$$\geq \inf\{x > 0: 0 = V(t, 0, -x\lambda, c, \sigma_0, \alpha)\}$$
$$= \lambda^{-1} \inf\{y > 0: 0 = V(0, 0, -y, c, \sigma_0, \alpha)\} = \lambda^{-1} k(t, c, \sigma_0, \alpha). \quad \square$$

B.3.5 Lipschitz continuity of k

Let $\lambda_\epsilon = \left(1 + \epsilon\alpha^{-2}\sigma_0^2\right)^{-1/2} < 1$ and note that by definition $\lambda_\epsilon \sigma_0 = \sigma_\epsilon$. We can thus use equations (6), (7) and (9) to get

$$k(\epsilon, c, \sigma_0, \alpha) = k(0, c, \sigma_\epsilon, \alpha) = \lambda_\epsilon k(0, c\lambda_\epsilon^{-3}, \sigma_0) \geq \lambda_\epsilon^4 k(0, c, \sigma_0, \alpha).$$

As a consequence we can bound the difference between the value of the barrier at time zero and at time ϵ from below

$$k(\epsilon, c, \sigma_0, \alpha) - k(0, c, \sigma_0, \alpha) \geq \left[\left(1 + \epsilon\alpha^{-2}\sigma_0^2\right)^{-2} - 1\right] k(0, c, \sigma_0, \alpha).$$

Dividing by ϵ and taking the limit $\epsilon \to 0$ yields that k_t the partial derivative of the boundary with respect to time satisfies

$$k_t(0, c, \sigma_0, \alpha) \geq -2\alpha^{-2}\sigma_0^2 k(0, c, \sigma_0, \alpha).$$

Since by equation (6) $k(t + \epsilon, c, \sigma_0, \alpha) = k(\epsilon, c, \sigma_t, \alpha)$ we have that

$$k_t(t, c, \sigma_0, \alpha) = \lim_{\epsilon \to 0} \frac{k(t + \epsilon, c, \sigma_0, \alpha) - k(t, c, \sigma_0, \alpha)}{\epsilon} = \lim_{\epsilon \to 0} \frac{k(\epsilon, c, \sigma_t, \alpha) - k(0, c, \sigma_t, \alpha)}{\epsilon}$$
$$= k_t(0, c, \sigma_t, \alpha) \geq -2\alpha^{-2}\sigma_t^2 k(0, c, \sigma_t, \alpha) = -2\alpha^{-2}\sigma_t^2 k(t, c, \sigma_0, \alpha) \geq -2\alpha^{-2}\sigma_0^2 k(0, c, \sigma_0, \alpha),$$

where the last equality follows from equation (6) and the last inequality follows since k and σ_t are decreasing in t. Thus, k_t is bounded from below; since its upper bound is zero, k is Lipschitz continuous in t. $\quad \square$

B.4 Proof of Theorem 5

If the agent stops at time t at the barrier $k(t, c, \sigma_0, \alpha)$, his posterior belief is that the true states are normally distributed with $\theta^i \sim \mathcal{N}(X_t^i, \sigma_t)$. Thus, the probability that the agent assigns to picking l

when r is optimal, conditional on stopping at t, is

$$
\begin{aligned}
\mathbb{P}\left[\theta^l < \theta^r \mid X_t^l - X_t^r = k(t, c, \sigma_0, \alpha)\right] &= \mathbb{P}\left[(\theta^l - \theta^r) - (X_t^l - X_t^r) \leq -k(t, c, \sigma_0, \alpha) \,\Big|\, X_t^l - X_t^r = k(t, c, \sigma_0, \alpha)\right] \\
&= \mathbb{P}\left[\frac{(\theta^l - \theta^r) - (X_t^l - X_t^r)}{\sqrt{2}\sigma_t} \leq -\frac{k(t, c, \sigma_0, \alpha)}{\sqrt{2}\sigma_t} \,\Big|\, X_t^l - X_t^r = k(t, c, \sigma_0, \alpha)\right] \\
&= \Phi\left(-\frac{1}{\sqrt{2}}k(t, c, \sigma_0, \alpha)\sigma_t^{-1}\right).
\end{aligned}
$$

From the symmetry of the problem, there is the same probability of mistakenly picking r instead of l. To show that the probability of being wrong increases over time, it remains to show that $k(t, c, \sigma_0, \alpha)\sigma_t^{-1}$ is decreasing in t. We have that

$$
\frac{\partial}{\partial \sigma_t}\left[k(0, \sigma_t, c, \alpha)\sigma_t^{-1}\right] = k_\sigma(0, c, \sigma_t, \alpha)\sigma_t^{-1} - k(0, c, \sigma_t, \alpha)\sigma_t^{-2}.
$$

We will now show that this is equal to $-3\,c\,k_c(0, \sigma_t, c, \alpha)\sigma_t^2$, which is nonnegative. To see that, we show that $k_\sigma(0, c, \sigma_0, \alpha)\sigma_0 = -3\,c\,k_c(0, c, \sigma_0, \alpha) + k(0, c, \sigma_0, \alpha)$. Set

$$
\beta_\epsilon \sigma_0 = \sigma_0 + \epsilon \Rightarrow \beta_\epsilon = 1 + \frac{\epsilon}{\sigma_0}.
$$

Inserting in equation. (7) gives

$$
\begin{aligned}
k(0, c, \sigma_0 \beta_\epsilon, \alpha) &= k(0, c, \sigma_0 + \epsilon, \alpha) = \beta_\epsilon\, k(0, c\beta_\epsilon^{-3}, \sigma_0, \alpha) \\
\Leftrightarrow k(0, c, \sigma_0 + \epsilon, \alpha) - k(0, c, \sigma_0, \alpha) &= k(0, c\beta_\epsilon^{-3}, \sigma_0, \alpha) - k(0, c, \sigma_0, \alpha) + \frac{\epsilon}{\sigma_0}k(0, c\beta_\epsilon^{-3}, \sigma_0, \alpha).
\end{aligned}
$$

Dividing by ϵ and taking the limit $\epsilon \to 0$ yields

$$
\begin{aligned}
k_\sigma(0, c, \sigma_0, \alpha) &= k_c(0, c, \sigma_0, \alpha)c\left[\lim_{\epsilon \to 0}\frac{\beta_\epsilon^{-3} - 1}{\epsilon}\right] + \frac{1}{\sigma_0}k(0, c, \sigma_0, \alpha) \\
&= k_c(0, c, \sigma_0, \alpha)c\left[-3\right]\frac{\partial \beta_\epsilon}{\partial \epsilon} + \frac{1}{\sigma_0}k(0, c, \sigma_0, \alpha) \\
&= -k_c(0, c, \sigma_0, \alpha)c\frac{3}{\sigma_0} + \frac{1}{\sigma_0}k(0, c, \sigma_0, \alpha) \\
\Leftrightarrow k_\sigma(0, c, \sigma_0, \alpha)\sigma_0 &= -3\,c\,k_c(0, c, \sigma_0, \alpha) + k(0, c, \sigma_0, \alpha).
\end{aligned}
$$

B.5 Proof of Fact 4

Let $\kappa := \mathbb{E}\left[\max\{\theta^l, \theta^r\}\right]$ and fix a stopping time τ. To show that

$$
\mathbb{E}\left[-\mathbf{1}_{\{X_\tau^l \geq X_\tau^r\}}(\theta^r - \theta^l)^+ - \mathbf{1}_{\{X_\tau^r > X_\tau^l\}}(\theta^l - \theta^r)^+ - c\tau\right] = \mathbb{E}\left[\max\{X_\tau^l, X_\tau^r\} - c\tau\right] + \kappa,
$$

the cost terms can be dropped. Let D be the difference between the expected payoff from the optimal decision and the expected payoff from choosing the correct action, $D := \mathbb{E}\left[\max\{X_\tau^l, X_\tau^r\}\right] -$

$\mathbb{E}\left[\max\{\theta^l, \theta^r\}\right]$. By decomposing the expectation into two events,

$$D = \mathbb{E}\left[\mathbf{1}_{\{X_\tau^l \leq X_\tau^r\}}(X_\tau^l - \max\{\theta^l, \theta^r\}) + \mathbf{1}_{\{X_\tau^r < X_\tau^l\}}(X_\tau^r - \max\{\theta^l, \theta^r\})\right].$$

Plugging in the definition of X_τ^i and using the law of iterated expectations,

$$D = \mathbb{E}\left[\mathbf{1}_{\{X_\tau^l \leq X_\tau^r\}}(\mathbb{E}[\theta^l|\mathcal{F}_\tau] - \max\{\theta^l, \theta^r\}) + \mathbf{1}_{\{X_\tau^r < X_\tau^l\}}(\mathbb{E}[\theta^r|\mathcal{F}_\tau] - \max\{\theta^l, \theta^r\})\right]$$

$$= \mathbb{E}\left[\mathbf{1}_{\{X_\tau^l \leq X_\tau^r\}}(\mathbb{E}[\theta^l|\mathcal{F}_\tau] - \mathbb{E}[\max\{\theta^l, \theta^r\}|\mathcal{F}_\tau]) + \mathbf{1}_{\{X_\tau^r < X_\tau^l\}}(\mathbb{E}[\theta^r|\mathcal{F}_\tau] - \mathbb{E}[\max\{\theta^l, \theta^r\}|\mathcal{F}_\tau])\right]$$

$$= \mathbb{E}\left[\mathbf{1}_{\{X_\tau^l \leq X_\tau^r\}}\mathbb{E}[-(\theta^r - \theta^l)^+|\mathcal{F}_\tau] + \mathbf{1}_{\{X_\tau^r < X_\tau^l\}}\mathbb{E}[-(\theta^l - \theta^r)^+|\mathcal{F}_\tau]\right]$$

$$= \mathbb{E}\left[-\mathbf{1}_{\{X_\tau^l \leq X_\tau^r\}}(\theta^r - \theta^l)^+ - \mathbf{1}_{\{X_\tau^r < X_\tau^l\}}(\theta^l - \theta^r)^+\right].$$

B.6   Proof of Fact 3

We rely on Bather's (1962) analysis of the Chernoff model, which by Fact 4 applies to our model. Bather studies a model with zero prior precision. Since such an agent never stops instantaneously, all that matters is his beliefs at $t > 0$, which are well defined even in this case, and given by $\hat{X}_t^i = t^{-1}Z_t^i$ and $\hat{\sigma}_t^{-2} = t\alpha^{-2}$. In Section 6, p. 619 Bather (1962) shows that

$$k\left(t, c, \infty, \frac{1}{\sqrt{2}}\right)\sqrt{t} = \frac{1}{4\,c\,t^{3/2}} + O\left(\frac{1}{t^3}\right).$$

which implies that

$$k\left(t, c, \infty, \frac{1}{\sqrt{2}}\right) = \frac{1}{4\,c\,t^2} + O\left(\frac{1}{t^{7/2}}\right).$$

Fix $\alpha > 0$. By equation (8) we have $k(t, c, \infty, \alpha) = \alpha\sqrt{2}k(t, \frac{1}{\alpha\sqrt{2}}c, \infty, \frac{1}{\sqrt{2}})$. Thus,

$$k\left(t, c, \infty, \alpha\right) = \frac{1}{2\,c\alpha^{-2}\,t^2} + O\left(\frac{1}{t^{7/2}}\right).$$

This implies that there exists $T, \beta > 0$ such that for all $t > T$ we have

$$\left|k\left(t, c, \infty, \alpha\right) - \frac{1}{2\,c\alpha^{-2}\,t^2}\right| \leq \frac{\beta}{t^{7/2}}.$$

Fix $\sigma_0 > 0$ and let $s := t - \alpha^2\sigma_0^{-2}$. This way, the agent who starts with zero prior precision and waits t seconds has the same posterior precision as the agent who starts with σ_0^2 and waits s

seconds.[26] Thus, by (6) we have $k(t, c, \infty, \alpha) = k(s, c, \sigma_0, \alpha)$, so

$$\left| k(s, c, \sigma_0, \alpha) - \frac{1}{2\, c\alpha^2\, (\alpha^{-2}s + \sigma_0^{-2})^2} \right| \leq \frac{\beta}{(s + \alpha^2\sigma_0^{-2})^{7/2}}.$$

Finally, since $b(s, c, \sigma_0, \alpha) = \alpha^2 k(s, c, \sigma_0, \alpha)\sigma_s^{-2}$, we have

$$\left| b(s, c, \sigma_0, \alpha) - \frac{1}{2\, c\alpha^2\, (\alpha^{-2}s + \sigma_0^{-2})} \right| \leq \frac{\beta}{(s + \alpha^2\sigma_0^{-2})^{5/2}}.$$

To see that (10) and (11) hold, notice that by equations (6), (8), (7) and (9) applied in that order, it follows that

$$\bar{k}(t, c, \sigma_0, \alpha) = \bar{k}(0, c, \sigma_t, \alpha) = \alpha\bar{k}(0, \alpha^{-1}c, \alpha^{-1}\sigma_t, 1) = \sigma_t\bar{k}(0, \alpha^2 c\sigma_t^{-3}, 1, 1)$$
$$= \alpha^{-2}c^{-1}\sigma_t^4\bar{k}(0, 1, 1, 1) = \frac{\kappa}{c\alpha^2(\sigma_0^{-2} + \alpha^{-2}t)^2},$$

where $\kappa = \bar{k}(0, 1, 1, 1)$. Since $\bar{b}(t, c, \sigma_0, \alpha) = \alpha^2\bar{k}(t, c, \sigma_0, \alpha)\sigma_t^{-2}$, it follows that $\bar{b}(t, c, \sigma_0, \alpha) = \frac{\kappa}{c(\sigma_0^{-2}+\alpha^{-2}t)}$. The fact that $\kappa = \frac{1}{2}$ follows from the proof of Fact 3, as any other constant would result in a contradiction as $t \to \infty$.

B.7   Proof of Theorem 6

Let $G = \{t_n\}_{n=1}^N$ be a finite set of times at which the agent is allowed to stop and denote by $\mathcal{T}$ all stopping times τ such that $\tau \in G$ almost surely. As we restrict the agent to stopping times in $\mathcal{T}$, the stopping problem becomes a discrete time optimal stopping problem. By Doob's optional sampling theorem we have that

$$\sup_\tau \mathbb{E}\left[\max\{X_\tau^l, X_\tau^r\} - d(\tau)\right] = \sup_\tau \mathbb{E}\left[\frac{1}{2}\max\{X_\tau^l - X_\tau^r, X_\tau^r - X_\tau^l\} + \frac{1}{2}(X_\tau^l + X_\tau^r) - d(\tau)\right]$$
$$= \sup_\tau \mathbb{E}\left[\frac{1}{2}|X_\tau^l - X_\tau^r| - d(\tau)\right] + \frac{1}{2}(X_0^l + X_0^r),$$

so any optimal stopping time also solves $\sup_\tau \mathbb{E}\left[|X_\tau^l - X_\tau^r| - 2\, d(\tau)\right]$. Define $\Delta_n = |X_{t_n}^l - X_{t_n}^r|$ for all $n = 1, \ldots, N$. Observe that $(\Delta_n)_{n=1,\ldots,N}$ is a one-dimensional discrete time Markov process. To prove that for every barrier there exists a cost function which generates b by Theorem 1 in Kruse and Strack (2015) it suffices to prove that:

1. there exists a constant C such that $\mathbb{E}[\Delta_{n+1}|\mathcal{F}_{t_n}] \leq C(1 + \Delta_n)$.

2. Δ_{n+1} is increasing in Δ_n in the sense of first order stochastic dominance

3. $z(n, y) = \mathbb{E}[\Delta_{n+1} - \Delta_n \mid \Delta_n = y]$ is strictly decreasing in y.

---

[26]To see this, observe that $\sigma_s^2 = \frac{1}{\sigma_0^{-2}+\alpha^{-2}s^2} = \frac{1}{\alpha^{-2}t} = \hat{\sigma}_t^2$.

Condition 1 keeps the value of continuing from exploding, which would be inconsistent with a finite boundary. Conditions 2 and 3 combined ensure that the optimal policy is a cut-off rule.

B.7.1 Certain-Difference DDM

Set $Z_t = Z_t^l - Z_t^r = (\theta'' - \theta')t + \sqrt{2}\alpha B_t$. Then

$$
\begin{aligned}
l_t &= \log\left(\frac{\mathbb{P}\left[\theta = \theta_l \mid \mathcal{F}_t\right]}{\mathbb{P}\left[\theta = \theta_r \mid \mathcal{F}_t\right]}\right) = \log\left(\frac{\mu}{1-\mu}\right) + \log\left(\frac{\exp(-(4\alpha^2 t)^{-1})(Z_t - (\theta'' - \theta')t)^2}{\exp(-(4\alpha^2 t)^{-1})(Z_t - (\theta' - \theta'')t)^2}\right) \\
&= \log\left(\frac{\mu}{1-\mu}\right) + \frac{Z_t(\theta'' - \theta')}{\alpha^2}.
\end{aligned}
$$

Denote by $p_n = \mathbb{P}[\theta = \theta_l \mid \mathcal{F}_{t_n}]$ the posterior probability that l is the better choice. The expected absolute difference of the two choices satisfies

$$
\begin{aligned}
\Delta_n &= |X_{t_n}^l - X_{t_n}^r| = |p_n(\theta'' - \theta') + (1 - p_n)(\theta' - \theta'')| \\
&= |(2p_n - 1)(\theta'' - \theta')| = 2(\theta'' - \theta')\left|p_n - \frac{1}{2}\right|.
\end{aligned}
$$

Let $\psi_n := [Z_{t_n}^l - Z_{t_n}^r] - [Z_{t_{n-1}}^l - Z_{t_{n-1}}^r]$ denote the change in the signal from t_{n-1} to t_n. We have that the log likelihood is given by $l_{n+1} = l_n + \alpha^{-2}(\theta'' - \theta')\psi_{n+1}$. We thus have

$$
\Delta_n = 2(\theta'' - \theta')\left|\frac{e^{l_n}}{1 + e^{l_n}} - \frac{1}{2}\right| = 2(\theta'' - \theta')\left(\frac{e^{|l_n|}}{1 + e^{|l_n|}} - \frac{1}{2}\right). \tag{16}
$$

(1): It is easily seen that $\mathbb{E}\left[\Delta_{n+1}|\mathcal{F}_{t_n}\right] \leq (\theta'' - \theta')$, so for C big enough, $\mathbb{E}[\Delta_{n+1}|\mathcal{F}_{t_n}] \leq C(1 + \Delta_n)$.
(2): To simplify notation we introduce $m_n = |l_n|$. The process $(m_n)_{n=1,\ldots,N}$ is Markov. More precisely, $m_{n+1} = |m_n + \alpha^{-2}(\theta'' - \theta')\psi_{n+1}|$ is folded normal with mean of the underlying normal distribution equal to

$$
\begin{aligned}
m_n + \alpha^{-2}(\theta'' - \theta')\mathbb{E}[\psi_{n+1}|l_n] &= |l_n| + \alpha^{-2}(\theta'' - \theta')\Delta_n(t_{n+1} - t_n) \\
&= m_n + \alpha^{-2}\left(\frac{2e^{m_n}}{1 + e^{m_n}} - 1\right)(\theta'' - \theta')^2(t_{n+1} - t_n) \tag{17}
\end{aligned}
$$

and variance

$$
\begin{aligned}
Var\left[m_n + \alpha^{-2}(\theta'' - \theta')\psi_{n+1}\right] &= \alpha^{-4}(\theta'' - \theta')^2 Var\left[\psi_{n+1}\right] \\
&= 2\alpha^{-4}(\theta'' - \theta')^2(t_{n+1} - t_n).
\end{aligned}
$$

As argued in part (2) of the uncertain difference case, a folded normal random variable increases in the sense of first order stochastic dominance in the mean of the underlying normal distribution. As (17) increases in m_n it follows that m_{n+1} increases in m_n in the sense of first order stochastic dominance. By (16) $m_n = |l_n|$ is increasing in Δ_n and Δ_{n+1} is increasing in $m_{n+1} = |l_{n+1}|$ this completes the argument.
(3): It remains to show that $z(n, \Delta_n)$ is decreasing in Δ_n. As $(p_n)_{n=1,\ldots,M}$ is a martingale, and

moreover conditioning on p is equivalent to conditioning on $1 - p$, we have that

$$z(n, \Delta_n) = \mathbb{E}\left[\Delta_{n+1} \,\Big|\, p_i = \frac{\Delta_i}{2(\theta'' - \theta')} + \frac{1}{2}\right] - \Delta_n$$

$$= 2(\theta'' - \theta')\mathbb{E}\left[|p_{n+1} - \frac{1}{2}|\,\Big|\, p_n = \frac{\Delta_n}{2(\theta'' - \theta')} + \frac{1}{2}\right] - \Delta_n$$

$$= 2(\theta'' - \theta')\mathbb{E}\left[p_{n+1} - \frac{1}{2}\,\Big|\, p_i = \frac{\Delta_i}{2(\theta'' - \theta')} + \frac{1}{2}\right]$$

$$+ 2(\theta'' - \theta')\mathbb{E}\left[2\max\left\{\frac{1}{2} - p_{n+1}, 0\right\}\,\Big|\, p_i = \frac{\Delta_n}{(\theta'' - \theta')} + \frac{1}{2}\right] - \Delta_n\,.$$

As p is a martingale we can replace p_{n+1} by p_n

$$z(n, \Delta_n) = 2(\theta'' - \theta')\mathbb{E}\left[p_n - \frac{1}{2}\,\Big|\, p_n = \frac{\Delta_n}{2(\theta'' - \theta')} + \frac{1}{2}\right]$$

$$+ 2(\theta'' - \theta')\mathbb{E}\left[2\max\left\{\frac{1}{2} - p_{n+1}, 0\right\}\,\Big|\, p_i = \frac{\Delta_n}{2(\theta'' - \theta')} + \frac{1}{2}\right] - \Delta_n$$

$$= \Delta_n + 2(\theta'' - \theta'))\mathbb{E}\left[2\max\left\{\frac{1}{2} - p_{n+1}, 0\right\}\,\Big|\, p_n = \frac{\Delta_n}{2(\theta'' - \theta')} + \frac{1}{2}\right] - \Delta_n$$

$$= 4(\theta'' - \theta')\mathbb{E}\left[\max\left\{\frac{1}{2} - p_{n+1}, 0\right\}\,\Big|\, p_n = \frac{\Delta_n}{2(\theta'' - \theta')} + \frac{1}{2}\right]\,.$$

The above term is strictly decreasing in Δ_n as p_{n+1} increases in the sense of first order stochastic dominance in p_n and p_n in the conditional expectation is increasing in Δ_n.

B.7.2   Uncertain-Difference DDM

Let us further define $\beta_i^2 = 2\sigma_{t_i}^2 - 2\sigma_{t_{i+1}}^2$. As $X_{t_i+1}^l - X_{t_i+1}^r$ is Normal distributed with variance β_i^2 and mean Δ_i we have that Δ_{i+1} is folded normal distributed with mean

$$\mathbb{E}_i\left[\Delta_{i+1}\right] = \beta_i\sqrt{\frac{2}{\pi}}e^{-\frac{\Delta_i^2}{2\beta_i^2}} + \Delta_i(1 - 2\Phi(\frac{-\Delta_i}{\beta_i}))\,,$$

where Φ denotes the normal cdf. Thus, the expected change in delta is given by

$$z(i, y) = \beta_i\sqrt{\frac{2}{\pi}}e^{-\frac{y^2}{2\beta_i^2}} - 2\,y\,\Phi(\frac{-\Delta_i}{\beta_i})\,.$$

(1): It is easily seen that $\mathbb{E}_i\left[\Delta_{i+1}\right] \leq \beta_i\sqrt{\frac{2}{\pi}} + \Delta_i$.

(2): As Δ_i is folded normal distributed we have that

$$\mathbb{P}_i(\Delta_{i+1} \leq y) = \frac{1}{2}\left[\mathrm{erf}\left(\frac{y + \Delta_i}{\beta_i}\right) + \mathrm{erf}\left(\frac{y - \Delta_i}{\beta_i}\right)\right]\,.$$

Taking derivatives gives that

$$\frac{\partial}{\partial \Delta_i}\mathbb{P}_i(\Delta_{i+1} \leq y) = \frac{1}{2}\left[e^{-\left(\frac{y+\Delta_i}{\beta_i}\right)^2} - e^{-\left(\frac{y-\Delta_i}{\beta_i}\right)^2}\right] = \frac{1}{2}e^{-\left(\frac{y-\Delta_i}{\beta_i}\right)^2}\left[e^{-\frac{4\Delta_i y}{\beta_i}} - 1\right] < 0\,.$$

As $\Delta_i = |X^l_{t_i} - X^r_{t_i}|$ it follows that $y \geq 0$ and hence, Δ_{i+1} is increasing in Δ_i in the sense of first order stochastic dominance.

(3): The derivative of the expected change of the process Δ equals

$$\frac{\partial}{\partial y}z(i,y) = \frac{\partial}{\partial y}\left(\beta_i\sqrt{\frac{2}{\pi}}e^{-\frac{y^2}{2\beta_i^2}} - 2\,y\,\Phi(\frac{-\Delta_i}{\beta_i}))\right) = -2\Phi(\frac{-y}{\beta_i}) < 0\,.$$

Hence, z is strictly decreasing in y. $\qquad\qquad\square$

References

Arrow, K., D. Blackwell, and M. Girshick (1949): "Bayes and minimax solutions of sequential decision problems," *Econometrica*, 17, 213–244.

Ballinger, T. P., and N. T. Wilcox (1997): "Decisions, error and heterogeneity," *The Economic Journal*, 107(443), 1090–1105.

Bather, J. A. (1962): "Bayes procedures for deciding the sign of a normal mean," in *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 58, pp. 599–620. Cambridge Univ Press.

Bogacz, R., E. Brown, J. Moehlis, P. Holmes, and J. D. Cohen (2006): "The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced-choice tasks.," *Psychological review*, 113(4), 700.

Busemeyer, J. R., and J. G. Johnson (2004): "Computational models of decision making," *Blackwell handbook of judgment and decision making*, pp. 133–154.

Busemeyer, J. R., and J. T. Townsend (1992): "Fundamental derivations from decision field theory," *Mathematical Social Sciences*, 23(3), 255–282.

Caplin, A., and M. Dean (2011): "Search, choice, and revealed preference," *Theoretical Economics*, 6(1), 19–48.

Caplin, A., and D. Martin (2014): "The Dual-Process Drift Diffusion Model: Evidence from Response Times," .

Cheremukhin, A., A. Popova, and A. Tutino (2011): "Experimental evidence on rational inattention," *Federal Reserve Bank of Dallas Working Paper*, 1112.

Chernoff, H. (1961): "Sequential tests for the mean of a normal distribution," in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, pp. 79–91. Univ of California Press.

Churchland, A. K., R. Kiani, and M. N. Shadlen (2008): "Decision-making with multiple alternatives," *Nature neuroscience*, 11(6), 693–702.

CLITHERO, J. A., AND A. RANGEL (2013): "Combining Response Times and Choice Data Using A Neuroeconomic Model of the Decision Process Improves Out-of-Sample Predictions," *mimeo.*

DITTERICH, J. (2006): "Stochastic models of decisions about motion direction: behavior and physiology," *Neural Networks*, 19(8), 981–1012.

DRUGOWITSCH, J., R. MORENO-BOTE, A. K. CHURCHLAND, M. N. SHADLEN, AND A. POUGET (2012): "The cost of accumulating evidence in perceptual decision making," *The Journal of Neuroscience*, 32(11), 3612–3628.

DUDLEY, R. M. (2002): *Real analysis and probability*, vol. 74. Cambridge University Press.

EDWARDS, W. (1965): "Optimal strategies for seeking information: Models for statistics, choice reaction times, and human information processing," *Journal of Mathematical Psychology*, 2(2), 312–329.

FEHR, E., AND A. RANGEL (2011): "Neuroeconomic foundations of economic choicerecent advances," *The Journal of Economic Perspectives*, 25(4), 3–30.

GABAIX, X., AND D. LAIBSON (2005): "Bounded rationality and directed cognition," *Harvard University.*

HANES, D. P., AND J. D. SCHALL (1996): "Neural control of voluntary movement initiation," *Science*, 274(5286), 427–430.

HEY, J. D. (1995): "Experimental investigations of errors in decision making under risk," *European Economic Review*, 39(3), 633–640.

——— (2001): "Does repetition improve consistency?," *Experimental economics*, 4(1), 5–54.

ISCOE, I., A. KREININ, AND D. ROSEN (1999): "An integrated market and credit risk portfolio model," *Algo Research Quarterly*, 2(3), 21–38.

KE, T., Z.-J. M. SHEN, AND J. M. VILLAS-BOAS (2013): "Search for Information on Multiple Products," Discussion paper, Working paper, University of California, Berkeley.

KRAJBICH, I., C. ARMEL, AND A. RANGEL (2010): "Visual fixations and the computation and comparison of value in simple choice," *Nature neuroscience*, 13(10), 1292–1298.

KRAJBICH, I., D. LU, C. CAMERER, AND A. RANGEL (2012): "The attentional drift-diffusion model extends to simple purchasing decisions," *Frontiers in psychology*, 3, 193.

KRAJBICH, I., AND A. RANGEL (2011): "Multialternative drift-diffusion model predicts the relationship between visual fixations and choice in value-based decisions," *Proceedings of the National Academy of Sciences*, 108(33), 13852–13857.

KRUSE, T., AND P. STRACK (2014): "Inverse Optimal Stopping," *arXiv preprint arXiv:1406.0209.*

——— (2015): "Optimal stopping with private information," *forthcoming in Journal of Economic Theory.*

LAMING, D. R. J. (1968): *Information theory of choice-reaction times.* Academic Press.

LINK, S., AND R. HEATH (1975): "A sequential theory of psychological discrimination," *Psychometrika*, 40(1), 77–105.

LIPTSER, R. S., AND A. SHIRYAEV (2001): "Statistics of stochastic processes," *Springer-Verlag*, 1.

LU, J. (2013): "Random Choice and Private Information," *Working paper, Princeton University.*

LUCE, R. D. (1986): *Response times.* Oxford University Press.

MILOSAVLJEVIC, M., J. MALMAUD, A. HUTH, C. KOCH, AND A. RANGEL (2010a): "The drift diffusion model can account for value-based choice response times under high and low time pressure," *Judgement & Decision Making*, 5, 437–449.

———— (2010b): "The drift diffusion model can account for value-based choice response times under high and low time pressure," *Judgement & Decision Making*, 5, 437–449.

NATENZON, P. (2013): "Random choice and learning," *Working paper, University of Washington.*

PESKIR, G., ET AL. (2002): "On integral equation arising in the first-passage problem for Brownian motion," *Journal Integral Equations and Applications*, 14, 397–423.

PESKIR, G., AND A. SHIRYAEV (2006): *Optimal stopping and free-boundary problems.* Springer.

RAND, D. G., J. D. GREENE, AND M. A. NOWAK (2012): "Spontaneous giving and calculated greed," *Nature*, 489(7416), 427–430.

RAPOPORT, A., AND G. J. BURKHEIMER (1971): "Models for deferred decision making," *Journal of Mathematical Psychology*, 8(4), 508–538.

RATCLIFF, R. (1978): "A theory of memory retrieval.," *Psychological review*, 85(2), 59.

RATCLIFF, R., A. CHERIAN, AND M. SEGRAVES (2003): "A comparison of macaque behavior and superior colliculus neuronal activity to predictions from models of two-choice decisions," *Journal of neurophysiology*, 90(3), 1392–1407.

RATCLIFF, R., AND G. MCKOON (2008): "The diffusion decision model: Theory and data for two-choice decision tasks," *Neural computation*, 20(4), 873–922.

REUTSKAJA, E., R. NAGEL, C. F. CAMERER, AND A. RANGEL (2011): "Search dynamics in consumer choice under time pressure: An eye-tracking study," *The American Economic Review*, 101(2), 900–926.

REVUZ, D., AND M. YOR (1999): *Continuous martingales and Brownian motion*, vol. 293. Springer.

RUBINSTEIN, A. (2007): "Instinctive and cognitive reasoning: A study of response times," *The Economic Journal*, 117(523), 1243–1259.

SHADLEN, M. N., T. D. HANKS, A. K. CHURCHLAND, R. KIANI, AND T. YANG (2006): "The speed and accuracy of a simple perceptual decision: a mathematical primer," *Bayesian brain: Probabilistic approaches to neural coding*, pp. 209–37.

SHADLEN, M. N., AND R. KIANI (2013): "Decision Making as a Window on Cognition," *Neuron*, 80(3), 791–806.

SHIRYAEV, A. N. (1969): *Optimal stopping rules.* Russian Edition published by Nauka.

——— (2007): *Optimal stopping rules*. Springer.

SMITH, P. L. (2000): "Stochastic dynamic models of response time and accuracy: A foundational primer," *Journal of mathematical psychology*, 44(3), 408–463.

STONE, M. (1960): "Models for choice-reaction time," *Psychometrika*, 25(3), 251–260.

SWENSSON, R. G. (1972): "The elusive tradeoff: Speed vs accuracy in visual discrimination tasks," *Perception & Psychophysics*, 12(1), 16–32.

VICKERS, D. (1970): "Evidence for an accumulator model of psychophysical discrimination," *Ergonomics*, 13(1), 37–58.

VUL, E., N. GOODMAN, T. L. GRIFFITHS, AND J. B. TENENBAUM (2014): "One and done? Optimal decisions from very few samples," *Cognitive science*, 38(4), 599–637.

WALD, A. (1947): *Sequential analysis*. John Wiley & Sons.

WEBB, R. (2013): "Dynamic constraints on the distribution of stochastic choice: Drift Diffusion implies Random Utility," *Available at SSRN 2226018*.

WOODFORD, M. (2014): "An Optimizing Neuroeconomic Model of Discrete Choice," Discussion paper, National Bureau of Economic Research.