Optimal Learning under Robustness and Time-Consistency*

Larry G. Epstein   Shaolin Ji

March 1, 2018

Abstract

We model learning in a continuous-time Brownian setting where there is prior ambiguity. The associated model of preference values robustness and is time-consistent. The model is applied to study optimal learning when the choice between actions can be postponed, at a per-unit-time cost, in order to observe a signal that provides information about an unknown parameter. The corresponding optimal stopping problem is solved in closed-form in two specific settings: Ellsberg's two-urn thought experiment expanded to allow learning before the choice of bets, and a robust version of the classical problem of sequential testing of two simple hypotheses about the unknown drift of a Wiener process. In both cases, the link between robustness and the demand for learning is the focus.

Key words: ambiguity, robust decisions, learning, partial information, optimal stopping, sequential testing of simple hypotheses, Ellsberg Paradox, recursive utility, time consistency, model uncertainty

---

*Department of Economics, Boston University, lepstein@bu.edu and Zhongtai Securities Institute of Financial Studies, Shandong University, jsl@sdu.edu.cn. Ji gratefully acknowledges the financial support of the National Natural Science Foundation of China (award No. 11571203). We received helpful comments from Tomasz Strzalecki and excellent research assistance from Haodong Liu and Fernando Payro.

1. Introduction

We consider a decision-maker (DM) choosing between actions whose payoffs are uncertain because they depend on both exogenous randomness and on an unknown parameter θ. She can postpone the choice of action so as to learn about θ by observing the realization of a signal modeled by a Brownian motion with drift. Because of a per-unit-time cost of sampling, which can be material or cognitive, she faces an optimal stopping problem. A key feature is that DM does not have sufficient information to arrive at a single prior about θ, that is, there is ambiguity about θ. Therefore, prior beliefs are represented by a nonsingleton set of probability measures, and DM seeks to make robust choices of both stopping time and action by solving a maxmin problem. In addition, she is forward-looking and reasons by backward induction, as in the continuous-time version of maxmin utility given by Chen and Epstein (2002). One contribution herein is to extend the latter model to accommodate learning. As a result, we capture robustness to ambiguity (or model uncertainty), learning and time-consistency. The other contribution is to investigate optimal learning in two specific settings, outlined next, where the corresponding optimal stopping problems are solved explicitly and the effects of ambiguity on optimal learning are determined.

The first specific context begins with Ellsberg's metaphorical thought experiment: *There are two urns, each containing balls that are either red or blue, where the "known" or risky urn contains an equal number of red and blue balls, while no information is provided about the proportion of red balls in the "unknown" or ambiguous urn. DM must choose between betting on the color drawn from the risky urn or from the ambiguous urn.* The intuitive behavior highlighted by Ellsberg is the choice to bet on the draw from the risky urn no matter the color, which behavior is paradoxical for subjective expected utility theory, or indeed, for any model in which beliefs are represented by a single probability measure. Ellsberg's paradox is often taken as a normative critique of the Bayesian model and of the view that the single prior representation of beliefs is implied by rationality (e.g., Gilboa 2009, 2015; Gilboa et al. 2012). Here we add to the thought experiment by including a possibility to learn. Specifically, we allow DM to postpone her choice so that she can observe realizations of a diffusion process whose drift is equal to the proportion of red in the ambiguous urn. Under specific parametric restrictions we completely describe the optimal joint learning and betting strategy. In particular, we show that it is optimal to reject the opportunity to learn if and only if ambiguity aversion (suitably measured) exceeds a cut-off level. The rationality of no learning suggests that one needs to reexamine and qualify the common presumption that ambiguity would fade away, or at least diminish, in the presence of learning opportunities (Marinacci 2002). It can also explain experimental findings (Trautman and Zeckhauser 2013) that some subjects neglect opportunities to learn about an ambiguous urn even at no visible (material) cost. In addition, our model is suggestive of laboratory experiments that could provide further evidence on the connection between ambiguity and the demand for learning.

The second application is to the classical problem of sequential testing of two simple hypotheses about the unknown drift of a Wiener process. Wald (1947) introduced the sequential probability ratio test (SPRT), where desired error probabilities determine the cut-off values for deciding when to stop sampling and accept one of the hypotheses; Peskir and Shiryaev (2006, Ch. 6) analyse a Bayesian subjectivist approach and derive SPRT as the solution to an optimal stopping problem. We extend the latter analysis to accommodate situations where the statistician/analyst does not have sufficient information to justify reliance on a single prior, and we show that a desire for robustness lengthens optimal sampling relative to what would be chosen by any "compatible" Bayesian. In particular, under robustness it may be optimal to continue sampling even given a realized sample at which ALL compatible Bayesians would choose to stop. Because this application is clearly prescriptive, we emphasize that the Chen-Epstein model has axiomatic foundations in the sense that its discrete-time counterpart model (Epstein and Schneider 2003) is axiomatic, which foundations endow the present model with corresponding credentials as a prescriptive model.

It is well-known that modeling a concern with ambiguity and robust decision-making leads to "nonlinear" objective functions, which, in a dynamic setting and in the absence of commitment, can lead to time-inconsistency issues. See Peskir (2017) for analyses of optimal stopping problems featuring such time inconsistency. A similar issue arises also in a risk context where there is a known objective probability law, but where preference does not conform to von Neumann-Morgenstern's expected utility theory (Ebert and Strack 2017). Such models are problematic in some normative contexts. It is not clear why one would ever *prescribe* to a decision-maker (who is unable or unwilling to commit) that she *should* adopt a criterion function that would imply time-inconsistent plans and that she *should* then resolve these inconsistencies by behaving strategically against her future selves (as is commonly assumed). In descriptive contexts, one might view time-inconsistent models as appealing because they capture some form of bounded rationality. But such a view seems (to us) misplaced if one adopts the sophisticated game-theoretic approach to predicting behavior (as, for example, in the two works just cited). The latter assumption does not seem cognitively less demanding than assuming that preference is arrived at after reasoning about the future through backward induction, which underlies the recursive Chen-Epstein model of utility.

It is important to understand that, roughly speaking, time-consistency is the requirement that a contingent plan (e.g., a stopping strategy) that is optimal ex ante remain optimal conditional on every subsequent realization *assuming there are no surprises or unforeseen events*. A possible argument against such consistency, (that is sometimes expressed in the statistics literature), is that surprises are inevitable and thus that any prescription should take that into account rather than excluding their possibility. We would agree that a sophisticated decision-maker would expect that surprises may occur while (necessarily) being unable to describe what form they could take. However, to the best of our knowledge there currently does not exist a convincing model in the economics, statistics or psychology literatures of how such an individual should (or would) behave, that is, how the awareness that she may be missing something in her perception of the future should (or would) affect current behavior. That leaves time-consistency as a sensible guiding principle with the understanding that reoptimization can (and should) occur if there is a surprise.

The recursive maxmin model has been used in macroeconomics and finance (e.g., Epstein and Schneider 2010) and also in robust multistage stochastic optimization (e.g., Shapiro (2016) and the references therein, including to the closely related literature on conditional risk measures). Shapiro focuses on a property of sets of measures, called rectangularity following Epstein and Schneider (2003), that underlies recursivity of utility and time-consistency. Most of the existing literature deals with a discrete-time setting. The theoretical literature on learning under ambiguity is sparse and limited to passive learning (e.g., Epstein and Schneider 2007, 2008). We are not aware of any work, whether in a discrete or continuous-time setting, that deals with *optimal* learning under robustness and time-consistency. With regard to hypothesis testing, this paper adds to the literature on robust Bayesian statistics (Berger 1984,1985,1994; Rios-Insua and Ruggeri 2000), which is largely restricted to a static environment. Walley (1991) goes further and considers both a prior and a single posterior stage, but not sequential hypothesis testing. For a frequentist approach to robust sequential testing see Huber (1965).

The paper proceeds as follows. The next section describes the model of utility extending Chen-Epstein to accommodate learning. Readers who are primarily interested in the two applications can skip this relatively technical section and move directly to §3 and §4 where the Ellsberg and hypothesis testing contexts are studied respectively. Proofs (and added details about the Ellsberg setting) are provided in three appendices.

2. Recursive utility with learning

Here we outline the model of recursive utility under ambiguity due to Chen and Epstein (2002)–CE below– and then we describe how it can be adapted to include learning with partial information. The latter description is given in the simplest context adequate for the applications below. However, it should be clear that it can be adapted more generally.

Let $(\Omega, \mathcal{G}_\infty, P_0)$ be a probability space, and $W = (W_t)_{0 \leq t < \infty}$ a 1-dimensional Brownian motion which generates the filtration $\mathcal{G} = \{\mathcal{G}_t\}_{t \geq 0}$, with $\mathcal{G}_t \nearrow \mathcal{G}_\infty$. (All probability spaces are taken to be complete and all related filtrations are augmented in the usual sense.) The measure P_0 is a reference measure whose role is only to define null events.

CE define a set of predictive priors $\mathcal{P}_0$ on $(\Omega, \mathcal{G}_\infty)$ through specification of their densities with respect to P_0. (We adopt the common practice of distinguishing terminologically between beliefs about the state space, referred to as predictive priors, and beliefs about parameters, which are referred to as priors.) To do so, they take as an additional primitive a (suitably adapted) set-valued process (Ξ_t). (Technical restrictions are that $\Xi_t : \Omega \rightsquigarrow K \subset \mathbb{R}^d$ for some compact set K independent of t, $0 \in \Xi_t(\omega)\ dt \otimes dP_0\ a.s.$, and that each Ξ_t is convex- and compact-valued.) Define the associated set of real-valued processes by

$$\Xi = \{\eta = (\eta_t) \mid \eta_t(\omega) \in \Xi_t(\omega)\ dt \otimes dP_0\ a.s.\}.$$

Then each $\eta \in \Xi$ defines a probability measure on $\mathcal{G}_\infty$, denoted P^η, that is equivalent to P_0 and given by

$$\frac{dP^\eta}{dP_0} \mid_{\mathcal{G}_t} = \exp\{-\int_0^t \eta_s^2 ds - \int_0^t \eta_s dW_s\} \text{ for all } t.$$

Accordingly, each $\eta_t(\omega) \in \Xi_t(\omega)$ can be thought of roughly as defining conditional beliefs about $\mathcal{G}_{t+dt}$, and $\Xi_t(\omega)$ is called the set of *density generators* at (t, ω). By the Girsanov Theorem,

$$dW_t^\eta = \eta_t dt + dW_t \tag{2.1}$$

is a Brownian motion under P^η, which thus can be understood as an alternative hypothesis about the drift of the driving process W (the drift is 0 under P_0). Finally,

$$\mathcal{P}_0 \equiv \{P^\eta : \eta \in \Xi\}. \tag{2.2}$$

The set $\mathcal{P}_0$ is used to define a time 0 utility function on a suitable set of random payoffs denominated in utils. In order to model in the sequel the choice of how long to learn (or sample), we consider a set of stopping times τ, that is, each τ is an adapted $\mathbb{R}_+$-valued and $\{\mathcal{G}_t\}$-adapted random variable defined on Ω, that is, $\{\omega : \tau(\omega) > t\} \in \mathcal{G}_t$ for every t. For each such τ, utility is defined on the set $L(\tau)$ of real-valued random variables given by

$$L(\tau) = \{\xi \mid \xi \text{ is } \mathcal{G}_\tau\text{-measurable and } \sup_{Q \in \mathcal{P}_0} E_Q \mid \xi \mid < \infty\}.$$

The time 0 utility of any $\xi \in L(\tau)$ is given by

$$U_0(\xi) = \inf_{Q \in \mathcal{P}_0} E_Q \xi = -\sup_{Q \in \mathcal{P}_0} E_Q[-\xi]. \tag{2.3}$$

It is natural to consider also conditional utilities at each (t, ω), where

$$U_t(\xi) = \operatorname*{ess\,inf}_{Q \in \mathcal{P}_0} E_Q[\xi \mid \mathcal{G}_t]. \tag{2.4}$$

In words, $U_t(\xi)$ is the utility of ξ at time t conditional on the information available then and given the state ω (the dependence of $U_t(\xi)$ on ω is suppressed notationally). The special construction of $\mathcal{P}_0$ delivers the following counterpart of the law of total probability (or law of iterated expectations): For each ξ, and $0 \leq t < t'$,

$$U_t(\xi) = \underset{Q \in \mathcal{P}_0}{\operatorname{ess\,inf}} E_Q \left[U_{t'}(\xi) \mid \mathcal{G}_t \right]. \tag{2.5}$$

This recursivity ultimately delivers the time-consistency of optimal choices.

The components P_0, W, (Ξ_t) and $\{\mathcal{G}_t\}$ are primitives in CE. Next we specify them in terms of the deeper primitives of a model that includes learning about an unknown parameter $\theta \in \Theta \subset \mathbb{R}$.

Specifically, begin with a measurable space $(\Omega, \mathcal{F})$, a filtration $\{\mathcal{F}_t\}$, $\mathcal{F}_t \nearrow \mathcal{F}_\infty \subset \mathcal{F}$, and a collection $\{P^\mu : \mu \in \mathcal{M}_0\}$ of pairwise equivalent probability measures on $(\Omega, \mathcal{F})$. Though θ is an unknown deterministic parameter, for mathematical precision we view θ as a random variable on $(\Omega, \mathcal{F})$. Further, for each $\mu \in \mathcal{M}_0$, P^μ induces the distribution μ for θ via $\mu(A) = P^\mu(\{\theta \in A\})$ for all Borel measurable $A \subset \Theta$. Accordingly, $\mathcal{M}_0$ can be viewed as a set of priors on Θ, and its nonsingleton nature indicates ambiguity about θ. There is also a standard Brownian motion $B = (B_t)$, with generated filtration $\{\mathcal{F}_t^B\}$, such that B is independent of θ under each P^μ. B is the Brownian motion driving the signals process $Z = (Z_t)$ according to

$$Z_t = \int_0^t \theta ds + \int_0^t \sigma dB_s = \theta t + \sigma B_t, \tag{2.6}$$

where σ is a known positive constant. Because only realizations of Z_t are observable, take $\{\mathcal{G}_t\}$ to be the filtration generated by Z. Assuming knowledge of the signal structure, Bayesian updating of $\mu \in \mathcal{M}_0$ gives the posterior μ_t at time t. Thus prior-by-prior Bayesian updating leads to the set-valued process $(\mathcal{M}_t)$ of posteriors on θ.

Proceed to specify the other CE components P_0, W and (Ξ_t).

Step 1. Take $\mu \in \mathcal{M}_0$. By standard filtering theory (Liptser and Shiryaev 1977, Theorem 8.3), if we replace the unknown parameter θ by the estimate $\widehat{\theta}_t^\mu = \int \theta d\mu_t$, then we can rewrite (2.6) in the form

$$\begin{aligned} dZ_t &= \hat{\theta}_t^\mu (Z_t) dt + \sigma (dB_t + \frac{\theta - \hat{\theta}_t^\mu (Z_t)}{\sigma} dt) \\ &= \hat{\theta}_t^\mu (Z_t) dt + \sigma d\tilde{B}_t^\mu, \end{aligned} \tag{2.7}$$

where the innovation process $(\tilde{B}_t^\mu)$ is a standard $\{\mathcal{G}_t\}$-adapted Brownian motion on $(\Omega, \mathcal{G}_\infty, P^\mu)$. Thus $(\tilde{B}_t^\mu)$ takes the same role as (W_t^η) in CE (see (2.1) above). Rewrite (2.7) as

$$d\tilde{B}_t^\mu = -\frac{1}{\sigma} \hat{\theta}_t^\mu (Z_t) dt + \frac{1}{\sigma} dZ_t$$

which suggests that (Z_t/σ) (resp. $(-\hat{\theta}_t^{\mu}(Z_t)/\sigma)$) can be chosen as the Brownian motion (W_t) (resp. the drift (η_t)) in (2.1).

Step 2. Find a reference probability measure P_0 on $(\Omega, \mathcal{G}_\infty)$ under which (Z_t/σ) is a $\{\mathcal{G}_t\}$-adapted Brownian motion on $(\Omega, \mathcal{G}_\infty)$. Fix $\overline{\mu} \in \mathcal{M}_0$ and define P_0 by:

$$\begin{aligned} \tfrac{dP_0}{dP^{\overline{\mu}}} \mid_{\mathcal{G}_t} &= \exp\{-\tfrac{1}{2\sigma^2}\textstyle\int_0^t (\hat{\theta}_s^{\overline{\mu}}(Z_s))^2 ds - \tfrac{1}{\sigma}\int_0^t \hat{\theta}_s^{\overline{\mu}}(Z_s)\, d\tilde{B}_s^{\overline{\mu}}\} \\ &= \exp\{\tfrac{1}{2\sigma^2}\textstyle\int_0^t (\hat{\theta}_s^{\overline{\mu}}(Z_s))^2 ds - \tfrac{1}{\sigma^2}\int_0^t \hat{\theta}_s^{\overline{\mu}}(Z_s)\, dZ_s\}. \end{aligned}$$

By Girsanov's Theorem, (Z_t/σ) is a $\{\mathcal{G}_t\}$-adapted Brownian motion under P_0.

Step 3. Viewing P_0 as a reference measure, perturb it. For each $\mu \in \mathcal{M}_0$, define P_0^{μ} on $(\Omega, \mathcal{G}_\infty)$ by

$$\frac{dP_0^{\mu}}{dP_0} \mid_{\mathcal{G}_t} = \exp\{-\frac{1}{2\sigma^2}\int_0^t (\hat{\theta}_s^{\mu}(Z_s))^2 ds + \frac{1}{\sigma^2}\int_0^t \hat{\theta}_s^{\mu}(Z_s)\, dZ_s\}.$$

By Girsanov, $d\tilde{B}_t^{\mu} = -\frac{1}{\sigma}\hat{\theta}_t^{\mu}(Z_t)\, dt + \frac{1}{\sigma} dZ_t$ is a Brownian motion under P_0^{μ}.

In general, $P^{\mu} \neq P_0^{\mu}$. However, they induce the identical distribution for Z. This is because $(\tilde{B}_t^{\mu})$ is a $\{\mathcal{G}_t\}$-adapted Brownian motion under both P^{μ} and P_0^{μ}. Therefore, by the uniqueness of weak solutions to SDEs, the solution Z_t of (2.7) on $(\Omega, \mathcal{F}_\infty, P^{\mu})$ and the solution Z' of (2.7) on $(\Omega, \mathcal{G}_\infty, P_0^{\mu})$ have identical distributions. (Argue as in Oksendal (2005, Example 8.6.9). See his Section 5.3 for discussion of weak versus strong solutions of SDEs.) Given that only the distribution of signals matters in our model, there is no reason to distinguish between the two probability measures. Thus we apply CE to the following components: W and P_0 defined in Step 2, and Ξ_t given by

$$\Xi_t = \{-\hat{\theta}_t^{\mu}/\sigma : \mu \in \mathcal{M}_0, \widehat{\theta}_t^{\mu} = \int \theta d\mu_t\}. \tag{2.8}$$

In summary, taking these specifications for P_0, W, (Ξ_t) and $\{\mathcal{G}_t\}$ in the CE model yields a set $\mathcal{P}_0$ of predictive priors, and a corresponding utility function, that capture prior ambiguity about the parameter θ (through $\mathcal{M}_0$), learning as signals are realized (through updating to the set of posteriors $\mathcal{M}_t$), and robust (maxmin) and time-consistent decision-making (because of (2.5)). We use this model in the two applications that follow.

Remark 1. *We add a few remarks about related literature. Cheng and Reidel (2013) describe how CE can be applied to study optimal stopping, but they do not discuss learning. CE suggest (but do not prove) that their framework can accommodate passive learning. We are aware of two papers that explicitly address passive learning in the CE framework–Choi (2016) and Miao (2009)–whose models are much different than* *the above. Two core distinguishing features of Choi's model are: (i) his set of priors $\mathcal{M}_0$ consists exclusively of Dirac, or dogmatic, measures which naturally do not admit Bayesian updating; and (ii) ambiguity affects learning primarily because there are multiple-likelihoods, reflecting the assumption that the signal structure is not well understood. See the related discrete-time work of Epstein and Schneider (2007, 2008) for the distinction between prior ambiguity about an unknown parameter, as in our model, and ambiguity about the signal structure (or the likelihood function, as in Choi). Our focus on prior ambiguity derives from our objective–trying to understand the connection between ambiguity and (optimal) learning in the situation most favorable for learning which is that the signal structure is well understood.*

Miao focuses on partial information and filtering in the presence of ambiguity. In his approach, application of CE is immediate and partial information does not make much difference for the analysis. He applies classical filtering for a reference model and then adds time- and history-invariant ambiguity to the updated reference measure. There is no interaction between filtering and ambiguity; for example, the dependence of estimates on the prior μ as in (2.8) is absent.

3. Optimal learning and Ellsberg's urns

3.1. The setting

There are two urns each containing balls that are either red or blue: a risky urn in which the proportion of red balls is $\frac{1}{2}$ and an ambiguous urn in which the color composition is unknown. Denote by $\theta+\frac{1}{2}$ the unknown proportion of red balls, where $\theta \in \Theta = \left[-\frac{1}{2}, \frac{1}{2}\right]$ is the bias towards red: $\theta > 0$ indicates more red than blue, $\theta < 0$ indicates the opposite, and $\theta = 0$ indicates an equal number as in the risky urn. (We suppose that the number of balls in the ambiguous urn is large and treat θ as a continuous variable.) There is ambiguity about θ modeled by $\mathcal{M}_0 \subset \Delta(\Theta)$.

Before choosing between bets, DM is given the opportunity to postpone her choice so that she can learn about θ by observing realizations of the signal process Z given by (2.6). The underlying state space Ω, the filtration $\{\mathcal{G}_t\}$ generated by Z, and other notation are as in §2. Unless specified otherwise, all processes below are taken to be $\{\mathcal{G}_t\}$-adapted even where not stated explicitly.

There is a constant per-unit-time cost $c > 0$ of learning. If DM stops learning at t, then her conditional expected payoff (in utils) is X_t; think of X_t as the indirect utility she can attain by choosing optimally between the bets available at t. Her choice of when to stop is described by a stopping time (or strategy) τ; the set of all stopping strategies is Γ. DM is forward-looking and has time 0 beliefs about future signals given by the set $\mathcal{P}_0 \subset \Delta(\Omega, \mathcal{G}_\infty)$ described in the previous section. (As noted there, $\mathcal{P}_0$ is a continuous-time counterpart of the "rectangular" set of predictive priors introduced and axiomatized in a discrete-time setting by Epstein and Schneider (2003); in particular, it models a sophisticated forward-looking agent who can be thought of as reasoning by backward induction.) Thus as a maxmin agent she chooses an optimal stopping strategy τ^* by solving

$$\max_{\tau \in \Gamma} \min_{P \in \mathcal{P}_0} E_P \left(X_\tau - c\tau \right). \tag{3.1}$$

For the setting of choosing between bets on urns, X_t takes a specific form. Bets have prizes 1 and 0, and are evaluated according to maxmin with utility index u which, without loss of generality, is normalized to satisfy

$$u(0) = 0,\ u(1) = 1.$$

Then the time t-conditional utility of betting on red (blue) from the ambiguous urn is $\min_{\mu \in \mathcal{M}_t} E\mu$ $(\min_{\mu \in \mathcal{M}_t} E^*\mu)$, where

$$E\mu \equiv \int (\tfrac{1}{2} + \theta) d\mu \text{ and } E^*\mu \equiv \int \left(\tfrac{1}{2} - \theta\right) d\mu.$$

The bet on red (or blue) from the risky urn has utility $\frac{1}{2}$.

DM can choose between betting on the draw from the risky or ambiguous urn and also on drawing red or blue. Thus, if she makes her betting choice at time t, her payoff is given by

$$X_t = \max \left\{ \min_{\mu \in \mathcal{M}_t} E\mu,\ \min_{\mu \in \mathcal{M}_t} E^*\mu,\ \tfrac{1}{2} \right\}. \tag{3.2}$$

In order to obtain closed-form solutions to the optimal stopping problem, we specialize prior beliefs about the bias and assume, for parameters $0 < \alpha < \frac{1}{2}$ and $0 < \epsilon < 1$, that

$$\mathcal{M}_0 = \{(1-m)\delta_{-\alpha} + m\delta_\alpha : \frac{1-\epsilon}{2} \leq m \leq \frac{1+\epsilon}{2}\}. \tag{3.3}$$

According to each prior, the urn is biased, (the proportion of red is either $\frac{1}{2} - \alpha$ or $\frac{1}{2} + \alpha$), but there is ambiguity about which direction for the bias is more likely. The result is that initially DM conforms to the intuitive ambiguity-averse behavior in Ellsberg's 2-urn experiment: she strictly prefers to bet on the risky urn to betting on either color from the ambiguous urn because

$$\min_{\mu \in \mathcal{M}_0} E\mu = \min_{\mu \in \mathcal{M}_0} E^*\mu = \frac{1}{2} - \epsilon\alpha < \frac{1}{2}. \tag{3.4}$$

The specification $\mathcal{M}_0$ involves the two parameters α and ϵ. We interpret ϵ as modeling ambiguity (aversion): the set $\mathcal{M}_0$ can be identified with the probability interval $\left[\frac{1-\epsilon}{2}, \frac{1+\epsilon}{2}\right]$ for the positive bias α, and this interval is larger if ϵ increases. At the extreme when $\epsilon = 0$, then $\mathcal{M}_0$ is the singleton according to which the two biases are equally likely, and DM is a Bayesian who faces uncertainty with variance α^2 about the true bias, but no ambiguity. We interpret α as measuring the degree of this prior uncertainty, or *prior variance*; ($\alpha = 0$ implies certainty that the composition of the ambiguous urn is identical to that of the risky urn). The model's other two parameters c and σ have obvious interpretations.

Bayesian updating of each prior yields the following set of posteriors (see Appendix A):

$$\mathcal{M}_t = \{(1-m)\delta_{-\alpha} + m\delta_\alpha : \underline{m}_t \leq m \leq \overline{m}_t\}, \tag{3.5}$$

where

$$\underline{m}_t = \frac{\frac{1-\epsilon}{1+\epsilon}\varphi(Z_t)}{1+\frac{1-\epsilon}{1+\epsilon}\varphi(Z_t)},\ \overline{m}_t = \frac{\frac{1+\epsilon}{1-\epsilon}\varphi(Z_t)}{1+\frac{1+\epsilon}{1-\epsilon}\varphi(Z_t)}, \tag{3.6}$$

and

$$\varphi(z) = \exp\left(2\alpha z/\sigma^2\right). \tag{3.7}$$

The probability interval $[\underline{m}_t, \overline{m}_t]$ for the positive bias α changes over time, with the response to the signal captured by the function φ. One obtains, therefore, that

$$\begin{aligned} \min_{\mu \in \mathcal{M}_t} E\mu &= (\tfrac{1}{2}+\alpha) - \frac{2\alpha}{1+\frac{1-\epsilon}{1+\epsilon}\varphi(Z_t)} \\ \min_{\mu \in \mathcal{M}_t} E^*\mu &= (\tfrac{1}{2}-\alpha) + \frac{2\alpha}{1+\frac{1+\epsilon}{1-\epsilon}\varphi(Z_t)} \end{aligned} \tag{3.8}$$

and hence,

$$X_t = X(Z_t) = \begin{cases} (\frac{1}{2}+\alpha) - \frac{2\alpha}{1+\frac{1-\epsilon}{1+\epsilon}\varphi(Z_t)} & \text{if } Z_t > \frac{\sigma^2}{2\alpha}\log(\frac{1+\epsilon}{1-\epsilon}) \\ (\frac{1}{2}-\alpha) + \frac{2\alpha}{1+\frac{1+\epsilon}{1-\epsilon}\varphi(Z_t)} & \text{if } Z_t < -\frac{\sigma^2}{2\alpha}\log(\frac{1+\epsilon}{1-\epsilon}) \\ \frac{1}{2} & \text{otherwise.} \end{cases} \tag{3.9}$$

3.2. Optimal stopping

We give an explicit solution to the optimal stopping problem (3.1), assuming (3.5), (3.9) and the construction of $\mathcal{P}_0$ detailed in §2. (Below "almost surely" qualifications should be understood, even where not stated explicitly, and as defined relative to any measure in $\mathcal{P}_0$.)

Let

$$l(r) = 2\log(\frac{r}{1-r}) - \frac{1}{r} + \frac{1}{1-r},\ r \in (0,1), \tag{3.10}$$

and define $\widehat{r}$ by

$$l(\widehat{r}) = \frac{2\alpha^3}{c\sigma^2}. \tag{3.11}$$

$\widehat{r}$ is uniquely defined thereby and $\frac{1}{2} < \widehat{r} < 1$, because $l(\cdot)$ is strictly increasing, $l(0) = -\infty$, $l(\frac{1}{2}) = 0$, and $l(1) = \infty$.

Theorem 3.1. *(i) $\tau^* = 0$ if and only if $\frac{1+\epsilon}{2} \geq \widehat{r}$, in which case $X_{\tau^*} = X_0 = \frac{1}{2}$.*
(ii) Let $\frac{1+\epsilon}{2} < \widehat{r}$. Then the optimal stopping time satisfies $\tau^* > 0$ and is given by

$$\tau^* = \min\{t \geq 0 : \ \mid Z_t \mid \geq \overline{z}\},$$

where

$$\overline{z} = \frac{\sigma^2}{2\alpha}\left[\log\frac{1+\epsilon}{1-\epsilon} + \log\frac{\overline{r}}{1-\overline{r}}\right], \tag{3.12a}$$

and $\overline{r}$, $\widehat{r} < \overline{r} < 1$, is the unique solution to the equation

$$l(r) + l(\frac{1+\epsilon}{2}) = \frac{4\alpha^3}{c\sigma^2}. \tag{3.13}$$

Moreover, on stopping either the bet on red is chosen (if $Z_{\tau^*} \geq \overline{z}$) or the bet on blue is chosen (if $Z_{\tau^*} \leq -\overline{z}$); the bet on the risky urn is never optimal at $\tau^* > 0$.

Part (i) characterizes conditions under which no learning is optimal. This case excludes the limiting Bayesian model with $\epsilon = 0$ for which some learning is necessarily optimal for all values of the remaining parameters. In fact, *it is optimal to reject learning if and only if ambiguity, as measured by ϵ, is suitably large.* Then the bet on the risky urn is chosen immediately and the opportunity to learn is declined. The cut-off value $2\widehat{r} - 1$ for ϵ is increasing in α and decreasing in c and σ. In the complementary case where some learning is chosen, (ii) shows that it is optimal to sample as long as the signal Z_t lies in the continuation interval $(-\overline{z}, \overline{z})$. When Z_t hits either endpoint, learning stops and DM bets on the ambiguous urn. Thus the *risky urn is chosen (if and) only if it is not optimal to learn.*

There is simple intuition for the noted features of the optimal strategy. First, consider the effect of ambiguity (large ϵ) on the incentive to learn. DM's prior beliefs admit only α and $-\alpha$ as the two possible values for the true bias. She will incur the cost of learning if she believes that she is likely to learn quickly which of these is true. She understands that she will come to accept α (or $-\alpha$) as being true given realization of sufficiently large positive (negative) values for Z_t. A difficulty is that she is not sure which probability law in her set $\mathcal{P}_0$ describes the signal process. As a conservative decision-maker, she bases her decisions on the worst-case scenario P^* in her set. Because she is trying to learn, the worst-case minimizes the probability of extreme, hence revealing, signal realizations, which, informally speaking, occurs if $P^*(\{dZ_t > 0\} \mid Z_t > 0)$ and $P^*(\{dZ_t < 0\} \mid Z_t < 0)$ are as small as possible. That is, if $Z_t > 0$, then the distribution of the increment dZ_t is computed using the posterior associated with that prior in $\mathcal{M}_0$

which assigns the largest probability $\frac{1+\epsilon}{2}$ to the negative bias $-\alpha$, while if $Z_t < 0$, then the distribution of the increment is computed using the posterior associated with the prior assigning the largest probability $\frac{1+\epsilon}{2}$ to the positive bias α. It follows that the prospect of learning from future signals is less attractive when viewed from the perspective of P^* the greater is ϵ. A second effect of an increase in ϵ is that it reduces the ex ante utility of betting on the ambiguous urn (3.4) and hence implies that signals in an increasingly large interval would not change betting preference. Consequently, a small sample is unlikely to be of value – only long samples are useful. Together, these two effects suggest existence of a cutoff value for ϵ beyond which no amount of learning is sufficiently attractive to justify its cost.

There remains the following question for smaller values of ϵ: why is it never optimal to try learning for a while and then, for some sample realizations, to stop and bet on the risky urn? The intuition, adapted from Fudenberg, Strack and Strzalecki (2017), is that this feature is a consequence of the specification $\mathcal{M}_0$ for the set of priors. To see why, suppose that Z_t is small for some positive t. A possible interpretation, particularly for large t, is that the true bias is small and thus that there is little to be gained by continuing to sample – DM might as well stop and bet on the risky urn. But this reasoning is excluded when, as in our specification, DM is certain that the bias is $\pm\alpha$. Then signals sufficiently near 0 must be noise and the situation is essentially the same as it was at the start. Hence, if stopping to bet on the risky urn were optimal at t, it would have been optimal also at time 0. This intuition is suggestive of the likely consequences of generalizing the specification of $\mathcal{M}_0$. Suppose, for example, that $\mathcal{M}_0$ is such that all its priors share a common finite support. We conjecture that then the predicted incompatibility of learning and betting on the risky urn would be overturned if and only if the zero bias point is in the common support.

See Appendix B for additional features of the optimal stopping strategy that are derived from the explicit solution given in the theorem.

4. A robust sequential hypothesis test

The setting and notation are slightly modified from the Ellsberg context. The major modifications are: the absence of an unambiguous urn or "outside option," and dropping the symmetric treatment of the two colors which here are the two hypotheses. We reparametrize slightly so that the parameter set is $\Theta = \{0, 1\}$ and the signal process is given by

$$Z_t = \theta\beta t + \sigma B_t, \tag{4.1}$$

with $\beta \neq 0$ and $\sigma > 0$ given. The task is to choose between the two statistical hypotheses

$$H_0 : \theta = 0 \text{ and } H_1 : \theta = 1.$$

The loss function specifies a zero loss if $d = \theta$, ($d \in \{0, 1\}$ denotes the decision adopted on stopping) and otherwise is given by

$$L(\theta = 1, d = 0) = a > 0,$$
$$L(\theta = 0, d = 1) = b > 0.$$

Initial beliefs are given by

$$\mathcal{M}_0 = \{\mu^m = (1 - m)\delta_0 + m\delta_1 : \underline{m} \leq m \leq \overline{m}\}, \tag{4.2}$$

and prior-by-prior Bayesian updating gives the time t set of posteriors $\mathcal{M}_t$,

$$\mathcal{M}_t = \{\mu_t^m = (1 - m)\delta_0 + m\delta_1 : \underline{m}_t \leq m \leq \overline{m}_t\}.$$

Here the largest and smallest posterior probabilities for H_1 are

$$\overline{m}_t = \frac{\frac{\overline{m}}{1-\overline{m}}\varphi(t, Z_t)}{1 + \frac{\overline{m}}{1-\overline{m}}\varphi(t, Z_t)}, \ \underline{m}_t = \frac{\frac{\underline{m}}{1-\underline{m}}\varphi(t, Z_t)}{1 + \frac{\underline{m}}{1-\underline{m}}\varphi(t, Z_t)},$$

where

$$\varphi(t, z) = \exp\{\frac{\beta}{\sigma^2}\left(z - \frac{\beta t}{2}\right)\}.$$

The minimum expected loss if sampling is stopped at t is

$$X_t = \min\left\{a\overline{m}_t, b\left(1 - \underline{m}_t\right)\right\}. \tag{4.3}$$

This leads to the following optimal stopping problem:

$$\min_{\tau} \max_{P \in \mathcal{P}_0} E_P\left(\tau + X_\tau\right). \tag{4.4}$$

(The set $\mathcal{P}_0$ of predictive priors is constructed as described in the previous sections.)

To describe the solution, define the increasing function $l(\cdot)$ on $(0, 1)$ by

$$l(r) = \frac{2\sigma^2}{\beta^2}[2\log(\frac{r}{1-r}) - \frac{1}{r} + \frac{1}{1-r}].$$

For perspective, consider first the special Bayesian case ($\mathcal{M}_0 = \{\mu\}$, hence $\mathcal{M}_t = \{\mu_t\}$, $\mu_t(1) = m_t$). Define $r_1^B < r_2^B$ by

$$l(r_1^B) = l(\frac{b}{a+b}) - a, \ \ l(r_2^B) = l(\frac{b}{a+b}) + b.$$

Theorem 4.1 (Peskir and Shiryaev 2006). *In the Bayesian case, it is optimal to continue at t if and only if*

$$r_1^B < m_t < r_2^B. \tag{4.5}$$

Otherwise, it is optimal to accept H_1 or H_0 according as $m_t \geq r_2^B$ or $m_t \leq r_1^B$ respectively.

The cut-off values in the general case are defined as follows. Let

$$\tilde{Z}_t = \frac{\beta t}{2} + \frac{\sigma^2}{\beta} \log\left(\frac{b-a}{2a}\frac{1-\underline{m}}{\underline{m}} + \sqrt{\frac{(b-a)^2}{4a^2}\left(\frac{1-\underline{m}}{\underline{m}}\right)^2 + \frac{b}{a}\frac{1-\overline{m}}{\overline{m}}\frac{1-\underline{m}}{\underline{m}}}\right), \tag{4.6}$$

and define $r_1 < r_2$ by

$$l(r_1) = l\left(\overline{m}_t\left(\tilde{Z}_t\right)\right) - a,\ l(r_2) = l\left(\underline{m}_t\left(\tilde{Z}_t\right)\right) + b.$$

Theorem 4.2. *In the general model, it is optimal to continue at t if and only if*

$$[\underline{m}_t, \overline{m}_t] \cap (r_1, r_2) \neq \varnothing. \tag{4.7}$$

Otherwise, it is optimal to accept H_1 or H_0 according as $\underline{m}_t \geq r_2$ or $\overline{m}_t \leq r_1$ respectively.

Compare the two results assuming $\overline{m} \neq \underline{m}$ and that the Bayesian's prior probability for H_1 lies in the interval $[\underline{m}, \overline{m}]$, (which is what we mean in §1 and below by a "compatible Bayesian"). Then it is easy to see from $r_1 < r_1^B < r_2^B < r_2$ that the desire for robustness leads to the optimality of longer sampling. Note that the *robustness-seeking DM may choose to continue sampling even given a realized sample at which ALL compatible Bayesians would choose to stop* (this occurs if $r_1 < \overline{m}_t < r_1^B$ or $r_2^B < \underline{m}_t < r_2$). In this sense, focussing on multiple Bayesian agents alone understates the value of sampling.

Proof of the theorem is similar to that of Theorem 3.1 and is available upon request. Note that the stopping conditions could be expressed alternatively using the signal process Z_t as in Theorem 3.1 which would make the similarity between the two theorems even more apparent.

Remark 2. *Time-consistency in the present context is closely related to the Stopping Rule Principle – that the stopping rule should have no effect on what is inferred from observed data and hence on the decision taken after stopping (Berger 1985). It is well-known that: (i) conventional frequentist methods, based on ex ante fixed sample size significance levels, violate this Principle and permit the analyst to sample to a foregone conclusion when data-dependent stopping rules are permitted; and (ii) Bayesian posterior odds analysis satisfies the Principle. Kadane, Schervish and Seidenfeld (1996)* *point to the law of iterated expectations as responsible for excluding foregone conclusions (if the prior is countably additive). Equation (2.5) is a nonlinear counterpart that we suspect plays a similar role in our model (though details are beyond the scope of this paper).*

A. Appendix: Proof Of Theorem 3.1

First, note that the formula (3.5) describing posteriors follows from Liptser and Shiryaev (1977, Theorem 9.1): given any $\mu = (1-m)\delta_{\theta_1} + m\delta_{\theta_2}$, then $\pi_t \equiv \mu_t(\theta_2)$ satisfies $\pi_0 = m$ and

$$d\pi_t \;=\; \tfrac{\theta_2-\theta_1}{\sigma^2}\pi_t(1-\pi_t)[dZ_t - (\theta_1(1-\pi_t)+\theta_2\pi_t)dt].$$

The solution is

$$\pi_t = \frac{m}{1-m}\varphi(t,Z_t) \Big/ \left[1 + \frac{m}{1-m}\varphi(t,Z_t)\right],$$

where

$$\varphi(t,z) = \exp\{\frac{\theta_2-\theta_1}{\sigma^2}z - \frac{1}{2\sigma^2}(\theta_2^2-\theta_1^2)t\}.$$

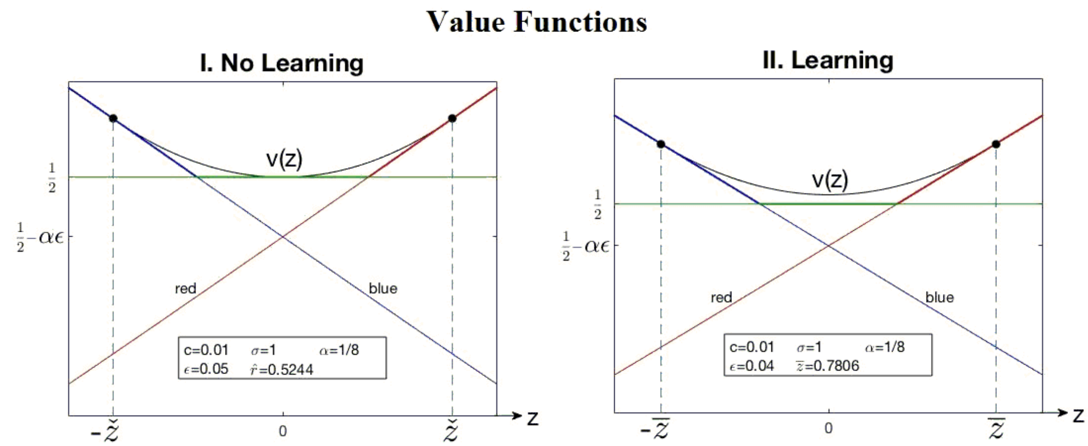
In particular, for the two extreme measures, $\mu = \overline{\mu}, \underline{\mu}$, satisfying, for all t,

$$\overline{\mu}_t(\alpha) = \overline{m}_t \text{ and } \underline{\mu}_t(\alpha) = \underline{m}_t,$$

one obtains the corresponding parameter estimates

$$\begin{aligned} \widehat{\theta}_t^{\overline{\mu}}(Z_t) &= \alpha - \frac{2\alpha}{1+\frac{1+\epsilon}{1-\epsilon}\varphi(Z_t)}, \text{ and} \\ \hat{\theta}_t^{\underline{\mu}}(Z_t) &= \alpha - \frac{2\alpha}{1+\frac{1-\epsilon}{1+\epsilon}\varphi(Z_t)}. \end{aligned} \tag{A.1}$$

Before proceeding to the formal proof, consider the Figure which illustrates both the theorem and elements in the proof below. The red (blue) curve represents the (minimum expected) payoff to a bet on red (blue) conditional on each signal z. The payoff, in green, to the bet on the risky urn equals $1/2$ for every z. The upper envelope of these three curves is the graph of $X(z)$, the maximum payoff possible if sampling ceases and a bet is chosen given z. Because v, the value function for the optimal stopping problem, includes the option of waiting longer before choosing between bets, it lies everywhere weakly above X, and coincides with X at values of z where further sampling is not optimal. Since $z=0$ at time 0, the (earliest) stopping time occurs at the smallest $\mid z \mid$ where v and X coincide, which occurs at 0 in I corresponding to part (i) of the theorem ($\check{z}$ is defined in the proof of (i) below), and at $\pm\overline{z}$ in II corresponding to part (ii) of the theorem. Note that at stopping points there is a smooth fit between v and X as is



common in the free-boundary approach to analysing optimal stopping problems (Peskir and Shiryaev 2006)). In the zero-learning case portrayed, the contact between v and the horizontal line at $\frac{1}{2}$ extends for a small interval (denoted $[-\widehat{z}, \widehat{z}]$ in the proof) about 0. The two figures share the parameter values $(c, \sigma, \alpha) = \left(.01, 1, \frac{1}{8}\right)$. They differ only in the value of ϵ, (.04 versus .05), which difference is significant because the no-learning cut-off value for ϵ is $2\widehat{r} - 1 = .0488$. Accordingly, there is no sampling when $\epsilon = .05$ and the expected optimal sample size equals .61 if $\epsilon = .04$ and the true bias is zero (by (B.1)).

Proof of Theorem 3.1: For both (i) and (ii), the strategy is to: (a) guess the P^* in $\mathcal{P}_0$ that is the worst-case scenario; (b) solve the classical optimal stopping problem given the single prior P^*; (c) show that the value function derived in (b) is also the value function for our problem (3.1); and (d) use the value function to derive τ^*. The process is aided by intuition derived from analysis of the modified optimal stopping problem where the bets on stopping are on a single fixed color, say red, and the choice is only between urns. Analysis of this problem is simpler because it is apparent that the worst-case, at every time and sample, corresponds to the measure in $\mathcal{M}_0$ that assigns the lowest (prior and posterior) probability to the bias towards red. (In our problem, in contrast, the identity of the worst-case prior varies with the sample.) Solution of the single-color problems for red and then blue, gives value functions g_1 and g_2 respectively, which, in turn, appear in the expressions (A.4) and (A.12) for the value functions for step (c).

Define g_1 and g_2 by, for $0 < y < 1$,

$$\begin{aligned} g_1(y; C_1, C_2) &= \frac{c\sigma^2}{2\alpha^2}(2y-1)\log(\frac{y}{1-y}) - C_1(y - \frac{1}{2}) + C_2, \\ g_2(y; C_1, C_2) &= \frac{c\sigma^2}{2\alpha^2}(2y-1)\log(\frac{y}{1-y}) + C_1(y - \frac{1}{2}) + C_2, \end{aligned} \tag{A.2}$$

where the constants C_1 and C_2 will be determined by the smooth-contact conditions discussed in connection with the Figure. In particular, they will differ between parts (i) and (ii).

Let P^* be the probability measure in $\mathcal{P}_0$ which has density generator process (η_t),

$$-\eta_t = (\hat{\theta}_t^{\overline{\mu}}/\sigma)\mathbf{1}_{Z_t \leq 0} + (\hat{\theta}_t^{\underline{\mu}}/\sigma)\mathbf{1}_{Z_t > 0}.$$

It will be shown that P^* is the worst-case scenario in $\mathcal{P}_0$.

Proof of (ii): Consider the classical optimal stopping problem under the measure P^*,

$$\max_{\tau} E_{P^*}[X(Z_\tau) - c\tau], \tag{A.3}$$

where $X(\cdot)$ is defined in (3.9). To describe the value function v for this problem, define

$$v(z) = \begin{cases} \frac{1}{2} - \alpha + \frac{2\alpha}{1+\frac{1+\epsilon}{1-\epsilon}\varphi(z)} & \text{if } z < -\overline{z} \\ g_1(1 - \frac{1}{1+\frac{1+\epsilon}{1-\epsilon}\varphi(z)}; C_1, C_2) & \text{if } -\overline{z} \leq z < 0 \\ g_2(1 - \frac{1}{1+\frac{1-\epsilon}{1+\epsilon}\varphi(z)}; C_1, C_2) & \text{if } 0 \leq z < \overline{z} \\ \frac{1}{2} + \alpha - \frac{2\alpha}{1+\frac{1-\epsilon}{1+\epsilon}\varphi(z)} & \text{if } \overline{z} \leq z, \end{cases} \tag{A.4}$$

where

$$C_1 = 2\alpha - \frac{c\sigma^2}{2\alpha^2} l(\bar{r}), \;\; C_2 = \frac{1}{2} + \frac{c\sigma^2}{4\alpha^2}\frac{(2\bar{r}-1)^2}{\bar{r}(1-\bar{r})}.$$

Lemma A.1. *v is the value function of the classical optimal stopping problem (A.3), i.e., for any $t \geq 0$,*

$$v(z) = \max_{\tau \geq t} E_{P^*}[X(Z_\tau) - c(\tau - t) \mid Z_t = z].$$

Further, v satisfies the following HJB equation

$$\max\{X(z) - v(z), -c + \frac{1}{2}\sigma^2 v_{zz}(z) + f(z, -sgn(z)\epsilon) v_z(z)\} = 0, \tag{A.5}$$

where $sgn(z) = 1$ if $z \geq 0$, $= -1$ if $z < 0$, and

$$f(z,p) = \alpha - \frac{2\alpha}{1 + \frac{1+p}{1-p}\varphi(z)}. \tag{A.6}$$

Finally, v also satisfies

$$sgn(v_z(z)) = sgn(z), \text{ and} \tag{A.7}$$

$$-c + \frac{1}{2}\sigma^2 v_{zz}(z) + f(z, -sgn(z)\epsilon)v_z(z) = 0 \quad \forall z \in (-\overline{z}, \overline{z}), \tag{A.8}$$

For the proof, first verify that v defined in (A.4) satisfies the HJB equation (A.5), and then apply El Karoui et al. (1997, Theorems 8.5, 8.6). Alternatively, a proof can be constructed along the lines of Peskir and Shiryaev (2006, Ch. 6).

Next we prove that, for $t \geq 0$,

$$v(z) = \max_{\tau \geq t} \min_{P \in \mathcal{P}_0} E_P[X(Z_\tau) - c(\tau - t) \mid Z_t = z],$$

that is, v is the value function of our optimal stopping problem (3.1). Since $v(z)$ is time invariant, we prove only the case $t = 0$.

By Lemma A.1,

$$v(z) = \max_\tau E_{P^*}[X(Z_\tau) - c\tau] \geq \max_\tau \min_{P \in \mathcal{P}_0} E_P[X(Z_\tau) - c\tau].$$

To prove the opposite inequality, consider the stopping time

$$\tau^* = \inf\{t \geq 0 : \mid Z_t \mid \geq \overline{z}\}.$$

For $t \leq \tau^*$, by Ito's formula, (A.5), (A.7) and (A.8),

$$\begin{aligned} dv(Z_t) &= \frac{1}{2}\sigma^2 v_{zz}(Z_t)dt + v_z(Z_t)dZ_t \\ &= [c - f(Z_t, -\mathrm{sgn}(Z_t)\epsilon)v_z(Z_t)]dt + v_z(Z_t)dZ_t \\ &= [c - f(Z_t, -\mathrm{sgn}(v_z(Z_t))\epsilon)v_z(Z_t)]dt + v_z(Z_t)dZ_t. \end{aligned} \tag{A.9}$$

Each $P = P^\zeta \in \mathcal{P}_0$ corresponds to a density generator process $f(t, Z_t, \zeta_t)$ where (ζ_t) is a $\{\mathcal{G}_t\}$-adapted process taking values in $[-\epsilon, \epsilon]$. Set

$$W_t^\zeta = \frac{1}{\sigma} Z_t + \frac{1}{\sigma}\int_0^t f(Z_s, \zeta_s)ds.$$

Then (W_t^ζ) is a Brownian motion under P^ζ and

$$\begin{aligned} & dv(Z_t) \\ = \; & [c + (f(Z_t, \zeta_t) - f(Z_t, -\mathrm{sgn}(v_z(Z_t))\epsilon))\, v_z(Z_t)]dt + \sigma v_z(Z_t)dW_t^\zeta. \end{aligned} \tag{A.10}$$

Because $f(z,p)$ is increasing in p,

$$(f(Z_t,\zeta_t) - f(Z_t, -\mathrm{sgn}(v_z(Z_t))\epsilon))\, v_z(Z_t) \geq 0.$$

Taking expectation in (A.10) under P^{ζ}, we have

$$\begin{aligned} v(z) &\leq E_{P^{\zeta}}[v(Z_{\tau^*}) - c\tau^*] \\ &= E_{P^{\zeta}}[X(Z_{\tau^*}) - c\tau^*]. \end{aligned}$$

The above inequality is due to

$$E_{P^{\zeta}}[\int_0^{\tau^*} \sigma v_z(Z_t) dW_t^{\zeta}] = 0,$$

which is guaranteed by

$$\max_{Q\in\mathcal{P}_0} E_Q[\tau^*] < \infty; \tag{A.11}$$

see Peskir and Shiryaev (2006, Theorem 21.1) for the classical case. In our setting, (A.11) is implied by the boundedness of X_t because:

$$\begin{aligned} -\infty &< \max_{\tau\in\Gamma} \min_{P\in\mathcal{P}_0} E_P\left(X_\tau - c\tau\right) = \max_{\tau\in\Gamma}[-\max_{P\in\mathcal{P}_0} E_P\left(c\tau - X_\tau\right)] \\ &\leq \max_{\tau\in\Gamma}[\max_{P\in\mathcal{P}_0} E_P\left(X_\tau\right) - \max_{P\in\mathcal{P}_0} E_P\left(c\tau\right)] \Longrightarrow \max_{Q\in\mathcal{P}_0} E_Q[\tau^*] < \infty. \end{aligned}$$

Finally, because P^{ζ} can be any measure in $\mathcal{P}_0$, deduce that

$$\begin{aligned} v(z) &\leq \min_{P\in\mathcal{P}_0} E_P[X(Z_{\tau^*}) - c\tau^*] \\ &\leq \max_{\tau} \min_{P\in\mathcal{P}_0} E_P[X(Z_\tau) - c\tau]. \end{aligned}$$

Conclude that v is the value function for our optimal stopping problem and that τ^* is the optimal stopping time. Note that the time 0 signal $Z_0 = 0$ falls in the continuation region.

To complete the proof of statement (ii), let $\overline{\overline{z}}$ be given by

$$\overline{\overline{z}} = \frac{\sigma^2}{2\alpha}\log(\frac{1+\epsilon}{1-\epsilon}) < \overline{z}.$$

It follows from (3.2) and (3.5) that at any given t, not necessarily an optimal stopping time, betting on the ambiguous urn is preferred to betting on the risky urn iff $\mid Z_t \mid \geq \overline{\overline{z}}$. Thus at $\tau^* > 0$,

$$\mid Z_{\tau^*} \mid = \overline{z} > \overline{\overline{z}},$$

and betting on the ambiguous urn is optimal on stopping.

Remark 3. *The preceding implies that P^* is indeed the minimizing measure because the minimax property is satisfied:*

$$\begin{aligned}
\max_{\tau} E_{P^*} X(Z_\tau) &= \max_{\tau} \min_{P\in\mathcal{P}_0} E_P X(Z_\tau) \leq \\
\min_{P\in\mathcal{P}_0} \max_{\tau} E_P X(Z_\tau) &\leq \max_{\tau} E_{P^*} X(Z_\tau) \Longrightarrow \\
\min_{P\in\mathcal{P}_0} \max_{\tau} E_P X(Z_\tau) &= \max_{\tau} \min_{P\in\mathcal{P}_0} E_P X(Z_\tau).
\end{aligned}$$

Proof of (i): The proof is similar to that of part (ii). The only difference is that, as illustrated in panel I of the Figure, there is contact between the value function and X in an interval surrounding 0. This leads to the new constants

$$C_3 = \alpha,\ \ C_4 = \frac{1}{2} + \frac{c\sigma^2}{2\alpha^2}\left(\frac{1}{2\hat{r}(1-\hat{r})} - 2\right),$$

and to the value function v given by

$$v(z) = \begin{cases} \frac{1}{2} - \alpha + \frac{2\alpha}{1+\frac{1+\epsilon}{1-\epsilon}\varphi(z)} & \text{if } z \leq -\check{z} \\ g_3(1 - \frac{1}{1+\frac{1+\epsilon}{1-\epsilon}\varphi(z)}; C_3, C_4) & \text{if } -\check{z} < z < -\widehat{z} \\ \frac{1}{2} & -\widehat{z} \leq z \leq \widehat{z} \\ g_4(1 - \frac{1}{1+\frac{1-\epsilon}{1+\epsilon}\varphi(z)}; C_3, C_4) & \text{if } \widehat{z} < z < \check{z} \\ \frac{1}{2} + \alpha - \frac{2\alpha}{1+\frac{1-\epsilon}{1+\epsilon}\varphi(z)} & \text{if } \check{z} \leq z, \end{cases} \tag{A.12}$$

where

$$\begin{aligned}
\check{z} &= \frac{\sigma^2}{2\alpha}\left[\log\frac{1+\epsilon}{1-\epsilon} + \log\frac{\widehat{r}}{1-\widehat{r}}\right], \\
\widehat{z} &= \frac{\sigma^2}{2\alpha}\left[\log\frac{1+\epsilon}{1-\epsilon} + \log\frac{1-\widehat{r}}{\widehat{r}}\right].
\end{aligned}$$

The continuation region for this case is $(-\check{z}, -\widehat{z}) \cup (\widehat{z}, \check{z})$. Note that $\frac{1+\epsilon}{2} \geq \widehat{r}$ is equivalent to $\bar{r} \leq \frac{1+\epsilon}{2}$ which is also equivalent to $\widehat{z} \geq 0$. Thus $\frac{1+\epsilon}{2} \geq \widehat{r}$ implies that

$$-\check{z} \leq -\widehat{z} < \widehat{z} \leq \check{z}.$$

The significance of the interval $[-\widehat{z}, \widehat{z}]$ is that DM should stop and bet on risky urn when Z_τ first enters the interval. In our context, this occurs at time 0 because $Z_0 = 0$. ■

B. Appendix: More details on the Ellsberg urns

A corollary of Theorem 3.1 elaborates on the properties of the optimal stopping time. Given two stopping strategies τ_1 and τ_2, say that τ_1 *stops later* if

$$\{\omega \in \Omega : \tau_1(\omega) \leq t\} \subset \{\omega \in \Omega : \tau_2(\omega) \leq t\}, \text{ for every } t.$$

If both strategies have the form in the theorem with critical values $\overline{z}_1$ and $\overline{z}_2$ respectively, then the preceding is equivalent to $\overline{z}_1 \geq \overline{z}_2$. Denote by P^θ the probability distribution of (Z_t) if θ is the true bias.

Corollary B.1. *(a) DM stops sampling later in each of the following cases:*
(a.1) c falls. (a.2) ϵ increases in the interval $[0, 2\widehat{r} - 1)$, where $\widehat{r}$ is defined in (3.11).
(a.3) σ and α both increase in such a way that α/σ^2 is constant.
(b) $P^\theta(\tau^* < \infty) = 1$ for every θ.
(c) For each θ, the mean delay time according to P^θ is finite and given by

$$E^\theta \tau^* = \begin{cases} (\overline{z}/\sigma)^2 \left[\frac{\tanh(\theta\overline{z}/\sigma^2)}{\theta\overline{z}/\sigma^2}\right] & \text{if } \theta \neq 0 \\ (\overline{z}/\sigma)^2 & \text{if } \theta = 0. \end{cases} \tag{B.1}$$

(d) For each $\theta \neq 0$,

$$P^\theta(\{\theta Z_{\tau^*} > 0\}) = \frac{1}{1 + \exp\left(-\frac{2|\theta|}{\sigma^2}\overline{z}\right)}.$$

Proof. (a.1) $\ell(\cdot)$ increasing implies that $\widehat{r}$ is decreasing in c. There exists $\widehat{c}$ such that $c < \widehat{c}$ iff $\epsilon < 2\widehat{r} - 1 \Longrightarrow l(\frac{1+\epsilon}{2}) < \frac{4\alpha^3}{c\sigma^2}$, which implies that both r and $\overline{z}$ increase as c falls. For $c \geq \widehat{c}$, part (i) of the theorem gives $\tau^* = 0$.
(a.2) $\overline{z}$ is increasing in ϵ: $\ell'(r) = \frac{1}{r^2(1-r)^2}$, $\frac{d\overline{z}}{d\epsilon} > 0$ iff $\frac{2r_2}{1-\epsilon}\ell'(r) > \frac{1+\epsilon}{1-r_2}\frac{1}{2}\ell'\left(\frac{1+\epsilon}{2}\right)$ iff $(1-\epsilon)(1+\epsilon) > \overline{r}(1-\overline{r})$. But $\frac{1}{2} < \frac{1+\epsilon}{2} < \widehat{r} < \overline{r} \Longrightarrow \frac{1+\epsilon}{2} \cdot \frac{1-\epsilon}{2} > \widehat{r}(1-\widehat{r}) > \overline{r}(1-\overline{r}) \Longrightarrow (1-\epsilon)(1+\epsilon) > 4\overline{r}(1-\overline{r}) > \overline{r}(1-\overline{r})$
(a.3) If $\ell\left(\frac{1+\epsilon}{2}\right) \geq \frac{4\alpha^3}{c\sigma^2}$, then $\frac{1+\epsilon}{2} \geq \widehat{r}$ and $\tau^* = 0$. Next restrict attention to parameter values satisfying $\ell\left(\frac{1+\epsilon}{2}\right) < \frac{4\alpha^3}{c\sigma^2}$ and consider an increase in α with α/σ^2 held constant. In this region, $\overline{r} > \frac{1}{2}$ and $\overline{z}$ is an increasing function of $\overline{r}$, which in turn is an increasing function of α^2, hence of α.

(b) and (c) follow from well-known results regarding hitting times of Brownian motion with drift (see Borodin and Salminen (2015), for example). Here the question concerns the distribution of the time at which Z_t first hits $\pm\overline{z}$, assuming case (ii) of the theorem where some sampling is optimal. To prove (d), apply the optional stopping theorem to the martingale $e^{-2\theta Z_t/\sigma^2}$. ■

That lower cost leads to longer sampling, as in (a.1), is not surprising. (a.2) is more interesting. When ϵ is in the indicated interval, part (ii) of the theorem applies and $\tau^* > 0$. In that case, greater ambiguity leads to a worst-case scenario that renders the signal structure less informative, hence requiring a longer sample for learning enough to improve the choice between bets. But eventually, when ϵ reaches $2\widehat{r} - 1$, the sample size needed to learn is too long to justify the cost, and the response time drops to zero.

Consider (a.3). The separate comparative static effects of σ and α are indeterminate. For example, an increase in σ has two opposite effects. A larger signal variance implies a smaller response (in absolute value) to any realized signal when updating (recall (3.7)). Therefore, any given impact on beliefs requires a stronger signal, hence also a larger sample. However, looking forward, a larger signal variance implies that less can be gained from future learning, which argues for a smaller sample. The net effect is indeterminate without further assumptions. Similarly for the effects of α, though the directions of each of the noted effects are reversed. However, when both parameters change in such a way that the ratio α/σ^2 is constant, then only the second forward-looking effect (of an increase in α) applies and DM decides later.

Interpretation of (b) and (c) is clear. For (d), note that $\theta Z_{\tau^*} > 0$ if and only if the bet on red (blue) is chosen on stopping if $\theta > 0$ ($\theta < 0$). Thus (d) gives the probability, if $\theta \neq 0$ is the true bias, of choosing the correct bet on stopping. That probability increases with ϵ and $\mid \theta \mid$, and declines with c and σ.

The proof of Theorem 3.1 yields a closed-form expression for the value function associated with the optimal stopping problem. In particular, the value at time 0 satisfies (from (A.4) and (A.12)),

$$v_0 - \tfrac{1}{2} = \begin{cases} 0 & \text{if } \frac{1+\epsilon}{2} \geq \widehat{r} \\ \frac{c\sigma^2}{2\alpha^2}\left[\frac{1}{2\overline{r}(1-\overline{r})} - \frac{2}{(1+\epsilon)(1-\epsilon)}\right] & \text{if } \frac{1+\epsilon}{2} < \widehat{r}. \end{cases} \tag{B.2}$$

Since the payoff $\frac{1}{2}$ is the best available without learning, $v_0 - \frac{1}{2}$ is the value of the learning option. In the region $\frac{1+\epsilon}{2} < \widehat{r}$, its value declines with ϵ; and it equals zero when $\frac{1+\epsilon}{2} \geq \widehat{r}$. In contrast, it can be shown that the value of learning is increasing in the prior variance α.

References

[1] J.O. Berger, The robust Bayesian viewpoint, pp. 63-124 in *Robustness of Bayesian Analyses*, J.B. Kadane ed. Elsevier, 1984.

[2] J.O. Berger, *Statistical Decision Theory and Bayesian Analysis.* Springer, 1985.

[3] J.O. Berger, An overview of robust Bayesian analysis (with discussion), *Test* 3 (1994), 5-124.

[4] A. Borodin and P. Salminen, *Handbook of Brownian Motion–Facts and Formulae*, 2nd ed. Birkhauser, 2015.

[5] Z. Chen and L.G. Epstein, Ambiguity, risk and asset returns in continuous time, *Econometrica* 70 (2002), 1403-1443.

[6] X. Cheng and F. Reidel, Optimal stopping under ambiguity in continuous time, *Math. Finan. Econ.* 7 (2013), 29-68.

[7] H. Chernoff, Sequential tests for the mean of a normal distribution, pp. 79-91 in *Proc. Fourth Berkeley Symp. on Math. Statist. and Probab.* Vol 1. U. Cal. Press, 1961.

[8] H. Choi, Learning under ambiguity: portfolio choice and asset returns, 2016.

[9] S. Ebert and P. Strack, Never, ever getting started: on prospect theory without commitment, 2017.

[10] N. El Karoui, C. Kapoundijian, E. Pardoux, S. Peng and M. Quenez, Reflected solutions of backward SDE's and related obstacle problems for PDE's, *Ann. Probab.* 25 (1997), 702-737.

[11] D. Ellsberg, Risk, ambiguity, and the Savage axioms, *Quart. J. Econ.* 75 (1961), 643-669.

[12] L.G. Epstein and M. Schneider, Recursive multiple-priors, *J. Econ. Theory* 113 (2003), 1-31.

[13] L.G. Epstein and M. Schneider, Learning under ambiguity, *Rev. Econ. Stud.* 74 (2007), 1275-1303.

[14] L.G. Epstein and M. Schneider, Ambiguity, information quality and asset pricing. *J. Finance* 63 (2008), 197-228.

[15] L.G. Epstein and M. Schneider, Ambiguity and asset markets, *Ann. Rev. Finan. Econ.* 2 (2010), 315-346.

[16] D. Fudenberg, P. Strack and T. Strzalecki, Speed accuracy and the optimal timing of choices, 2017.

[17] I. Gilboa, *Theory of Decision under Uncertainty.* Cambridge U. Press 2009.

[18] I. Gilboa, Rationality and the Bayesian paradigm: an expository note, *J. Econ. Method.* 22 (2015), 312-334.

[19] I. Gilboa, A. Postlewaite and D. Schmeidler, Rationality of belief or: why Savage's axioms are neither necessary nor sufficient for rationality, *Synthese* 187 (2012), 11-31.

[20] I. Gilboa and D. Schmeidler, Maxmin expected utility with non-unique prior, *J. Math. Econ.* 18 (1989), 141-153.

[21] P.J. Huber, A robust version of the probability ratio test, *Ann. Math. Stat.* 36 (1965), 1753-1758.

[22] J.B. Kadane, M.J. Schervish and T. Seidenfeld, Reasoning to a foregone conclusion, *JASA* 91 (1996), 1228-1235.

[23] R. Liptser and A. Shiryaev, *Statistics of Random Processes I: General Theory.* Berlin: Springer, 1977.

[24] M. Marinacci, Learning from ambiguous urns, *Statistical Papers* 43 (2002), 143-51.

[25] J. Miao, Ambiguity, risk and portfolio choice under incomplete information, *Ann. Econ. Finan.* 10 (2009), 257-279.

[26] B. Oksendal, *Stochastic Differential Equations,* 6th ed. Springer, 2005.

[27] G. Peskir, Nonlinear optimal stopping and nonlinear optimal control, presented at *Optimal Stopping in Complex Environments*, Bielefeld 2017.

[28] G. Peskir and A. Shiryaev, *Optimal Stopping and Free-Boundary Problems*. Springer, 2006.

[29] D. Rios-Insua and F. Ruggeri, *Robust Bayesian Analysis*. Springer, 2000.

[30] A. Shapiro, Rectangular sets of probability measures, *Oper. Res.* 64 (2016), 528-541.

[31] S.T. Trautman and R.J. Zeckhauser, Shunning uncertainty: the neglect of learning opportunities, *Games Econ. Behav.* 79 (2013), 44-55.

[32] A. Wald, *Sequential Analysis*. Wiley, 1947.

[33] P. Walley, *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, 1991.