Optimal Learning and Ellsberg's Urns*

Larry G. Epstein   Shaolin Ji

July 28, 2017

Abstract

We consider the dynamics of learning under ambiguity when learning is costly and is chosen optimally. The setting is Ellsberg's two-urn thought experiment modified by allowing the agent to postpone her choice between bets so that she can learn about the composition of the ambiguous urn. Signals are modeled by a diffusion process whose drift is equal to the true bias of the ambiguous urn and they are observed at a constant cost per unit time. The resulting optimal stopping problem is solved and the effect of ambiguity on the extent of learning is determined. It is shown that rejection of learning opportunities can be optimal for an ambiguity averse agent even given a small cost.

Key words: ambiguity, learning, partial information, optimal stopping, drift-diffusion model

*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

1.1. Objectives and outline

This paper considers the connection between learning and ambiguity. A common view is that ambiguity is a short-run phenomenon that would fade away given the opportunity for learning. For example, consider the metaphorical example of Ellsberg's 2-urn 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, and no information is provided about the proportion of red balls in the "unknown" or ambiguous urn. A decision-maker (DM) must choose between betting on the color drawn from the risky urn or from the ambiguous urn. It is intuitive that if she can first sample with replacement repeatedly, then as her sample increases she will become increasingly confident in her assessment of the urn's composition, and in the limit, she will behave as if she knew the true composition of the urn. Marinacci (2002) provides formal confirmation of this intuition under the assumption that DM's prior beliefs are well-specified in the sense that the true probability law lies in their "support," and Marinacci and Massari (2016) show that, under suitable assumptions, ambiguity fades away even if prior beliefs are misspecified.

Both the noted intuition and all related formal studies of which we are aware rely on the assumption that sampling is costless and exogenous. Here we reconsider the connection between learning and ambiguity when learning is costly and is chosen optimally to balance its cost with the benefit of improved decision-making. One would expect that a positive cost would limit the extent of learning chosen optimally. However, intuition regarding other questions – for instance, the impact of prior ambiguity on the extent of learning – is not clear-cut. On the one hand, there is an incentive to learn in order to reduce ambiguity; and on the other hand, ambiguity may reduce how much can be learned because it limits the sharpness of inferences that can be drawn from signals. Indeed, if the second effect dominates, then could DM find it optimal to reject any sampling, no matter how short? If so, then the presumption that ambiguity would fade away, or even diminish, in the presence of learning opportunities must be reexamined and qualified.

To address the preceding and related questions, we modify Ellsberg's 2-urn thought experiment by allowing DM to postpone her choice between bets so that she can learn about the composition of the ambiguous urn by observing realizations of a signal. Signals are modeled by a diffusion process Z_t whose drift is equal to the true bias towards red. The longer the interval over which Z_t is observed the larger and more informative is her sample, but there is a per-unit-time cost $c > 0$ to sampling; the cost can be material or cognitive. Thus the choice of how much to learn is the solution to an optimal stopping problem. We assume that DM: (i) knows the distribution, conditional on the bias, of future signals (which is the situation most favorable for learning); (ii) deals with ambiguity by solving a maxmin problem (Gilboa and Schmeidler 1989); and (iii) is forward-looking and solves her problem by backward induction (as in the continuous-time version of maxmin in Chen and Epstein (2002)). Under these assumptions and 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 latter depends positively on the cost c but can be "moderate" even if c is small depending on other parameters of preference. (See Appendix B for a numerical illustration of this point.)

Two further contributions merit explicit mention. The first concerns our use of the drift diffusion model (DDM). As stated by Milosavljevic et al (2010, p. 437), "the drift diffusion model is one of the cornerstones of modern psychology ... and, increasingly, of behavioral neuroscience." This paper is a first step in introducing the DDM into the literature on choice under ambiguity, specifically in providing a model that has a foundation in optimal learning theory. An important component in doing so is the demonstration (Appendix A) that the Chen and Epstein (2002) continuous-time model of preferences under ambiguity can be extended to accommodate learning and partial information.

The second (potential) contribution is to robustify the classical Bayesian approach to sequential testing of two simple hypotheses about the unknown drift of a Wiener process (Wald 1947, Ch. 6 of Peskir and Shiryaev 2006). The hypothesis testing problem is isomorphic to the variation of this paper's model in which the risky urn is removed and DM chooses, after stopping, between betting on a red or blue draw from the ambiguous urn; and the analysis herein is readily adapted.

Related literature is discussed next. The main body of the paper is contained in Section 2 where technical details are minimized. Appendix A provides a formal description of the model including the relevant technicalities of continuous-time stochastic calculus. Appendix B proves our main result (Theorem 2.1) about the solution of the optimal stopping problem.

1.2. Related literature

The theoretical literature on learning under ambiguity is sparse and limited to passive learning; for example, see Epstein and Schneider (2007,8) and the papers cited in the opening paragraph for discrete-time models, and see Miao (2009) and Choi (2016) for continuous-time models (Appendix A contains more on this literature). One way to differentiate from this paper is that, roughly speaking, the existing literature focuses on the question *"how does exogenous learning affect ambiguity?"* while we shift the focus to the question *"how does ambiguity affect the choice of how much to learn?"*

There is some (mixed) evidence, derived from urns-based experiments, regarding the effects of learning on ambiguity aversion. Abdellaoui et al (2016) examine empirically how the degree of ambiguity aversion varies with the number of signal realizations the DM is permitted to observe prior to choice. They report that with a smaller sample, "people exhibit more ambiguity aversion for likely events, but are more ambiguity seeking for unlikely events. Moreover, though ambiguity attitude becomes less pronounced as sample size increases, it does not vanish." Our model assumes maxmin behavior and thus precludes ambiguity seeking. Nicholls, Romm and Zimper (2015) report that giving subjects statistical information (they are told the results of previous draws) does not reduce violations of Savage's sure-thing-principle relative to that of a control group. Trautman and Zeckhauser (2013) report that their subjects neglected opportunities to learn about an ambiguous urn even at no visible cost. As noted, when learning opportunities are described as in this paper, and when there is a cognitive cost to learning, even arbitrarily small, then seeming neglect can be fully rational.

Our interpretation thus far is that the learning process is observable to the modeler. An alternative is to view the model as describing DM's unobservable private information (Lu 2016), thought process or deliberations, (with c being a cognitive cost), which underlie the time delay before choosing between bets. Decision or response times have been used to distinguish between "intuitive" and "deliberative or reasoned" choices (for example, see Rubinstein (2007, 2013, 2016) and the references therein). Since we model a sophisticated forward-looking agent, interpreting quick response times as reflecting an intuitive decision-maker would require a strictly "as if" view of the model. In addition, while our model predicts that ambiguity averse behavior declines with response time, Butler and Guiso (2013, 2014) argue that intuitive choice leads to less ambiguity aversion. Rubinstein (2013) finds only a weak connection between response time and Ellsberg-type behavior.

Lastly in this introduction we mention Fudenberg, Strack and Strzalecki (2017), which triggered our interest in the present topic. They consider an optimal stopping problem under risk where a Bayesian DM decides how long to sample before making a binary decision, and they focus particularly on the correlation between speed and accuracy. There is some overlap in intuition as indicated below. However, our focus on ambiguity and differing specifications for prior beliefs lead to a substantially different analysis.

2. The model

2.1. The framework

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.)

Before choosing between bets, DM is given the opportunity to postpone her choice so that she can learn about θ by observing realizations of a signal process $Z = (Z_t)$ given by

$$Z_t = \int_0^t \theta ds + \int_0^t \sigma dB_s = \theta t + \sigma B_t. \tag{2.1}$$

Here $\sigma > 0$ and $B = (B_t)$ is a standard Brownian motion. The underlying state space is Ω. Because DM observes only realizations of Z_t, the information available through time is represented by the filtration generated by Z, denoted $\{\mathcal{G}_t\}$.[1]

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. A stopping strategy 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. The set of stopping strategies is Γ. DM is forward-looking and has time 0 beliefs about future signals

[1] Unless specified otherwise, all processes below are taken to be $\{\mathcal{G}_t\}$-adapted even where not stated explicitly. Abbreviate $\cup_{t\geq 0}\mathcal{G}_t$ by $\mathcal{G}_\infty$.

given by the set $\mathcal{P}_0 \subset \Delta(\Omega, \mathcal{G}_\infty)$. Thus as a maxmin agent she chooses an optimal stopping strategy τ^* by solving

$$\max_{\tau \in \Gamma} \min_{P \in \mathcal{P}_0} \left(E_P X_\tau - c\tau \right) \ . \tag{2.2}$$

We proceed to describe $\mathcal{P}_0$ and X_t in greater detail.

Initial beliefs about θ are given by the set of priors $\mathcal{M}_0 \subset \Delta(\Theta)$. Assume knowledge of the signal structure. Then prior-by-prior Bayesian updating leads to the set-valued (and $\{\mathcal{G}_t\}$-adapted) process $(\mathcal{M}_t)$ of posteriors on θ. It in turn induces a corresponding set-valued process of *predictive posteriors*, measures on $(\Omega, \mathcal{G}_\infty)$. However, when used in a maxmin model this process leads to violation of dynamic consistency. Thus, following Chen and Epstein (2002),[2] assume that DM uses the preceding posteriors to assess likelihoods only about the "next step" (that is, $\mathcal{G}_{t+dt}$), and then uses backward induction to arrive at sets $\mathcal{P}_t$ that describe beliefs about the entire future (that is, $\mathcal{G}_\infty$); mechanically, backward induction amounts to pasting together all possible selections of one-step-ahead beliefs. This construction captures prior ambiguity about the parameter θ through $\mathcal{M}_0$, learning through updating $\mathcal{M}_0$ to $\mathcal{M}_t$, knowledge of the signal structure (2.1) and backward induction reasoning built into the sets $\mathcal{P}_t$ of predictive posteriors, including in particular into $\mathcal{P}_0$. Appendix A provides a detailed and formal description.

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

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

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,

[2] The cited paper deals with a continuous-time framework. The corresponding discrete-time model and its axiomatic foundations are provided in Epstein and Schneider (2003).

her payoff is given by

$$X_t = \max\left\{\min_{\mu\in\mathcal{M}_t} E\mu, \min_{\mu\in\mathcal{M}_t} E^*\mu, \tfrac{1}{2}\right\} . \tag{2.3}$$

2.2. Parametric specification

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[3]

$$\mathcal{M}_0 = \{(1-m)\delta_{-\alpha} + m\delta_\alpha : \frac{1-\epsilon}{2} \leq m \leq \frac{1+\epsilon}{2}\} . \tag{2.4}$$

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

$$\min_{\mu\in\mathcal{M}_0} E\mu = \min_{\mu\in\mathcal{M}_0} E^*\mu = \frac{1}{2} - \epsilon\alpha < \frac{1}{2} . \tag{2.5}$$

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 B):

$$\mathcal{M}_t = \{(1-m)\delta_{-\alpha} + m\delta_\alpha : \underline{m}_t \leq m \leq \overline{m}_t\} , \tag{2.6}$$

where

$$\underline{m}_t = \frac{\frac{1-\epsilon}{1+\epsilon}\varphi(Z_t)}{1 + \frac{1-\epsilon}{1+\epsilon}\varphi(Z_t)} , \overline{m}_t = \frac{\frac{1+\epsilon}{1-\epsilon}\varphi(Z_t)}{1 + \frac{1+\epsilon}{1-\epsilon}\varphi(Z_t)} , \tag{2.7}$$

[3]The results below are unchanged if it is assumed that $\mathcal{M}_0$ consists of only the two extreme priors, those that correspond to $m = \frac{1+\epsilon}{2}$ and $m = \frac{1-\epsilon}{2}$ respectively.

and

$$\varphi(z) = \exp\left(2\alpha z/\sigma^2\right) \ . \tag{2.8}$$

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. Evidently, the response is a time-invariant function of Z_t with elasticity depending positively on the ratio α/σ^2, which is decreasing in the signal variance and increasing in the prior variance.

One obtains, therefore, that

$$\begin{aligned} \min_{\mu\in\mathcal{M}_t} E\mu &= (\tfrac{1}{2}+\alpha) - \frac{2\alpha}{1+\frac{1-\epsilon}{1+\epsilon}\varphi(Z_t)} \\ \min_{\mu\in\mathcal{M}_t} E^*\mu &= (\tfrac{1}{2}-\alpha) + \frac{2\alpha}{1+\frac{1+\epsilon}{1-\epsilon}\varphi(Z_t)} \end{aligned} \tag{2.9}$$

and hence,

$$X_t = X(Z_t) = \begin{cases} (\frac{1}{2}+\alpha) - \frac{2\alpha}{1+\frac{1-\epsilon}{1+\epsilon}\varphi(Z_t)} & \text{if } Z_t > \frac{\sigma^2}{2\alpha}\log(\frac{1+\epsilon}{1-\epsilon}) \\ (\frac{1}{2}-\alpha) + \frac{2\alpha}{1+\frac{1+\epsilon}{1-\epsilon}\varphi(Z_t)} & \text{if } Z_t < -\frac{\sigma^2}{2\alpha}\log(\frac{1+\epsilon}{1-\epsilon}) \\ \frac{1}{2} & \text{otherwise.} \end{cases} \tag{2.10}$$

2.3. Optimal stopping

Our main result gives the solution to the optimal stopping problem (2.2), assuming (2.6), (2.10) and the construction of $\mathcal{P}_0$ that adapts Chen and Epstein (2002) and is detailed in Appendix A.[4]

Let

$$l(r) = 2\log(\frac{r}{1-r}) - \frac{1}{r} + \frac{1}{1-r}, \quad r\in(0,1) \ , \tag{2.11}$$

and define $\widehat{r}$ by

$$l(\widehat{r}) = \frac{2\alpha^3}{c\sigma^2} \ . \tag{2.12}$$

$\widehat{r}$ 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 $l(1) = \infty$.

[4] Below the qualifications "almost surely" or "with probability 1" should be understood even where not stated explicitly. They are defined using any measure P in $\mathcal{P}_0$; the choice of P does not matter because the measures in $\mathcal{P}_0$ are pairwise mutually absolutely continuous (that is, equivalent).

Theorem 2.1. *(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

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

where

$$\overline{z} = \frac{\sigma^2}{2\alpha}\left[\log\frac{1+\epsilon}{1-\epsilon} + \log\frac{\overline{r}}{1-\overline{r}}\right] , \tag{2.13a}$$

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

$$l(r) + l(\frac{1+\epsilon}{2}) = \frac{4\alpha^3}{c\sigma^2} . \tag{2.14}$$

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. Note that 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. Moreover, rejection of opportunities to learn about the ambiguous urn can be rationalized even for arbitrarily small cost c. 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 region* or 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 (2.5) 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.[5]

To conclude discussion of the theorem, consider another perspective on the no-learning result. One might be inclined to understand it by comparing optimal stopping under ambiguity with two Bayesian optimal stopping problems where, respectively, the single priors are $\underline{\mu}_0$ and $\overline{\mu}_0$ given by $\underline{\mu}_0(\alpha) = \frac{1-\epsilon}{2} = \underline{m}_0$ and $\overline{\mu}_0(\alpha) = \frac{1+\epsilon}{2} = \overline{m}_0$. These are the two extreme priors in $\mathcal{M}_0$. Accordingly, one

[5] One might instead deviate from finite-support priors e.g., Fudenberg, Strack and Strzalecki (2017) assume normal priors.

might conjecture that there is a close connection between rejection of learning by these two fictional Bayesian agents and rejection of learning in our model with ambiguity. In fact, one can show that rejection of learning by Bayesian agents with single priors $\underline{\mu}_0$ and $\overline{\mu}_0$ respectively implies rejection of learning under ambiguity as in (i), *but that the converse is false*. (Formally, the Bayesian problems are classical and are special cases of Theorem 21.1 in Peskir and Shiryaev (2006), from which one can derive that no learning is optimal for both Bayesian agents if and only if $\frac{1+\epsilon}{2} > r^B$, where $l(r^B) = \frac{4\alpha^3}{c\sigma^2}$, and hence $r^B > \overline{r} > \widehat{r}$.) In particular, ambiguity about the bias corresponding to taking $\mathcal{M}_0$ equal to the convex hull of $\{\underline{\mu}_0, \overline{\mu}_0\}$ leads to more rejection of learning in general than what is implied by either $\underline{\mu}_0$ or $\overline{\mu}_0$. This is because, as described above, the worst-case scenario under ambiguity involves switching between the posteriors for $\underline{\mu}_0$ and $\overline{\mu}_0$ depending on the realized signal, rather than using either one throughout, which means that DM perceives the learning opportunity strictly less favorably than does either Bayesian agent.

A corollary describes some comparative statics results. Given two stopping strategies τ_1 and τ_2, say that τ_1 *stops later* if, for every t,

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

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 $\overline{z}_1 \geq \overline{z}_2$.

Corollary 2.2. *DM stops sampling later in each of the following cases:*
(1) c falls. (2) ϵ increases in the interval $[0, 2\widehat{r} - 1)$, where $\widehat{r}$ is defined in (2.12).
(3) σ and α both increase in such a way that α/σ^2 is constant.

That lower cost leads to longer sampling is not surprising. The second result 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, as described above, 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.

Turn to (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 (2.8)). 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.

Further properties of τ^* 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. Denote by P^θ the probability distribution of (Z_t) if θ is the true bias. Then DM stops in finite time almost surely for every θ,

$$P^\theta\left(\tau^* < \infty\right) = 1 \; .$$

In addition, the mean delay time according to P^θ is finite and given by

$$E^\theta \tau^* = \begin{cases} \left(\overline{z}/\sigma\right)^2 \left[\frac{\tanh\left(\theta\overline{z}/\sigma^2\right)}{\theta\overline{z}/\sigma^2}\right] & \text{if } \theta \neq 0 \\ \left(\overline{z}/\sigma\right)^2 & \text{if } \theta = 0 \; . \end{cases} \tag{2.15}$$

In particular, $E^\theta\tau^*$ falls as the absolute value of the true bias $\mid \theta \mid$ increases.

If $\theta > 0$, then the correct choice on stopping is to choose the bet on red, which occurs if Z_t hits $\overline{z}$ before hitting $-\overline{z}$; and symmetrically for bets on blue. It follows that[6]

$$P^\theta\left(\{\text{ correct choice of bet }\}\right) = \frac{1}{1+\exp\left(-\frac{2|\theta|}{\sigma^2}\overline{z}\right)} \;, \text{ if } \theta \neq 0 \; .$$

The probability of making the correct choice increases with ϵ and $\mid \theta \mid$, and declines with c and σ.

The proof of the theorem yields a closed-form expression for the value function associated with the optimal stopping problem. In particular, the value at time 0 satisfies (from (B.2) and (B.9)),

$$v_0 - \tfrac{1}{2} = \begin{cases} 0 & \text{if } \frac{1+\epsilon}{2} \geq \widehat{r} \\ \frac{c\sigma^2}{2\alpha^2}\left[\frac{1}{2\overline{r}(1-\overline{r})} - \frac{2}{(1+\epsilon)(1-\epsilon)}\right] & \text{if } \frac{1+\epsilon}{2} < \widehat{r} \; . \end{cases} \tag{2.16}$$

[6]Apply the optional stopping theorem to the martingale $e^{-2\theta Z_t/\sigma^2}$.

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 $\frac{1+\epsilon}{2} \geq \widehat{r}$. (The intuition for the negative relation between the value of learning and initial ambiguity was given above.) In contrast, the value of learning is increasing with the prior variance α (Appendix C).

3. Concluding remarks

Being a first step, the model is very special. We exploited its simplicity in order to solve the optimal stopping problem in closed-form and to provide a number of comparative statics predictions regarding the effects on optimal learning of prior ambiguity (ϵ), prior variance (α), signal variance (σ) and the learning cost (c). In addition, and perhaps surprisingly at first glance, we showed that the model can rationalize rejection of learning opportunities, even given ambiguity averse choices in Ellsberg's two-urn thought experiment and even for a small cost of learning.

Naturally, specificity and simplicity of the model raise questions, left for future research, about generality and robustness of the results. For example, we speculated briefly on the role of the specification of initial beliefs. Some readers may wonder about robustness to the assumption of maxmin utility. In particular, the smooth ambiguity model is a widely-used alternative for which smoothness (differentiability) is often highlighted as an important advantage, and it is available in a dynamic recursive form (Klibanoff, Marinacci and Mukerji 2009). However, it is not useful in our context because, as Skiadas (2013) has shown, in the continuous-time limit and given Brownian uncertainty, it is observationally equivalent to recursive expected utility (Kreps and Porteus 1978) thus implying indifference to ambiguity.

A. Appendix: Construction of $\mathcal{P}_0$

We show how the approach of Chen and Epstein (2002)–CE below–can be applied in a setting with learning and partial information. The construction that follows is described in part for the parametric specification in Section 2.2. However, it should be clear that it can be adapted much more generally.

Let $(\Omega, \mathcal{G}_\infty)$ with filtration $\mathcal{G} = \{\mathcal{G}_t\}_{0 \leq t < \infty}$ be defined as in Section 2.1. Fix the volatility $\sigma > 0$ as in (2.1). Suppose that under a reference probability measure P_0, $W = (W_t)_{0 \leq t < \infty}$ is a Brownian motion which generates the filtration $\{\mathcal{G}_t\}$. 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. Thus they take as primitive a set-valued process (Ξ_t), where $\Xi_t(\omega)$ is called the set of density generators at (t, ω); each $\eta_t(\omega) \in \Xi_t(\omega)$ can be thought of roughly as defining conditional beliefs about $\mathcal{G}_{t+dt}$. The associated set of processes is

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

Then

$$\mathcal{P}_0 = \{P^\eta : \eta \in \Xi\}\ , \tag{A.1}$$

where

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

By the Girsanov Theorem,

$$dW_t^\eta = \eta_t dt + dW_t \tag{A.2}$$

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

The components P_0, W, (Ξ_t) and $\{\mathcal{G}_t\}$ are primitives in CE. Here we specify them in terms of the primitives of our model; notably, (Ξ_t) models the effect on beliefs of past signal realizations, that is, learning. A noteworthy feature of the specification is that the signal process Z will serve as our reference Brownian motion under a suitably defined measure P_0.

For the purposes of this section, we take as given a measurable space $(\Omega, \mathcal{F})$ and 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 Borel measurable $A \subset \Theta$. 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 signals Z_t as in (2.1) and signals generate the subfiltration $\{\mathcal{G}_t\}$. (Adopt also other notation from Section 2.) All probability spaces are taken to be complete and all related filtrations are augmented in the usual sense.

Step 1. Take $\mu \in \mathcal{M}_0$. By standard filtering theory (Theorem 8.3 in Lipster and Shiryaev), if we replace the unknown parameter θ by the estimate $\widehat{\theta}_t^\mu = E\mu_t - \frac{1}{2}$, then we can rewrite (2.1) in the form

$$\begin{aligned} dZ_t &= \hat{\theta}_t^{\mu}(Z_t)\,dt + \sigma(dB_t + \frac{\theta - \hat{\theta}_t^{\mu}(Z_t)}{\sigma}dt) \\ &= \hat{\theta}_t^{\mu}(Z_t)\,dt + \sigma d\tilde{B}_t^{\mu}\ , \end{aligned} \tag{A.3}$$

where the innovation process $(\tilde{B}_t^{\mu})$ is a standard $\{\mathcal{G}_t\}$-adapted Brownian motion on $(\Omega, \mathcal{G}_\infty, P^\mu)$. This applies in particular for the two extreme measures, $\mu = \overline{\mu}, \underline{\mu}$, satisfying (recall (2.6) and (2.7)), for all t,

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

for which the corresponding parameter estimates are

$$\begin{aligned} \widehat{\theta}_t^{\overline{\mu}}(Z_t) &= \alpha - \frac{2\alpha}{1 + \frac{1+\epsilon}{1-\epsilon}\varphi(Z_t)}\ \text{, and} \\ \hat{\theta}_t^{\underline{\mu}}(Z_t) &= \alpha - \frac{2\alpha}{1 + \frac{1-\epsilon}{1+\epsilon}\varphi(Z_t)}\ . \end{aligned} \tag{A.4}$$

Since $(\tilde{B}_t^{\mu})$ is a standard $\mathcal{G}_t$-adapted Brownian motion on $(\Omega, \mathcal{G}_\infty, P^\mu)$, $(\tilde{B}_t^{\mu})$ takes the same role as (W_t^{η}) in CE. Rewrite (A.3) 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 CE.

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 $(\Omega, \mathcal{G}_\infty)$. Define P_0 by:

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

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

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

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 (A.3) on $(\Omega, \mathcal{F}_\infty, P^\mu)$ and the solution Z' of (A.3) on $(\Omega, \mathcal{G}_\infty, P_0^\mu)$ have identical distributions.[7] 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 $\Xi_t = \{\hat{\theta}_t^\mu/\sigma : \mu \in \mathcal{M}_0\}$ defined in (A.4). This produces the set $\mathcal{P}_0$ via (A.1) that we use in our analysis.

Remark 1. *Chen and Epstein (2002) do not discuss learning explicitly, but suggest that their framework can accommodate any model of passive learning.*[8] *We are aware of two papers that explicitly address 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 how learning affects prior ambiguity about urn composition, for example, 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 model, 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 (through CE's "κ-ignorance" specification) 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 (A.4) is absent in his model.

[7] Argue as in Example 8.6.9 of Oksendal (2005). See his Section 5.3 for discussion of weak versus strong solutions of SDEs.

[8] Cheng and Reidel (2013) describe how CE can be extended to study optimal stopping. They do not discuss learning in any detail.

B. Appendix: Proof of Theorem 2.1

First, note that the formula (2.6) describing posteriors follows from Theorem 9.1 in Lipster and Shiryaev (1977): 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

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

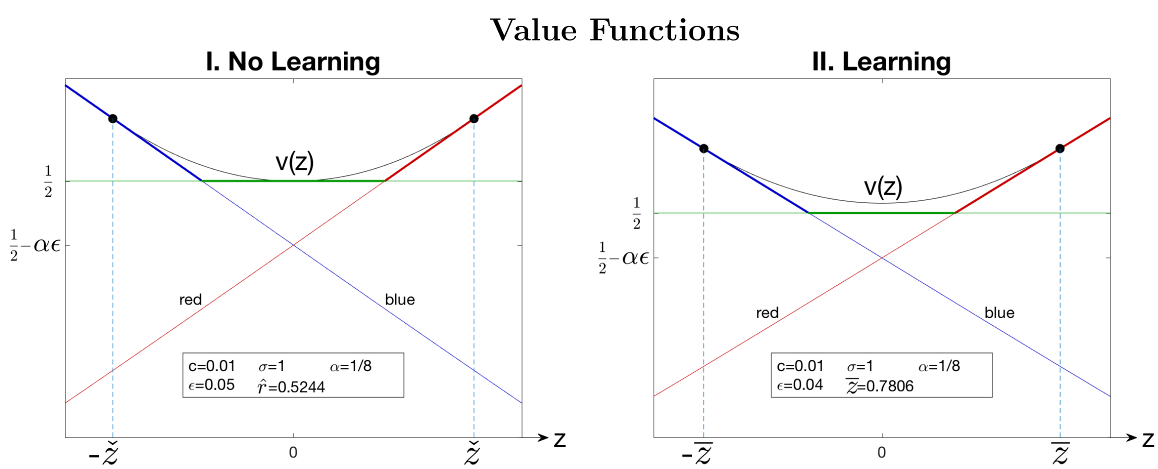
The solution is

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

where

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

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 sample size when $\epsilon = .04$ equals .61 if the true bias is zero (by (2.15)).

Proof of Theorem 2.1 (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 (2.2); 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 expression in (B.2) for the value function for step (c).

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

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

and consider the classical optimal stopping problem under the measure P^*,

$$\max_{\tau} E_{P^*}[X(Z_\tau) - c\tau] \ , \tag{B.1}$$

where $X(\cdot)$ is defined in (2.10).

It follows from received results for classical problems (Peskir and Shiryaev (2006, Ch. 6)) that the value function v for the above problem has the form:

$$\begin{aligned} C_1 &= 2\alpha - \frac{c\sigma^2}{2\alpha^2} l(\bar{r}) \, , \\ C_2 &= \frac{1}{2} + \frac{c\sigma^2}{4\alpha^2} \frac{(2\bar{r}-1)^2}{\bar{r}(1-\bar{r})}, \end{aligned}$$

$$\begin{aligned} g_1(y) &= \frac{c\sigma^2}{2\alpha^2}(2y-1)\log(\frac{y}{1-y}) - C_1(y-\frac{1}{2}) + C_2, \ 0<y<1, \\ g_2(y) &= \frac{c\sigma^2}{2\alpha^2}(2y-1)\log(\frac{y}{1-y}) + C_1(y-\frac{1}{2}) + C_2, \ 0<y<1, \end{aligned}$$

$$v(z) = \begin{cases} \frac{1}{2} - \alpha + \frac{2\alpha}{1+\frac{1+\epsilon}{1-\epsilon}\varphi(z)} & \text{if } z < -\overline{z} \\ g_1(1 - \frac{1}{1+\frac{1+\epsilon}{1-\epsilon}\varphi(z)}) & \text{if } -\overline{z} \le z < 0 \\ g_2(1 - \frac{1}{1+\frac{1-\epsilon}{1+\epsilon}\varphi(z)}) & \text{if } 0 \le z < \overline{z} \\ \frac{1}{2} + \alpha - \frac{2\alpha}{1+\frac{1-\epsilon}{1+\epsilon}\varphi(z)} & \text{if } \overline{z} \le z. \end{cases} \tag{B.2}$$

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

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

Further, under the assumption $\frac{1+\epsilon}{2} < \widehat{r}$, v satisfies the following HJB equation

$$\max\{X(z) - v(z), -c + \frac{1}{2}\sigma^2 v_{zz}(z) + f(z, -sgn(z)\epsilon)v_z(z)\} = 0 \, , \tag{B.3}$$

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

$$f(z,p) = \alpha - \frac{2\alpha}{1 + \frac{1+p}{1-p}\varphi(z)} \, . \tag{B.4}$$

Finally, v also satisfies

$$sgn(v_z(z)) = sgn(z) \text{ , and} \tag{B.5}$$

$$-c + \frac{1}{2}\sigma^2 v_{zz}(z) + f(z, -sgn(z)\epsilon)v_z(z) = 0 \quad \forall z \in (-\overline{z}, \overline{z}), \tag{B.6}$$

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 (2.2). Since $v(z)$ is time invariant, we prove only the case $t = 0$.

By Lemma B.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, (B.3), (B.5) and (B.6),

$$\begin{aligned} dv(Z_t) &= \frac{1}{2}\sigma^2 v_{zz}(Z_t)dt + v_z(Z_t)dZ_t \\ &= [c - f(Z_t, -\text{ sgn } (Z_t)\epsilon)v_z(Z_t)]dt + v_z(Z_t)dZ_t \\ &= [c - f(Z_t, -\text{ sgn } (v_z(Z_t))\epsilon)v_z(Z_t)]dt + v_z(Z_t)dZ_t \ . \end{aligned} \tag{B.7}$$

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

$$\begin{aligned} & dv(Z_t) \\ = \ & [c + (f(Z_t, \zeta_t) - f(Z_t, -\text{ sgn } (v_z(Z_t))\epsilon))\, v_z(Z_t)]dt + \sigma v_z(Z_t)dW_t^\zeta. \end{aligned} \tag{B.8}$$

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

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

Taking expectation in (B.8) under P^ζ, we have

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

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 (2.3) and (2.6) 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 2. *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): Assume $\widehat{r} \leq \frac{1+\epsilon}{2}$. Define

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

$$\begin{aligned} g_3(y) &= \frac{c\sigma^2}{2\alpha^2}(2y-1)\log(\frac{y}{1-y}) - C_3(y-\frac{1}{2}) + C_4,\ 0<y<1, \\ g_4(y) &= \frac{c\sigma^2}{2\alpha^2}(2y-1)\log(\frac{y}{1-y}) + C_3(y-\frac{1}{2}) + C_4,\ 0<y<1, \end{aligned}$$

$$v(z)=\begin{cases}\frac{1}{2}-\alpha+\frac{2\alpha}{1+\frac{1+\epsilon}{1-\epsilon}\varphi(z)} & \text{if } z\le -\check{z}\\ g_3(1-\frac{1}{1+\frac{1+\epsilon}{1-\epsilon}\varphi(z)}) & \text{if } -\check{z}<z<-\widehat{z}\\ \frac{1}{2} & -\widehat{z}\le z\le \widehat{z}\\ g_4(1-\frac{1}{1+\frac{1-\epsilon}{1+\epsilon}\varphi(z)}) & \text{if } \widehat{z}<z<\check{z}\\ \frac{1}{2}+\alpha-\frac{2\alpha}{1+\frac{1-\epsilon}{1+\epsilon}\varphi(z)} & \text{if } \check{z}\le z.\end{cases} \tag{B.9}$$

where

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

A similar analysis to that used for (ii) implies that $v(z)$ is the value function and the continuation region for this case is

$$(-\check{z},-\widehat{z})\cup(\widehat{z},\check{z})\ .$$

Note that $\frac{1+\epsilon}{2}\ge\widehat{r}$ is equivalent to $\bar{r}\le\frac{1+\epsilon}{2}$ which is also equivalent to $\widehat{z}\ge 0$. Thus $\frac{1+\epsilon}{2}\ge\widehat{r}$ implies that

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

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 $Z_0=0$. ■

C. Appendix: Miscellaneous

Proof of Corollary 2.2: 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 $\bar{z}$ increase as c falls. For $c\ge\widehat{c}$, part (i) of the theorem gives $\tau^*=0$.
2. $\bar{z}$ is increasing in ϵ: $\ell'(r)=\frac{1}{r^2(1-r)^2}$, $\frac{d\bar{z}}{d\epsilon}>0$ 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 $(1-\epsilon)(1+\epsilon)>\bar{r}(1-\bar{r})$. But $\frac{1}{2}<\frac{1+\epsilon}{2}<\widehat{r}<\bar{r}\Longrightarrow\frac{1+\epsilon}{2}\cdot\frac{1-\epsilon}{2}>\widehat{r}(1-\widehat{r})>\bar{r}(1-\bar{r})\Longrightarrow(1-\epsilon)(1+\epsilon)>4\bar{r}(1-\bar{r})>\bar{r}(1-\bar{r})$
3. If $\ell\left(\frac{1+\epsilon}{2}\right)\ge\frac{4\alpha^3}{c\sigma^2}$, then $\frac{1+\epsilon}{2}\ge\widehat{r}$ 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, $\bar{r}>\frac{1}{2}$ and $\bar{z}$ is an increasing function of $\bar{r}$, which in turn is an increasing function of α^2, hence of α. ■

Verify that v_0 in (2.16) is increasing in α: Consider two values for α, $\alpha_1 < \alpha_2$, and prove that $v_0^{\alpha_1} \leq v_0^{\alpha_2}$. Use that

$$\widehat{r}^{\alpha_1} < \widehat{r}^{\alpha_2} \,.$$

If $\frac{1+\epsilon}{2} \geq \widehat{r}^{\alpha_2}$, then there is no learning for either α and both values equal $\frac{1}{2}$. If $\widehat{r}^{\alpha_1} \leq \frac{1+\epsilon}{2} < \widehat{r}^{\alpha_2}$, then $v_0^{\alpha_1} = \frac{1}{2} < v_0^{\alpha_2}$.

The remaining possibility is $\frac{1+\epsilon}{2} < \widehat{r}^{\alpha_1}$. Then (2.16) applies for both α_1 and α_2. In the obvious notation, for $\alpha = \alpha_1$, α_2,

$$v_0^{\alpha} - \tfrac{1}{2} = \frac{c\sigma^2}{2\alpha^2}[\frac{1}{2\overline{r}^{\alpha}(1-\overline{r}^{\alpha})} - \frac{2}{(1+\epsilon)(1-\epsilon)}] \,,$$

and, by (2.14),

$$v_0^{\alpha} - \tfrac{1}{2} = \frac{2\alpha}{l(\overline{r}^{\alpha}) + l(\frac{1+\epsilon}{2})}[\frac{1}{2\overline{r}^{\alpha}(1-\overline{r}^{\alpha})} - \frac{2}{(1+\epsilon)(1-\epsilon)}].$$

For $x \in (\frac{1}{2}, 1)$, define

$$k(x) = \frac{1}{l(z) + l(\frac{1+\epsilon}{2})}[\frac{1}{2x(1-x)} - \frac{2}{(1+\epsilon)(1-\epsilon)}].$$

Since $\overline{r}^{\alpha} > \widehat{r}^{\alpha} > \frac{1}{2}$, $\overline{r}^{\alpha}$ is increasing in α and $k\left(x\right) > 0$, it suffices to prove that $k'(x) > 0$. Using $l'(x) = \frac{1}{x^2(1-x)^2}$, compute that $k'(x) =$

$$\begin{aligned}
& \frac{1}{(l(x) + l(\frac{1+\epsilon}{2}))^2} \frac{1}{x^2(1-x)^2}[(x - \frac{1}{2})(l(x) + l(\frac{1+\epsilon}{2})) - \frac{1}{2x(1-x)} + \frac{2}{(1+\epsilon)(1-\epsilon)}] \\
= \; & \frac{1}{(l(x) + l(\frac{1+\epsilon}{2}))^2} \frac{1}{x^2(1-x)^2}[(2x-1)(\log(\frac{x}{1-x}) + \frac{1}{2}l(\frac{1+\epsilon}{2})) + 2(\frac{1}{(1+\epsilon)(1-\epsilon)} - 1)].
\end{aligned}$$

Since $\frac{1}{(1+\epsilon)(1-\epsilon)} > 1$ and $x > \frac{1}{2}$, conclude that $k'\left(x\right) > 0$. ■

References

[1] M. Abdellaoui, B. Hill, E. Kemel, and H. Maafi, The evolution of ambiguity attitudes through learning, 2016.

[2] A. Baillon, H. Bleichrodt, U. Keskin, O. L'Haridon and C. Li, The effect of learning on ambiguity attitudes, *Manag. Sc.* forthcoming.

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

[4] J. V. Butler, L. Guiso and T. Jappelli, Manipulating reliance on intuition reduces risk and ambiguity aversion, 2013.

[5] J. V. Butler, L. Guiso and T. Jappelli, The role of intuition and reasoning in driving aversion to risk and ambiguity, *Theory Decision* 77 (2014), 455-484.

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

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

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

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

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

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

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

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

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

[15] P. Klibanoff, M. Marinacci and S. Mukerji, Recursive smooth ambiguity preferences, *J. Econ. Theory* 144 (2009), 930-976.

[16] D. Kreps and E. Porteus, Temporal resolution of uncertainty and dynamic choice theory, *Econometrica* 46 (1978), 185–200.

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

[18] J. Lu, Random choice and private information, *Econometrica* 84 (2016), 1983-2027.

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

[20] M. Marinacci and F. Massari, Learning from ambiguous and misspecified models, 2016.

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

[22] M. Milosavljevic, J. Malmaud, A. Huth, C. Koch and A. Rangel, The drift diffusion model can account for the accuracy and reaction time of value-based choices under high and low time pressure, *Judgement and Decision Making* 5 (2010), 437-449.

[23] N. Nicholls, A.T. Romm and A. Zimper, The impact of statistical learning on violations of the sure-thing principle, *J. Risk Uncert.* 50 (2015), 97-115.

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

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

[26] A. Rubinstein, A typology of players: between instinctive and contemplative, *Quart. J. Econ.* (2016), 1275-1303.

[27] A. Rubinstein, Response time and decision making: a "free" experimental study, *Judgement and Decision Making* 8 (2013), 540-551.

[28] A. Rubinstein, Instinctive and cognitive reasoning: a study of response times, *Econ. J.* 117 (2007), 1243-1259.

[29] C. Skiadas, Smooth ambiguity aversion toward small risks and continuous-time recursive utility, *J. Pol. Econ.* 121 (2013), 775-792.

[30] S.T. Trautman and G. van de Kuilen, Ambiguity attitudes, in G. Keren and G. Wu (eds.), *The Wiley Blackwell Handbook of Judgement and Decision Making*, Blackwell, 2015.

[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.