



Operations Research

Publication details, including instructions for authors and subscription information:
<http://pubsonline.informs.org>

Optimal Learning Under Robustness and Time-Consistency

Larry G. Epstein, Shaolin Ji

To cite this article:

Larry G. Epstein, Shaolin Ji (2020) Optimal Learning Under Robustness and Time-Consistency. Operations Research

Published online in Articles in Advance 16 Mar 2020

. <https://doi.org/10.1287/opre.2019.1899>

Full terms and conditions of use: <https://pubsonline.informs.org/Publications/Librarians-Portal/PubsOnLine-Terms-and-Conditions>

This article may be used only for the purposes of research, teaching, and/or private study. Commercial use or systematic downloading (by robots or other automatic processes) is prohibited without explicit Publisher approval, unless otherwise noted. For more information, contact permissions@informs.org.

The Publisher does not warrant or guarantee the article's accuracy, completeness, merchantability, fitness for a particular purpose, or non-infringement. Descriptions of, or references to, products or publications, or inclusion of an advertisement in this article, neither constitutes nor implies a guarantee, endorsement, or support of claims made of that product, publication, or service.

Copyright © 2020, INFORMS

Please scroll down for article—it is on subsequent pages





With 12,500 members from nearly 90 countries, INFORMS is the largest international association of operations research (O.R.) and analytics professionals and students. INFORMS provides unique networking and learning opportunities for individual professionals, and organizations of all types and sizes, to better understand and use O.R. and analytics tools and methods to transform strategic visions and achieve better outcomes.

For more information on INFORMS, its publications, membership, or meetings visit <http://www.informs.org>

Optimal Learning Under Robustness and Time-Consistency

 Larry G. Epstein,^a Shaolin Ji^b
^aDepartment of Economics, Boston University, Boston, Massachusetts 02215; ^bZhongtai Securities Institute of Financial Studies, Shandong University, 250100 Jinan, China

Contact: lepstein@bu.edu,  <https://orcid.org/0000-0002-1145-7407> (LGE); jsl@sdu.edu.cn,  <https://orcid.org/0000-0001-7420-2738> (SJ)

Received: March 2, 2018

Revised: March 3, 2019

Accepted: May 31, 2019

Published Online in Articles in Advance:
 March 16, 2020

Subject Classifications: decision analysis;
 statistics: Bayesian; utility/preference: theory

Area of Review: Special Issue: Mathematical
 Models of Individual and Group Decision Making
 in Operations Research (in honor of Kenneth
 Arrow)

<https://doi.org/10.1287/opre.2019.1899>
Copyright: © 2020 INFORMS

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. It 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, with a focus on 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 studied.

History: This paper has been accepted for the *Operations Research* Special Issue Honoring Kenneth Arrow: Mathematical Models of Individual and Group Decision Making in Operations Research.

Funding: S. Ji received financial support from the National Natural Science Foundation of China [Award Nos. 11571203 and 11871458] and from the National Key Research and Development Project of China [Award No. 2018YFA0703901].

Supplemental Material: The e-companion is available at <https://doi.org/10.1287/opre.2019.1899>.

Keywords: ambiguity • robust decisions • learning • partial information • optimal stopping • sequential testing of simple hypotheses • Ellsberg Paradox • recursive utility • time-consistency • model uncertainty

1. Introduction

We consider a decision maker (DM) choosing between three actions whose payoffs are uncertain because they depend on both exogenous randomness and on an unknown parameter θ , $\theta = \theta_0$, or θ_1 . 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 the 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 dynamically consistent 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 the above setting, with particular focus on two special cases that extend classical models. 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 (1961) 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.* The 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, a behavior that 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; and Gilboa et al. 2012). Here, we add to the thought experiment by including a possibility to learn. Specifically, we allow the 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 can be optimal to reject learning completely, and, if some learning is optimal, then it is never optimal to bet on the risky urn after stopping.* 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. The seminal papers, both using a discrete-time framework, are Wald (1945, 1947), who shows that the sequential probability ratio test (SPRT) provides an optimal trade-off between type I and type II errors, and Arrow et al. (1949), who derive SPRT from utility maximization using dynamic programming arguments. More recently, Peskir and Shiryaev (2006, chapter 6) employ a Bayesian subjectivist approach and derive SPRT as the solution to a continuous-time optimal stopping problem. We extend the latter analysis to accommodate situations where the DM, a statistician/analyst, does not have sufficient information to justify reliance on a single prior. We show that it is optimal to stop if every “compatible” Bayesian (one whose prior is an element of the set of priors used by the robustness-seeking DM) would choose to do so. But the corresponding statement for “continue” is false: It may be optimal to stop under robustness, even given a realized sample at which all compatible Bayesians would choose to continue. In this sense, “sensitivity analysis” overstates the robustness value of sampling.

We view our model as normative, a perspective that is most evident in the hypothesis testing context. Time-consistency of preference has obvious prescriptive appeal. 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, 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.

A brief review of other relevant literature concludes this introduction. The classical Bayesian model of sequential decision making, including, in particular, applications to inference and experimentation, are discussed in Howard (1970) and the references therein. The maxmin model of ambiguity-averse preference is axiomatized in a static setting in Gilboa and Schmeidler (1989) (which owes an intellectual debt to the Arrow and Hurwicz (1972) model of decision making under ignorance) and in a multiperiod discrete-time framework in Epstein and Schneider (2003), where time-consistency is one of the key axioms. Optimal stopping problems have been studied in the absence of time-consistency. 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 (Peskir 2017). 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 2018, Huang et al. 2020). Such models are problematic in 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). 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). 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).

Closest to the present paper is the literature on bandit problems with ambiguity and robustness (Caro and Das Gupta 2015, Li 2019). Both papers model endogenous learning (or experimentation) by maxmin dynamically consistent agents. Their models differ from ours in that they assume discrete time, an exogenously given horizon, and also in the nature of experimentation. In our model, the once-and-for-all choice of action and resulting payoff come after all learning has ceased, whereas in bandit problems, action choice and flow payoffs are continuous and intertwined with learning (for example, the cost of experimentation is the implied reduction in current flow payoffs). Consequently, their analyses and characterizations are much different—for example, their focus on the existence of a suitable Gittins index has no counterpart in our model.

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 applications can skip this relatively technical section and move directly to Section 3, where the “applied” optimal stopping problems are studied. The (more) general optimal stopping problem is solved in Section 4 (Theorem 4), thereby providing a unifying perspective on the two applications and some indication of the robustness of the results therein. Proofs are contained in the e-companion to this paper.

2. Recursive Utility with Learning

For background regarding time-consistency in the maxmin framework, consider first the following informal outline that anticipates the specific setting of this paper. DM faces uncertainty about a payoff-relevant state space Ω due to uncertainty about the value of a parameter $\theta \in \Theta$. Each θ determines a unique probability law on Ω , but there is prior ambiguity about the parameter that is represented by a nonsingleton set \mathcal{M}_0 of priors on Θ . As time proceeds, the DM learns about the parameter through observation of a signal whose increments are independent and identically distributed (i.i.d.) conditional on θ . At issue is how to model beliefs about Ω —that is, the set \mathcal{P}_0 of predictive priors. (Throughout, 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.) A seemingly natural approach is to take \mathcal{P}_0 to be the set of all measures that can be obtained by combining some prior μ_0 in \mathcal{M}_0 with the given conditionally i.i.d. likelihood. Learning is modeled through the set of posteriors \mathcal{M}_t at t obtained via prior-by-prior Bayesian updating of \mathcal{M}_0 and a corresponding set \mathcal{P}_t of predictive posteriors is

obtained as above. Finally, at each $t \geq 0$, \mathcal{P}_t guides choice according to the maxmin model. The point, however, is that time-consistency is violated: In general, ex ante optimal plans do not remain optimal according to updated beliefs. The reason is straightforward. Behavior at t depends on the worst-case posterior μ_t in \mathcal{M}_t , but worst cases at different nodes need not belong to same prior μ_0 . This is in contrast with the ex ante perspective expressed via \mathcal{P}_0 , where a single worst-case prior μ_0 determines the entire ex ante optimal plan. To restore dynamic consistency, one can enlarge \mathcal{P}_0 by adding to it all measures obtained by pasting together alien posteriors, leading to a “rectangular” set that is closed with respect to further pasting. One can think of the enlarged set as capturing *both* the subjectively possible probability laws *and* backward induction reasoning by the DM.

See Epstein and Schneider (2003) for further discussion and axiomatic foundations in a discrete-time framework and Chen and Epstein (2002)—CE below—for a continuous-time formulation that we outline next. 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 one-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 . 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

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

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

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

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,

$$dW_t^\eta = \eta_t dt + dW_t, \quad (1)$$

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,

$$\mathcal{P}_0 \equiv \{P^\eta : \eta \in \Xi\}. \quad (2)$$

(The “pasting” referred to above is accomplished through the fact that Ξ is constructed by taking all selections from the Ξ_t s.)

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

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

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

$$U_0(\xi) = \inf_{Q \in \mathcal{P}_0} E_Q \xi = -\sup_{Q \in \mathcal{P}_0} E_Q[-\xi]. \quad (3)$$

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

$$U_t(\xi) = \operatorname{ess\,inf}_{Q \in \mathcal{P}_0} E_Q[\xi \mid \mathcal{G}_t]. \quad (4)$$

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 $0 \leq t < t'$,

$$U_t(\xi) = \operatorname{ess\,inf}_{Q \in \mathcal{P}_0} E_Q[U_{t'}(\xi) \mid \mathcal{G}_t]. \quad (5)$$

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 (Ω, \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 (Ω, \mathcal{F}) . Although θ is an unknown deterministic parameter, for mathematical precision, we view θ as a random variable on (Ω, \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

$$Z_t = \int_0^t \theta ds + \int_0^t \sigma dB_s = \theta t + \sigma B_t, \quad (6)$$

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 $\hat{\theta}_t^\mu = \int \theta d\mu_t$, then we can rewrite (6) in the form

$$\begin{aligned} dZ_t &= \hat{\theta}_t^\mu(Z_t)dt + \sigma \left(dB_t + \frac{\theta - \hat{\theta}_t^\mu(Z_t)}{\sigma} dt \right) \\ &= \hat{\theta}_t^\mu(Z_t)dt + \sigma d\tilde{B}_t^\mu, \end{aligned} \quad (7)$$

where the innovation process (\tilde{B}_t^μ) is a standard $\{\mathcal{G}_t\}$ -adapted Brownian motion on $(\Omega, \mathcal{G}_\infty, P^\mu)$. Thus, (\tilde{B}_t^μ) takes the same role as (W_t^η) in CE (see (1) above). Rewrite (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) (respectively (resp.), $(-\hat{\theta}_t^\mu(Z_t)/\sigma)$) can be chosen as the Brownian motion (W_t) (resp., the drift (η_t)) in (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 $\bar{\mu} \in \mathcal{M}_0$ and define P_0 by:

$$\begin{aligned} \frac{dP_0}{d\bar{\mu}} \Big|_{\mathcal{G}_t} &= \exp \left\{ -\frac{1}{2\sigma^2} \int_0^t (\hat{\theta}_s^{\bar{\mu}}(Z_s))^2 ds - \frac{1}{\sigma} \int_0^t \hat{\theta}_s^{\bar{\mu}}(Z_s) d\tilde{B}_s^{\bar{\mu}} \right\} \\ &= \exp \left\{ \frac{1}{2\sigma^2} \int_0^t (\hat{\theta}_s^{\bar{\mu}}(Z_s))^2 ds - \frac{1}{\sigma^2} \int_0^t \hat{\theta}_s^{\bar{\mu}}(Z_s) dZ_s \right\}. \end{aligned}$$

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

$$\begin{aligned} \frac{dP_0^\mu}{dP_0} \Big|_{\mathcal{G}_t} &= \exp \left\{ -\frac{1}{2\sigma^2} \int_0^t (\hat{\theta}_s^\mu(Z_s))^2 ds \right. \\ &\quad \left. + \frac{1}{\sigma^2} \int_0^t \hat{\theta}_s^\mu(Z_s) dZ_s \right\}. \end{aligned}$$

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^μ) is a $\{\mathcal{G}_t\}$ -adapted Brownian motion under both P^μ and P_0^μ . Therefore, by the uniqueness of weak solutions to stochastic differential equations (SDEs), the solution Z_t of (7) on $(\Omega, \mathcal{F}_\infty, P^\mu)$ and the solution Z_t^μ of (7) on $(\Omega, \mathcal{G}_\infty, P_0^\mu)$ have identical distributions. (Argue as in Oksendal (2005, example 8.6.9).) 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

$$\Xi_t = \left\{ -\hat{\theta}_t^\mu / \sigma : \mu \in \mathcal{M}_0, \hat{\theta}_t^\mu = \int \theta d\mu_t \right\}. \quad (8)$$

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 (5)). We use this model in the optimal stopping problems that follow. The only remaining primitive is \mathcal{M}_0 , which is specified to suit the particular setting of interest.

As indicated, the key technical step in our extension of CE is in adopting the weak formulation rather than their strong formulation. For readers who may be unfamiliar with this distinction, we suggest Oksendal (2005, section 5.3) for discussion of weak versus strong solutions of SDEs and Zhang (2017, chapter 9). The latter exposit both the technical advantages of the weak formulation and its economic rationale, notably in models with imperfect information (such as here, where given (6), Z is observed but not B) or asymmetric information (such as in principal-agent models). In our context, the weak formulation is suggested if one views B not as modeling a physical noise or shock, but rather as a way to specify that the *distribution* of $(Z_t - \theta t)/\sigma$ is standard normal (conditional on θ).

Remark 1. We add a few remarks about related literature. Cheng and Riedel (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 (8) is absent.

3. Optimal Learning

3.1. The Framework and General Problem

The DM must choose an action from the set $A = \{a_0, a_1, a_2\}$. Payoffs are uncertain and depend on an unknown parameter θ . Before choosing an action, the DM can learn about θ by observing realizations of the signal process Z given by (6), where σ is a known positive constant. There is a constant per-unit-time cost $c > 0$ of learning. (The underlying state space Ω , the filtration $\{\mathcal{G}_t\}$ generated by Z , and other notation are as in Section 2. Unless specified otherwise, all processes below are taken to be $\{\mathcal{G}_t\}$ -adapted even where not stated explicitly.)

If the 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 from A . The 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. Her choice of when to stop is described by a stopping time (or strategy) τ , which is restricted to be uniformly integrable ($\sup_{Q \in \mathcal{P}_0} E_Q \tau < \infty$); the set of all stopping strategies is Γ . As a maxmin agent, she chooses an optimal stopping strategy τ^* by solving

$$\max_{\tau \in \Gamma} \min_{P \in \mathcal{P}_0} E_P(X_\tau - c\tau). \quad (9)$$

It remains to specify \mathcal{M}_0 , which determines \mathcal{P}_0 as described in Section 2, and X_t .

We assume that all priors μ in \mathcal{M}_0 have binary support $\Theta = \{\theta_0, \theta_1\}$, $\theta_0 < \theta_1$. Specifically, let

$$\mathcal{M}_0 = \{\mu^m = (1 - m)\delta_{\theta_0} + m\delta_{\theta_1} : \underline{m} \leq m \leq \bar{m}\}. \quad (10)$$

Therefore, \mathcal{M}_0 can be identified with the probability interval $[\underline{m}_0, \overline{m}_0]$ for the larger parameter value θ_1 . Let $0 < \underline{m}_0 < \overline{m}_0 < 1$.

Bayesian updating of each prior yields the following set of posteriors at t ,

$$\mathcal{M}_t = \{(1-m)\delta_{\theta_0} + m\delta_{\theta_1} : \underline{m}_t \leq m \leq \overline{m}_t\}, \quad (11)$$

where, by Liptser and Shiryaev (1977, theorem 9.1),

$$\underline{m}_t = \frac{\frac{\underline{m}_0}{1-\underline{m}_0}\varphi(t, Z_t)}{1 + \frac{\underline{m}_0}{1-\underline{m}_0}\varphi(t, Z_t)}, \quad \overline{m}_t = \frac{\frac{\overline{m}_0}{1-\overline{m}_0}\varphi(t, Z_t)}{1 + \frac{\overline{m}_0}{1-\overline{m}_0}\varphi(t, Z_t)}, \quad (12)$$

and

$$\varphi(t, z) = \exp\left\{\frac{\theta_1 - \theta_0}{\sigma^2}z - \frac{1}{2\sigma^2}(\theta_1^2 - \theta_0^2)t\right\}. \quad (13)$$

Conditional on the parameter value, payoffs are given by $u(a_i, \theta_j)$, where each $u(a_i, \theta_j)$ is nonnegative. Think of $u(\cdot, \theta_j)$ as including the valuation of any risk remaining, even if θ_j is known to be true—for example, $u(a_i, \theta_j)$ could be the expected utility of the lottery implied by (a_i, θ_j) . Payoffs are assumed to satisfy: For each $i, j = 0, 1$, $i \neq j$,

$$u(a_i, \theta_j) = u(a_i, \theta_i) > u(a_j, \theta_i). \quad (14)$$

Thus, a_0 is better than a_1 given θ_0 , and the reverse given θ_1 , and the payoff to the better action is the same for both parameter values. The payoff to the third action a_2 does not depend on θ and can be thought of as a default or outside option. Its payoff is not ambiguous because incomplete confidence about θ is the only source of ambiguity in the model, but choice of a_2 may entail risk. Adopt the notation

$$u_2 = u(a_2, \theta_0) = u(a_2, \theta_1). \quad (15)$$

It is evident that action a_2 may be irrelevant if its payoff is sufficiently low—for example, if $u_2 = 0$. To exclude the trivial case where a_2 is always chosen, assume that

$$u_2 < u(a_i, \theta_i), \quad i = 0, 1.$$

Consider next payoffs conditional on time t beliefs about θ as represented by the set of posteriors \mathcal{M}_t . The Gilboa–Schmeidler utility of a_i is $\min_{\mu \in \mathcal{M}_t} \int u(a_i, \theta) d\mu$. Therefore, if DM chooses an optimal action at time t , then her payoff is

$$X_t = \max\left\{\min_{\mu \in \mathcal{M}_t} \int u(a_0, \theta) d\mu, \min_{\mu \in \mathcal{M}_t} \int u(a_1, \theta) d\mu, u_2\right\}. \quad (16)$$

The preceding completes specification of the optimal stopping problem (9). Its solution is described in Section 4 under two *alternative* additional assumptions:

Payoff symmetry $u(a_0, \theta_1) = u(a_1, \theta_0)$; and no risky option $u_2 \leq u(a_i, \theta_j)$, $i \neq j = 0, 1$

The first assumption adds to the symmetry contained in (14). Given (14), the second implies that action a_2 is (weakly) inferior to each of a_0 and a_1 conditional on either parameter value. Hence, it would never be chosen uniquely and can be ignored, leaving only two actions. These assumptions are satisfied, respectively, by the two special models upon which we focus: Ellsberg's urns (payoff symmetry) and hypothesis testing (no risky option). We focus on these first because they extend classic models in the literature and because they provide simply distinct insights into the connection between ambiguity and optimal learning.

3.2. Learning and Ellsberg's Urns

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. Thus, θ denotes the bias toward 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. The DM can choose between betting on the draw from the risky or ambiguous urn and also on drawing red or blue. In the absence of learning, the intuitive behavior highlighted by Ellsberg is to bet on the draw from the risky urn no matter the color. Here, we consider betting preference when an ambiguity-averse decision maker can defer the choice between bets until after learning optimally about θ .

Before proceeding, we address the possible concern that our model of learning—namely, the signal structure (6)—is unnatural in the Ellsberg context, where sampling with replacement from the ambiguous urn suggests itself. However, just as the Ellsberg choice problems are understood to be metaphorical and suggestive examples, we intend our model with learning to be viewed in the same spirit. For example, one can reinterpret our version of the Ellsberg setting as follows. An investor must choose between investing in a known (risky) stock or in an unknown/unfamiliar (ambiguous) stock. If she chooses the latter, she can go long (corresponding to betting on red) or short (corresponding to betting on blue). Before choosing, she can observe the price Z_t of the ambiguous stock, which evolves according to (6) and provides information about an underlying fundamental θ .

To resume with the formal analysis, we apply the model described above with particular specifications for its key primitives A , Θ , \mathcal{M}_0 , and u . For A , let a_2 denote a bet on the risky urn and let a_1 (a_0) denote the bet on drawing red (blue) from the ambiguous urn. (There is no need to differentiate between bets on red

and blue for the risky urn.) Take $\Theta = \{\theta_0, \theta_1\}$, where $\theta_0 + \theta_1 = 0$, or, equivalently, for some $0 < \alpha < \frac{1}{2}$,

$$\theta_0 = -\alpha, \theta_1 = \alpha. \quad (17)$$

Thus, only two possible biases, of equal size, are thought possible (the proportion of red is either $\frac{1}{2} - \alpha$ or $\frac{1}{2} + \alpha$). However, there is ambiguity about which direction for the bias is more likely. This ambiguity is modeled by \mathcal{M}_0 having the form in (10), where we assume in addition that the probability interval for α (the bias toward red) is such that $\underline{m}_0 + \bar{m}_0 = 1$, or, equivalently, for some $0 < \epsilon < 1$,

$$\underline{m}_0 = \frac{1 - \epsilon}{2}, \bar{m}_0 = \frac{1 + \epsilon}{2}. \quad (18)$$

Thus, the lowest probability for a bias toward blue equals that for red, implying indifference at time 0 between bets on red and blue. This assumption, and also the color symmetry in (17), are natural because information about the ambiguous urn gives no reason to distinguish between colors.

We are left with the two parameters α and ϵ . We interpret ϵ as modeling ambiguity (aversion): the probability interval $[\frac{1-\epsilon}{2}, \frac{1+\epsilon}{2}]$ for the bias toward red 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 the 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).

Finally, specify payoffs u . All bets have the same winning and losing prizes, denominated in utils, which can be normalized to 1 and 0, respectively. Given the composition of the ambiguous urn, then only risk is involved in every bet, and an expected utility calculation yields

$$\begin{aligned} u(a_0, -\alpha) &= u(a_1, \alpha) = \alpha + \frac{1}{2}, \\ u(a_0, \alpha) &= u(a_1, -\alpha) = \alpha - \frac{1}{2}, \text{ and } u_2 = \frac{1}{2}. \end{aligned} \quad (19)$$

The assumptions in Section 3.1 are readily verified.

For convenience of the reader, we include the implied expression for the conditional payoff $X_t = X(Z_t)$:

$$X(Z_t) = \begin{cases} \left(\frac{1}{2} + \alpha\right) - \frac{2\alpha}{1 + \frac{\sigma^2}{2\alpha}\varphi(Z_t)} & \text{if } Z_t > \frac{\sigma^2}{2\alpha} \log\left(\frac{1+\epsilon}{1-\epsilon}\right) \\ \left(\frac{1}{2} - \alpha\right) + \frac{2\alpha}{1 + \frac{\sigma^2}{2\alpha}\varphi(Z_t)} & \text{if } Z_t < -\frac{\sigma^2}{2\alpha} \log\left(\frac{1+\epsilon}{1-\epsilon}\right) \\ \frac{1}{2} & \text{otherwise,} \end{cases} \quad (20)$$

where $\varphi(z) = \exp(2\alpha z/\sigma^2)$. Thus, if Z_t is large positive (negative), then a bet on drawing red (blue) from the ambiguous urn is optimal. For intermediate values,

there is not enough evidence for a bias in either direction to compensate for the ambiguity, and betting on the risky urn is optimal. This is true in particular ex ante where $Z_0 = 0$, consistent with the intuitive ambiguity-averse behavior in Ellsberg's two-urn experiment without learning.

We give an explicit solution to the optimal stopping problem (9) satisfying (17)–(19). To do so, let

$$l(r) = 2 \log \left(\frac{r}{1-r} \right) - \frac{1}{r} + \frac{1}{1-r}, \quad r \in (0, 1), \quad (21)$$

and define \hat{r} by

$$l(\hat{r}) = \frac{2\alpha^3}{c\sigma^2}. \quad (22)$$

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

Theorem 1.

(i) $\tau^* = 0$ if and only if $\frac{1+\epsilon}{2} \geq \hat{r}$, in which case $X_{\tau^*} = X_0 = \frac{1}{2}$.

(ii) Let $\frac{1+\epsilon}{2} < \hat{r}$. Then, the optimal stopping time satisfies $\tau^* > 0$ and is given by

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

where

$$\bar{z} = \frac{\sigma^2}{2\alpha} \left[\log \frac{1+\epsilon}{1-\epsilon} + \log \frac{\bar{r}}{1-\bar{r}} \right] > 0, \quad (23)$$

and $\bar{r}, \hat{r} < \bar{r} < 1$, is the unique solution to the equation

$$l(r) + l\left(\frac{1+\epsilon}{2}\right) = \frac{4\alpha^3}{c\sigma^2}. \quad (24)$$

Moreover, on stopping, either the bet on red is chosen (if $Z_{\tau^*} \geq \bar{z}$) or the bet on blue is chosen (if $Z_{\tau^*} \leq -\bar{z}$); the bet on the risky urn is never optimal at $\tau^* > 0$. Finally, if $\epsilon < \epsilon' < 2\hat{r} - 1$, and if $\tau^{*'}$ is the corresponding optimal stopping time, then $\tau^{*' \geq \tau^*$.

The two cases are defined by the relative magnitudes of ϵ , parametrizing ambiguity, and \hat{r} , which is an increasing function of $\alpha^3/(c\sigma^2)$; in particular, through α , it depends positively on the payoff to knowing the direction of the true bias. Thus, (i) considers the case where ambiguity is large relative to payoffs (and taking also sampling cost and signal variance into account). Then, no learning is optimal, and the bet on the risky urn is chosen immediately. In contrast, some learning is necessarily optimal given small ambiguity (case (ii)), including in the limiting Bayesian model with $\epsilon = 0$. Thus, it is optimal to reject learning if and only if ambiguity, as measured by ϵ , is suitably large. In case (ii), it is optimal to sample as long as the signal Z_t lies in the continuation interval $(-\bar{z}, \bar{z})$. Two features of this learning region stand out.

First, 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*. The second noteworthy feature is that *sampling increases with greater ambiguity*, as measured by ϵ , though when ϵ reaches $2\hat{r} - 1$, then, by (i), it is optimal to reject any learning.

There is simple intuition for the preceding. First, consider the effect of ambiguity (large ϵ) on the incentive to learn. The 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 $P^*(\{dZ_t > 0\} | Z_t > 0)$ and $P^*(\{dZ_t < 0\} | Z_t < 0)$ are as small as possible. That is, if $Z_t > 0$, then the distribution of the increment dZ_t is computed by using the posterior associated with that prior in \mathcal{M}_0 , which assigns the largest probability $\frac{1+\epsilon}{2}$ to the negative bias $-\alpha$, whereas if $Z_t < 0$, then the distribution of the increment is computed by using the posterior associated with the prior assigning the largest probability $\frac{1+\epsilon}{2}$ to the positive bias α . It follows that, from the perspective of the worst-case scenario, the signal structure is less informative the greater ϵ . Accordingly, conditional on some learning being optimal, then it must be with the expectation of a long sampling period that increases in length with ϵ . A second effect of an increase in ϵ is that it reduces the ex ante utility of betting on the ambiguous urn 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. At the cutoff, here $2\hat{r} - 1$, the DM is just indifferent between stopping and learning for another instant.

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 (2018), 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—the DM might as well stop and bet on the risky urn. But this reasoning is excluded when, as in our specification, the 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 the zero bias point is in the common support.

Finally, using the closed-form solution in the theorem, we can give more concrete expression to the effect of ambiguity on optimal learning. Restrict attention to values of ϵ in $[0, 2\hat{r} - 1)$, where some learning is optimal, and denote by P^θ the probability distribution of (Z_t) if θ is the true bias. Then, by well-known results regarding hitting times of Brownian motion with drift (Borodin and Salminen 2015), the mean sample length according to P^θ is

$$E^\theta \tau^* = \begin{cases} (\bar{z}/\sigma)^2 \left[\frac{\tanh(\theta\bar{z}/\sigma^2)}{\theta\bar{z}/\sigma^2} \right] & \text{if } \theta \neq 0 \\ (\bar{z}/\sigma)^2 & \text{if } \theta = 0, \end{cases} \quad (25)$$

which is increasing in ϵ . Note also that $\theta Z_{\tau^*} > 0$ if and only if the bet on red (blue) is chosen on stopping if $\theta > 0$ ($\theta < 0$). Thus, the probability, if $\theta \neq 0$ is the true bias, of choosing the “correct” bet on stopping is given by

$$P^\theta(\{\theta Z_{\tau^*} > 0\}) = \frac{1}{1 + \exp\left(-\frac{2|\theta|\bar{z}}{\sigma^2}\right)}, \text{ if } \theta \neq 0,$$

which increases with ϵ . (To prove this equality, apply the optional stopping theorem to the P^θ -martingale $e^{-2\theta Z_t/\sigma^2}$.)

The proof of Theorem 1 yields a closed-form expression for the value function associated with the optional stopping problem. In particular, the value at time 0 satisfies (from EC.4 and EC.10 in the e-companion),

$$v_0 - \frac{1}{2} = \begin{cases} 0 & \text{if } \frac{1+\epsilon}{2} \geq \hat{r} \\ \frac{c\sigma^2}{4\alpha^2} \left[\frac{1}{\hat{r}(1-\hat{r})} - \frac{4}{(1+\epsilon)(1-\epsilon)} \right] & \text{if } \frac{1+\epsilon}{2} < \hat{r}. \end{cases} \quad (26)$$

Because the payoff $\frac{1}{2}$ is the best available without learning, $v_0 - \frac{1}{2}$ is the value of the learning option. It is positive for small $\epsilon < 2\hat{r} - 1$ and declines continuously to 0 as ϵ increases to the switch point. (Note that $\frac{1+\epsilon}{2} = \hat{r}$ implies that both are equal in turn to \hat{r} , and, hence, that v_0 is continuous at $\epsilon = 2\hat{r} - 1$.) This is consistent with intuition given above.

As a numerical example, let $(c, \sigma, \alpha) = (0.01, 1, \frac{1}{8})$, which gives 0.0488 as the cutoff for ϵ . Thus, learning is rejected if $\epsilon = 0.05$. For $\epsilon = 0.04$, however, $\tau^* > 0$ and $E\tau^* = 0.61$ under $P^{\theta=0}$. Neither of the values for ϵ is extreme: In the classic Ellsberg setting (with no learning), they imply probability equivalents for the bet on red equal to 0.4875 and 0.4900 for $\epsilon = 0.05$ and $\epsilon = 0.04$, respectively.

3.3. A Robust Sequential Hypothesis Test

The DM samples the signal process Z with the objective of then choosing between the two statistical hypotheses

$$H_0 : \theta = 0 \text{ and } H_1 : \theta = \beta,$$

where $\beta > 0$. The novelty relative to Arrow et al. (1949) and Peskir and Shiryaev (2006) is that there is prior ambiguity about the value of θ , and a robust decision procedure is sought.

The following specialization of the general model is adopted. Let $\Theta = \{0, \beta\}$. The actions a_0 and a_1 are accept H_0 and accept H_1 , respectively. A third action is absent because there is no “outside option”—one of the hypotheses must be chosen. (Formally, one could include a_2 and specify its payoff below to be zero, in which case it would never be chosen.) The set of priors \mathcal{M}_0 is as given in (10), corresponding to the probability interval $[\underline{m}_0, \overline{m}_0]$ for $\theta = \beta$. Finally, payoffs are given by

$$\begin{aligned} u(a_0, 0) &= u(a_1, \beta) = a + b, \\ u(a_0, \beta) &= b, u(a_1, 0) = a, \end{aligned}$$

where $a, b > 0$. (Payoffs in this context are usually specified in terms of a loss function that is to be minimized. The loss function L satisfying $L(a_0, 0) = L(a_1, \beta) = 0$, $L(a_0, \beta) = a$, and $L(a_1, 0) = b$, gives an equivalent reformulation.)

There are two differences in specification from the Ellsberg context. First, there is no counterpart of the risky urn when choosing between hypotheses. Second, although symmetry between colors is natural in the Ellsberg context, symmetry between hypotheses is not; thus, b need not equal a , and the probability interval $[\underline{m}_0, \overline{m}_0]$ need not be symmetric about $\frac{1}{2}$.

The optimal stopping problem (9) admits a closed-form solution. For perspective, consider first the special Bayesian case ($\mathcal{M}_0 = \{\mu\}$; hence, $\mathcal{M}_t = \{\mu_t\}$, $\mu_t(\beta) = m_t$). Denote by $\tilde{r}_B^\ell < \tilde{r}_B^R$ the solutions to (33) below, which in this context simplifies to

$$\begin{aligned} l(\tilde{r}_B^R) - l(\tilde{r}_B^\ell) &= \frac{a+b}{\tilde{c}} \\ \frac{1}{\tilde{r}_B^R(1-\tilde{r}_B^R)} - \frac{1}{\tilde{r}_B^\ell(1-\tilde{r}_B^\ell)} &= \frac{b-a}{\tilde{c}}. \end{aligned} \quad (27)$$

Then, we have the following classical result.

Theorem 2 (Shiryaev 2008, Peskir and Shiryaev 2006). *In the Bayesian case, for any prior probability m_0 , it is optimal to continue at t if and only if*

$$\tilde{r}_B^\ell < m_t < \tilde{r}_B^R. \quad (28)$$

Otherwise, it is optimal to accept H_1 or H_0 according as $m_t \geq \tilde{r}_B^R$ or $m_t \leq \tilde{r}_B^\ell$, respectively.

In the model with ambiguity, the cut-off values are \tilde{r}^ℓ and \tilde{r}^R , $\tilde{r}^\ell < \tilde{r}^R$, that solve the appropriate version of (33), and we have the following generalization of the classical result.

Theorem 3. *In the model with ambiguity, it is optimal to stop and accept H_1 or H_0 according as $\underline{m}_t \geq \tilde{r}^R$ or $\overline{m}_t \leq \tilde{r}^\ell$, respectively. Otherwise, it is optimal to continue.*

In addition, if $a = b$, then

$$\tilde{r}_B^\ell < \tilde{r}^\ell \text{ and } \tilde{r}^R < \tilde{r}_B^R. \quad (29)$$

Under the assumption of payoff symmetry ($a = b$), the theorem has noteworthy implications for the relation between the optimal stopping strategies for the Bayesian and the robustness-seeking DM. (We conjecture that (29) is valid, even if $a \neq b$, but a proof has escaped us.) If $m_0 \in [\underline{m}_0, \overline{m}_0]$, refer to a *compatible Bayesian*. The theorem implies:

1. If every compatible Bayesian stops and chooses a_i , then it is optimal also for the DM to stop and choose a_i , $i = 1, 2$.
2. If every compatible Bayesian continues, then it may still be optimal for the DM to stop.

In other words, the DM should accept a unanimous recommendation of compatible Bayesian experts if she is to stop and choose a specific action, but not necessarily if she is to continue. In this sense, “sensitivity analysis” overstates the robustness value of sampling.

The intuition is clear. Prior ambiguity leads to the signal structure being perceived as less likely to be informative (seen from the perspective of the worst-case measure P^* —see the outline below of the proof of Theorem 4), even though the signal structure itself is not ambiguous. In contrast, there is no counterpart given multiple Bayesian agents—each is confident in beliefs about θ and is certain that signal increments are conditionally i.i.d. Only the DM internalizes uncertainty about the probability law and discounts the benefits of learning accordingly.

Remark 2. As is made clear in Theorem 4, stopping conditions can be stated equivalently in terms of either the signal process (as in the Ellsberg model) or posteriors (as here). In the text, we have adopted the formulations that seem more natural for each particular setting. For example, the use of posteriors above facilitates comparison with the classical Bayesian result.

Remark 3. 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 et al. (1996) point to the law of iterated expectations as responsible for excluding foregone conclusions (if the prior is countably additive). Equation (5) is a non-linear counterpart that we suspect plays a similar role in our model (though details are beyond the scope of this paper).

4. A More General Theorem

In order to condense notation, we write u_{ij} in place of $u(a_i, \theta_j)$, $i, j = 0, 1$.

Theorem 4 below describes the solution to the optimal stopping problem in Section 3.1 assuming either payoff symmetry ($u_{01} = u_{10}$) or no risky option ($u_2 \leq \min\{u_{10}, u_{01}\}$). Payoff symmetry is satisfied in Theorem 1, but the latter assumes more, specifically ex ante indifference between a_0 and a_1 ($\underline{m}_0 + \bar{m}_0 = 1$) and $u_2 = \frac{1}{2}(u_{00} + u_{10})$. Thus, it is extended below by Theorem 4(a). The assumption of no risky option is the crucial element in the hypothesis testing example, and the corresponding optimal stopping problem is isomorphic to that in Theorem 4(b).

Both \bar{m}_t and \underline{m}_t defined in (12) are increasing functions of $\varphi(t, z_t)$. It follows that there exists a unique pair of probabilities $\underline{\pi}$ and $\bar{\pi}$ and a unique (deterministic) signal realization trajectory (\tilde{z}_t) satisfying, for every t ,

$$\underline{\pi} = \underline{m}_t(\tilde{z}_t), \bar{\pi} = \bar{m}_t(\tilde{z}_t), \text{ and} \\ \underline{\pi}u_{11} + (1 - \underline{\pi})u_{10} = \bar{\pi}u_{01} + (1 - \bar{\pi})u_{00}.$$

For example, $\tilde{z}_0 = 0$, $\underline{\pi} = \underline{m}_0$, and $\bar{\pi} = \bar{m}_0$ if and only if a_0 and a_1 are indifferent ex ante. More generally, a_0 and a_1 are indifferent conditional on the signal \tilde{z}_t at t and a_0 (a_1) is preferred at t if $Z_t < (>) \tilde{z}_t$.

Normalize the cost of learning to \hat{c} , $\hat{c} = 2c\sigma^2/(\theta_1 - \theta_0)^2$.

Optimal stopping strategies will be described in terms of several critical values that are, in turn, defined by using the functions l and \tilde{l} : For all r in $(0, 1)$,

$$l(r) = 2 \log \left(\frac{r}{1-r} \right) - \frac{1}{r} + \frac{1}{1-r} \\ \tilde{l}(r) = \log \left(\frac{r}{1-r} \right) + \frac{r}{1-r}.$$

Let (r_1^R, r_2^R) , (r_1^I, r_2^I) , (r^R, r^I) , and $(\tilde{r}^R, \tilde{r}^I)$ solve the following equations, respectively:

$$l(r_2^R) - l(r_1^R) = \frac{u_{11} - u_{10}}{\hat{c}} \\ \tilde{l}(r_2^R) - \tilde{l}(r_1^R) = \frac{u_2 - u_{10}}{\hat{c}}, \quad (30)$$

$$l(r_2^I) - l(r_1^I) = -\frac{u_{00} - u_{01}}{\hat{c}} \\ \tilde{l}(r_2^I) - \tilde{l}(r_1^I) = \frac{u_2 - u_{00}}{\hat{c}}, \quad (31)$$

$$l(r^R) - l(\underline{\pi}) = \frac{u_{11} - u_{10}}{\hat{c}} \\ l(r^I) - l(\bar{\pi}) = -\frac{u_{00} - u_{01}}{\hat{c}}, \quad (32)$$

$$l(\tilde{r}^R) - l(\underline{\pi}) = l(\tilde{r}^I) - l(\bar{\pi}) + \frac{u_{11} - u_{10} + u_{00} - u_{01}}{\hat{c}} \\ \tilde{l}(\tilde{r}^R) - \tilde{l}(\underline{\pi}) - \underline{\pi}(l(\tilde{r}^R) - l(\underline{\pi})) = \\ \tilde{l}(\tilde{r}^I) - \tilde{l}(\bar{\pi}) - \bar{\pi}(l(\tilde{r}^I) - l(\bar{\pi})). \quad (33)$$

(The latter reduces to (32) if payoff symmetry is satisfied.)

Define

$$u_2^{**} = \frac{\hat{c}}{2} \left[\frac{1}{r^I(1-r^I)} - \frac{1}{\bar{\pi}(1-\bar{\pi})} \right] + \frac{u_{00} - u_{01}}{2}. \quad (34)$$

Besides the existence and uniqueness assertions, the next lemma proves a number of properties that are important for the optimal stopping theorem to follow.

Lemma 1. *There exist unique solutions to (32) and (33), and the solutions to the latter satisfy*

$$\tilde{r}^I < \bar{\pi}, \tilde{r}^R > \underline{\pi}. \quad (35)$$

*If $u_2 \geq u_2^{**}$, then there exist unique solutions also to (30) and (31), and the solutions satisfy*

$$r_2^I < r_1^I, r_1^R < r_2^R, \underline{\pi} < r^R, r^I < \bar{\pi}.$$

If payoff symmetry is also satisfied, then:

$$\underline{\pi} + \bar{\pi} = 1 = r^I + r^R, \text{ and} \quad (36)$$

$$r_1^I \leq \bar{\pi} \iff r_1^R \geq \underline{\pi} \iff u_2 \geq u_2^{**}. \quad (37)$$

Define

$$\bar{f}(t, r) = \frac{\theta_1 + \theta_0}{2} t + \frac{\sigma^2}{\theta_1 - \theta_0} \log \left(\frac{1 - \bar{m}_0}{\bar{m}_0} \frac{r}{1-r} \right) \\ \underline{f}(t, r) = \frac{\theta_1 + \theta_0}{2} t + \frac{\sigma^2}{\theta_1 - \theta_0} \log \left(\frac{1 - \underline{m}_0}{\underline{m}_0} \frac{r}{1-r} \right).$$

Then, $\underline{m}_t(\underline{f}(t, r)) = r = \bar{m}_t(\bar{f}(t, r))$, and, for any r_1 and r_2 ,

$$\bar{f}(t, r_1) \leq \tilde{z}_t \iff r_1 \leq \bar{\pi} \\ \underline{f}(t, r_2) \geq \tilde{z}_t \iff r_2 \geq \underline{\pi}. \quad (38)$$

Finally, define three stopping times:

$$\begin{aligned}\tau_0 &\equiv \min\{t \geq 0 : Z_t \leq \bar{f}(t, r_2^l)\} \\ &= \min\{t \geq 0 : \bar{m}_t \leq r_2^l\}, \\ \tau_1 &\equiv \min\{t \geq 0 : Z_t \geq \underline{f}(t, r_2^R)\} \\ &= \min\{t \geq 0 : \underline{m}_t \geq r_2^R\}, \text{ and} \\ \tau_2 &\equiv \min\{t \geq 0 : \bar{f}(t, r_1^l) \leq Z_t \leq \underline{f}(t, r_1^R)\} \\ &= \min\{t \geq 0 : \bar{m}_t \geq r_1^l \text{ and } \underline{m}_t \leq r_1^R\}.\end{aligned}$$

Theorem 4. (a) Assume payoff symmetry ($u_{01} = u_{10}$).

(a.i) If $r_1^l \leq \bar{\pi}$, then the optimal stopping time τ^* is given by

$$\tau^* = \min\{\tau_i : i = 0, 1, 2\}.$$

Moreover, if $\tau^* = \tau_i$, then a_i is optimal on stopping. In particular, if there is ex ante indifference between a_0 and a_1 ($\underline{\pi} = \underline{m}_0$ and $\bar{\pi} = \bar{m}_0$), then $\tau^* = 0$ and a_2 is chosen.

(a.ii) If $r_1^l > \bar{\pi}$, then

$$\begin{aligned}\tau^* &= \min\{t \geq 0 : Z_t \leq \bar{f}(t, r^l) \text{ or } Z_t \geq \underline{f}(t, r^R)\} \\ &= \min\{t \geq 0 : \bar{m}_t \leq r^l \text{ or } \underline{m}_t \geq r^R\}.\end{aligned}$$

Moreover, a_0 is optimal on stopping if $Z_{\tau^*} \leq \bar{f}(\tau^*, r^l)$ (equivalently, if $\bar{m}_{\tau^*} \leq r^l$), a_1 is optimal if $Z_{\tau^*} \geq \underline{f}(\tau^*, r^R)$ (equivalently, if $\underline{m}_{\tau^*} \geq r^R$), and a_2 is never optimal.

(b) Assume $u_2 \leq \min\{u_{10}, u_{01}\}$. Then,

$$\begin{aligned}\tau^* &= \min\{t \geq 0 : Z_t \leq \bar{f}(t, \tilde{r}^l) \text{ or } Z_t \geq \underline{f}(t, \tilde{r}^R)\} \\ &= \min\{t \geq 0 : \bar{m}_t \leq \tilde{r}^l \text{ or } \underline{m}_t \geq \tilde{r}^R\}.\end{aligned}$$

Moreover, a_0 is optimal on stopping if $Z_{\tau^*} \leq \bar{f}(\tau^*, \tilde{r}^l)$ (equivalently if $\bar{m}_{\tau^*} \leq \tilde{r}^l$), a_1 is optimal if $Z_{\tau^*} \geq \underline{f}(\tau^*, \tilde{r}^R)$ (equivalently if $\underline{m}_{\tau^*} \geq \tilde{r}^R$), and a_2 is never optimal.

In (a), the distinction between the two subcases depends on the relative magnitudes of r_1^l and $\bar{\pi}$. From (31), it follows that r_1^l falls as u_2 increases, while $\bar{\pi}$ does not depend on u_2 . Therefore, (a.i) applies if the payoff u_2 to the unambiguous default is sufficiently large. The other factor leading to (a.i) is large $\bar{\pi}$, equivalently (by (36)) small $\underline{\pi}$, which is supported by \bar{m}_0 large and \underline{m}_0 small. Thus, (a.i) is supported also by large prior ambiguity.

In (a.i), $\tau^* = 0$ if either $\bar{m}_0 \leq r_2^l$ (prior beliefs are strongly biased toward θ_0 , and, hence, a_0 is chosen immediately), or $\underline{m}_0 \geq r_2^R$ (prior beliefs are strongly biased toward θ_1 and, hence, a_1 is chosen), or $\bar{m}_0 \geq r_1^l$ and $\underline{m}_0 \leq r_1^R$ (the worst-case probabilities of both θ_0 and θ_1 are both sufficiently low that neither a_0 nor a_1 is attractive enough to justify the cost of sampling, and,

hence, a_2 is chosen). That leaves continuation being optimal at time 0 if and only if prior beliefs are “intermediate” in the sense that

$$\begin{aligned}\text{either: } & \left[r_2^l < \bar{m}_0 < r_1^l \right] \text{ and } \underline{m}_0 < r_2^R, \\ \text{or: } & \left[r_1^R < \underline{m}_0 < r_2^R \right] \text{ and } \bar{m}_0 > r_2^l.\end{aligned}$$

This continuation region could be empty. Because learning is only about the payoffs to a_0 and a_1 , the situation at time 0 that is least favorable to learning is where there is ex ante indifference between a_0 and a_1 —then, a long and, hence, costly sample would likely be needed to modify the ex ante ranking of actions. In this case, therefore, it is optimal to reject learning and choose a_2 , as in Theorem 1. However, if, for example, a_1 is strictly preferred initially, then an incentive to learn is that a relatively short interval of sampling may be enough to decide between a_1 and a_2 . In addition, if \underline{m}_0 is sufficiently large, say, near 1, then near certainty that $\theta = \theta_1$ can lead to rejection of learning and the immediate choice of a_1 , rather than of a_2 as in the Ellsberg context.

In (a.ii), $\tau^* = 0$ if and only if $[\underline{m}_0, \bar{m}_0]$ is disjoint from (r^l, r^R) . Notably, the default action is not chosen, regardless of when sampling stops. Its payoff u_2 is too low (from (37), $u_2 < u_2^*$) compared with the expected payoff of choosing a_0 or a_1 , possibly after some learning. Moreover, even given some learning, it is not optimal to choose a_2 , regardless of the realized sample, as explained in discussion of Theorem 1. Under ex ante indifference, Lemma 1 implies that $\tau^* > 0$ in (a.ii). Combined with (a.i), we see that if there is ex ante indifference between a_0 and a_1 , then a_2 is chosen if and only if there is no learning, thus generalizing the result in the Ellsberg model. (The latter also assumes $u_2 = \frac{1}{2}(u_{00} + u_{10})$, which we see here is not needed for the preceding conclusion.)

Finally, consider (b), where the payoff to the unambiguous action is so low that it would never be chosen, regardless of prior beliefs and even in the absence of the option to learn. The optimal strategy is similar to that in (a.ii) in form and interpretation—only the critical values may differ to reflect the different assumptions about payoffs. Another comment about (b) is that when $\bar{m}_0 = \underline{m}_0$, then $\bar{\pi} = \underline{\pi}$ and equations (33) defining the critical values \tilde{r}^R and \tilde{r}^l become

$$\begin{aligned}l(\tilde{r}^R) - l(\tilde{r}^l) &= \frac{u_{11} - u_{10} + u_{00} - u_{01}}{\hat{c}}, \\ \tilde{l}(\tilde{r}^R) - \tilde{l}(\tilde{r}^l) &= \frac{u_{00} - u_{10}}{\hat{c}},\end{aligned}$$

which are equations (21.1.14) and (21.1.15) in Peskir and Shiryaev (2006).

Proof of the theorem is provided in the e-companion. Here, we comment briefly on the proof strategy.

The strategy is to: (1) guess the P^* in \mathcal{P}_0 that is the worst-case scenario; (2) solve the classical optimal stopping problem given the single prior P^* by applying the techniques of Peskir and Shiryaev (2006, chapter 6) and El Karoui et al. (1997, theorems 8.5, 8.6); (3) show that the value function derived in (2) is also the value function for our problem (9); and (4) use the value function to derive τ^* .

The intuition for the conjectured P^* was given in Section 3.2 for the Ellsberg context. In this more general context, it extends to the conjecture that P^* should make $P^*(\{dZ_t > 0\} | Z_t > \tilde{z}_t)$ and $P^*(\{dZ_t < 0\} | Z_t < \tilde{z}_t)$ as small as possible, by using \underline{m}_t when $Z_t > \tilde{z}_t$ and \bar{m}_t when $Z_t < \tilde{z}_t$. (See EC.1 in the e-companion for the precise definition of P^* .) The search for the value function v begins with the Hamilton–Jacobi–Bellman equation, which yields its functional form up to some constants to be determined by smooth contact conditions between v and the payoff function X (see Peskir and Shiryaev 2006 for this free-boundary approach to analyzing optimal stopping problems). A new ingredient relative to existing models stems from the nature of P^* , specifically from the fact that the relevant posterior probability at t switches between \underline{m}_t and \bar{m}_t as described, implying that the form of the value function differs between the regions $Z_t > \tilde{z}_t$ and $Z_t < \tilde{z}_t$. Thus, in addition to ensuring a smooth contact at stopping points, one must also be concerned with the smooth connection at \tilde{z}_t .

We elaborate on the latter point in order to highlight the technical novelty that arises from ambiguity. For concreteness, consider (a.ii), where a_2 is never chosen. Let y denote a posterior probability, computed by using \underline{m}_0 or \bar{m}_0 , depending on the subdomain, and let $V^R(y) : [\underline{\pi}, 1] \rightarrow [0, +\infty)$ and $V^l(y) : [0, \bar{\pi}] \rightarrow [0, +\infty)$ denote corresponding candidates for the value in the indicated regions. Then, the variational inequality and smooth contacts lead to the following free-boundary differential equation, in which $r^R \in (\underline{\pi}, 1]$ and $r^l \in [0, \bar{\pi})$ are also unknowns to be determined:

$$\begin{cases} V_{yy}^R(y) = \hat{c} \frac{1}{y^2(1-y)^2}, & y \in (\underline{\pi}, r^R) \\ V^R(r^R) = (u_{11} - u_{10})r^R + u_{10} \\ V_y^R(r^R) = (u_{11} - u_{10}) \\ V_{yy}^l(y) = \hat{c} \frac{1}{y^2(1-y)^2}, & y \in (r^l, \bar{\pi}) \\ V^l(r^l) = -(u_{00} - u_{01})r^l + u_{00} \\ V_y^l(r^l) = -(u_{00} - u_{01}), \end{cases} \quad (39)$$

and the (new) smooth contact conditions due to ambiguity ($\underline{\pi} < \bar{\pi}$):

$$\begin{cases} V^R(\underline{\pi}) = V^l(\bar{\pi}), \\ V_y^R(\underline{\pi}) = V_y^l(\bar{\pi}). \end{cases} \quad (40)$$

In (a.ii), payoff symmetry leads to the simplification $V_y^R(\underline{\pi}) = V_y^l(\bar{\pi}) = 0$, which leads to (32) becoming two separated equations. However, in (b), the connection is not trivial.

Acknowledgments

The authors are grateful for suggestions from two referees and for comments from Tomasz Strzalecki. An earlier version of this paper, titled “Optimal Learning and Ellsberg’s Urns,” was posted on ArXiv in August 2017.

References

Arrow KJ, Hurwicz L (1972) An optimality criterion for decision making under ignorance. Carter C, Ford J, eds. *Uncertainty and Expectations in Economics* (Basil Blackwell, Oxford, UK), 1–11.

Arrow KJ, Blackwell D, Girshick MA (1949) Bayes and minimax solutions of sequential decision problems. *Econometrica* 17(3/4): 214–244.

Berger JO (1984) The robust Bayesian viewpoint. Kadane J, ed. *Robustness in Bayesian Statistics* (North Holland, Amsterdam), 63–124.

Berger JO (1985) *Statistical Decision Theory and Bayesian Analysis* (Springer, New York).

Berger JO (1994) An overview of robust Bayesian analysis (with discussion). *TEST* 3(1):5–124.

Borodin A, Salminen P (2015) *Handbook of Brownian Motion—Facts and Formulae*, 2nd ed. (Birkhauser, Basel, Switzerland).

Caro F, Das Gupta A (2015) Robust control of the multi-armed bandit problem. *Ann. Oper.*, ePub ahead of print August 21, <https://doi.org/10.1007/s10479-015-1965-7>.

Chen Z, Epstein LG (2002) Ambiguity, risk and asset returns in continuous time. *Econometrica* 70(4):1403–1443.

Cheng X, Riedel F (2013) Optimal stopping under ambiguity in continuous time. *Math. Financial Econom.* 7(1):29–68.

Choi H (2016) Learning under ambiguity: Portfolio choice and asset returns. Working paper, City University of Hong Kong, Kowloon Tong, Hong Kong.

Ebert S, Strack P (2018) Never, ever getting started: On prospect theory without commitment Working paper, Frankfurt School of Finance and Management, Frankfurt, Germany.

El Karoui N, Kapoundjian C, Pardoux E, Peng S, Quenez M (1997) Reflected solutions of backward SDE’s and related obstacle problems for PDE’s. *Ann. Probab.* 25(2):702–737.

Ellsberg D (1961) Risk, ambiguity, and the Savage axioms. *Quart. J. Econom.* 75(4):643–669.

Epstein LG, Schneider M (2003) Recursive multiple-priors. *J. Econom. Theory* 113(1):1–31.

Epstein LG, Schneider M (2007) Learning under ambiguity. *Rev. Econom. Stud.* 74(4):1275–1303.

Epstein LG, Schneider M (2008) Ambiguity, information quality and asset pricing. *J. Finance* 63(1):197–228.

Epstein LG, Schneider M (2010) Ambiguity and asset markets. *Annual Rev. Financial Econom.* 2:315–346.

Fudenberg D, Strack P, Strzalecki T (2018) Speed accuracy and the optimal timing of choices. *Amer. Econom. Rev.* 108(2): 3651–3684.

Gilboa I (2009) *Theory of Decision Under Uncertainty* (Cambridge University Press, New York).

Gilboa I (2015) Rationality and the Bayesian paradigm. *J. Econom. Methods* 22(3):312–334.

Gilboa I, Schmeidler D (1989) Maxmin expected utility with non-unique prior. *J. Math. Econom.* 18(2):141–153.

Gilboa I, Postlewaite A, Schmeidler D (2012) Rationality of belief or: Why Savage’s axioms are neither necessary nor sufficient for rationality. *Synthese* 187(1):11–31.

- Howard RA (1970) Decision analysis: Perspectives on inference, decision, and experimentation. *Proc. IEEE* 58(5):632–643.
- Huang YJ, Nguyen-Huu A, Zhou XY (2020) General stopping behaviors of naive and non-committed sophisticated agents, with application to probability distortion. *Math. Finance* 30(1): 310–340.
- Huber PJ (1965) A robust version of the probability ratio test. *Ann. Math. Statist.* 36(6):1753–1758.
- Kadane JB, Schervish MJ, Seidenfeld T (1996) Reasoning to a foregone conclusion. *JASA* 91(435):1228–1235.
- Li J (2019) The K-armed bandit problem with multiple-priors. *J. Math. Econom.* 80(1):22–38.
- Liptser R, Shiryaev A (1977) *Statistics of Random Processes I: General Theory* (Springer, Berlin).
- Marinacci M (2002) Learning from ambiguous urns. *Statist. Papers* 43(1):143–151.
- Miao J (2009) Ambiguity, risk and portfolio choice under incomplete information. *Ann. Econom. Finance* 10(2):257–279.
- Oksendal B (2005) *Stochastic Differential Equations*, 6th ed. (Springer, Berlin).
- Peskir G (2017) Nonlinear optimal stopping and nonlinear optimal control. Presentation, Conference on Optimal Stopping in Complex Environments, Bielefeld, Germany.
- Peskir G, Shiryaev A (2006) *Optimal Stopping and Free-Boundary Problems* (Springer, Berlin).
- Rios-Insua D, Ruggeri F (2000) *Robust Bayesian Analysis* (Springer, New York).
- Shapiro A (2016) Rectangular sets of probability measures. *Oper. Res.* 64(2):528–541.
- Shiryaev A (2008) *Optimal Stopping Rules*, 2nd ed. Aries AB, trans. (Springer, Berlin).
- Trautman ST, Zeckhauser RJ (2013) Shunning uncertainty: The neglect of learning opportunities. *Games Econom. Behav.* 79:44–55.
- Wald A (1945) Sequential tests of statistical hypotheses. *Ann. Math. Statist.* 16(2):117–186.
- Wald A (1947) *Sequential Analysis* (Wiley, New York).
- Walley P (1991) *Statistical Reasoning with Imprecise Probabilities* (Chapman and Hall, London).
- Zhang J (2017) *Backward Stochastic Differential Equations* (Springer, New York).

Larry G. Epstein is a professor of economics at Boston University. His main research interest is in decision theory, including choice under risk and under ambiguity (model uncertainty), with specific emphasis on dynamic settings and models of learning and updating. He also studies applications to finance (asset pricing) and statistics.

Shaolin Ji is a professor of mathematics at Shandong University. His research interests are in mathematical finance, stochastic optimal control, and nonlinear expectation theory.