ECONtribute
MARKETS & PUBLIC POLICY

ECONtribute
Discussion Paper No. 194

Prolonged Learning and Hasty Stopping: the Wald Problem with Ambiguity

Sarah Auster Yeon-Koo Che Konrad Mierendorff

August 2022 www.econtribute.de

UNIVERSITÄT BONN Universität zu Köln

Funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany´s Excellence Strategy – EXC 2126/1-390838866 is gratefully acknowledged.

Prolonged Learning and Hasty Stopping: the Wald Problem with Ambiguity*

Sarah Auster Yeon-Koo Che Konrad Mierendorff†

August 29, 2022

Abstract

This paper studies sequential information acquisition by an ambiguity-averse decision maker (DM), who decides how long to collect information before taking an irreversible action. The agent optimizes against the worst-case belief and updates prior by prior. We show that the consideration of ambiguity gives rise to rich dynamics: compared to the Bayesian DM, the DM here tends to experiment excessively when facing modest uncertainty and, to counteract it, may stop experimenting prematurely when facing high uncertainty. In the latter case, the DM's stopping rule is non-monotonic in beliefs and features randomized stopping.

Keywords: Wald problem, ambiguity aversion, prolonged learning, preemptive stopping
JEL Classification Numbers: C61, D81, D83, D91

*We thank David Ahn, Nageeb Ali, Renee Bowen, Martin Cripps, Mark Dean, Francesc Dilmé, Daniel Gottlieb, Will Grimme, Yoram Halevy, Nenad Kos, Stephan Lauermann, Bart Lipman, Georg Noeldeke, Jawwad Noor, Pietro Ortoleva, Wolfgang Pesendorfer, Ludvig Sinander, Philipp Strack, Tomasz Strzalecki, Bruno Strulovici, Juuso Toikka, Yangfan Zhou and audiences in various seminars and conferences for helpful comments and discussions. Sarah Auster gratefully acknowledges funding from the German Research Foundation (DFG) through Germany's Excellence Strategy—EXC 2126/1—390838866 and CRC TR 224 (Project B02).

†Auster: Department of Economics, University of Bonn (email: auster.sarah@gmail.com); Che: Department of Economics, Columbia University (email: yeonkooche@gmail.com); Mierendorff: Department of Economics, University College London

1 Introduction

The problem of making a decision on an action after deliberating on its merits is ubiquitous in many situations of life. Examples include a firm contemplating a potential new project, a jury deliberating on a verdict, a citizen pondering over his/her political vote, or a scientist evaluating a hypothesis. The pervasiveness of such a problem makes the framework of central importance, as seen by the extensive research on the subject matter; see Wald (1947); Arrow et al. (1949); Moscarini and Smith (2001); Fudenberg et al. (2018); Che and Mierendorff (2019), among others. With the exception of a few recent papers, the vast literature studying the problem considers a Bayesian framework. The Bayesian model misses, however, a crucial aspect of the problem: the decision maker (DM) often lacks a clear model or beliefs about the uncertainty she is facing. In fact, Wald, often regarded as the first to formally introduce the framework, was mindful of the ambiguity and suggests a robust approach (Wald, 1947), which anticipates the maxmin decision rule axiomatized by Gilboa and Schmeidler (1989).

We adopt Gilboa and Schmeidler (1989) by considering a DM who has ambiguous beliefs about the state of world and evaluates her choices against the worst-case scenario in a set of plausible priors. Specifically, the state is either L or R and the DM must decide between two irreversible actions, ℓ and r. At each point, the DM can stop and choose an action. Alternatively, she can postpone her action and learn about the state. In the baseline model, the learning takes the form of a Poisson breakthrough that confirms state R. Imagine, for instance, a CEO of a biotech company seeking to establish the efficacy of a new drug or a theorist seeking to prove a theorem. The Bayesian optimal strategy in such a situation calls for looking for evidence of state R—efficacy of the drug or validity of the proof—for a duration of time, until a long stretch of unsuccessful attempts convinces her to declare R to be unlikely and stop for action ℓ.

Introducing ambiguity in a dynamic setting raises conceptual issues about updating and dynamic consistency. To avoid the latter, the existing literature often follows a recursive approach, requiring the DM to adopt a prior-by-prior updating rule on a *rectangular* set of priors (see Epstein and Schneider 2003). Rectangularity requires the set of priors to include all measures that can be obtained by pasting together alien posteriors, thereby guaranteeing the existence of a single worst-case belief that guides the DM's conditional choices. We follow a different approach: rather than assuming away dynamic inconsistency, we seek to understand its behavioral implications for a "sophisticated" DM who recognizes potential preference reversals and deals with them in a forward looking and rational manner. To this end, we maintain prior-by-prior updating, but allow for changes in the conditional worst-case scenario.

The paper shows that the consideration of ambiguity and time inconsistency gives rise to rich dynamics in sequential learning. Late in the process, ambiguity increases the DM's incentives to confirm the state, thereby prolonging experimentation. After a period of looking for R-evidence, the DM contemplates two possibilities: no such evidence exists because the state is L or instead the state is R and she simply got unlucky. As time progresses, the DM becomes increasingly convinced of the first. Following a sufficiently long period of futile search, the Bayesian DM thus stops to take action ℓ. The ambiguity-averse DM, on the other hand, continues to worry about the second possibility and, as a consequence, keeps looking for more proof. Ambiguity aversion thus causes the DM to continue experimenting well past the time at which the Bayesian counterpart would stop. In the context of our example, the biotech CEO or the theorist refuses to see the "writing on the wall," and gets hung up on proving the efficacy of the drug or the validity of the theorem. From the perspective of an outsider, the ambiguity-averse DM would thus be seen to exhibit prolonged indecision.

Early on, however, the DM is concerned about the opposite scenario, namely that R is relatively unlikely and the search for evidence will be in vain. From this vantage point, her later self's refusal to discount R sufficiently enough to stop experimentation in time is seen as misguided. Recognizing the peril of continued indecision, the DM must then intervene and stop before it is too late, namely before the change in the worst-case belief occurs. The DM adopts a mixed strategy in this case, randomizing between experimentation and stopping. She does so by stopping according to a Poison process at a rate chosen to hedge against both events. The stopping occurs around the point of belief switch, which is typically well before the Bayesian DM would find it reasonable to stop. Hence, to an outside observer, the DM would be seen to exhibit premature decisiveness.

Comparative statics yields further implications of ambiguity. An increase in ambiguity—as measured by the size of the initial set of priors—increases the tendency for the DM to prolong her experimentation later in the game, which increases her need to preemptively stop earlier in the process (and, naturally, the rate at which she does so). These effects, taken together, imply that increased ambiguity "blunts" the DM's ability to adapt her learning strategy to the true likelihoods of states. Strikingly, this means that, for a sufficiently large ambiguity, the DM's experimentation time *increases* rather than decreases as the true likelihood of L increases. The combined effects also imply that the stopping time is more dispersed with larger ambiguity.

Finally, we consider two interesting extensions of the baseline model. First, we allow the DM to choose between two different news sources, one generating R-evidence, as before, and the other generating L-evidence. We show that excessive learning and randomized stopping also arise with sufficiently large ambiguity and sufficiently small information cost. In addition, the DM may "split" her attention between the two news sources in order to hedge, even when a Bayesian DM would not do so under any prior. Second, we consider the case of incremental learning with diffusion signals. Although some details of the solution change, the main features of the model—excessive learning and preemption via randomized stopping—persist in this case.

In sum, our analysis captures two behavioral traits—protracted indecision and hasty, often unwarranted, decisiveness. The other main contribution of the paper is methodological. Analyzing dynamic decision-making under ambiguity is a priori difficult because the associated time inconsistency problem makes the dynamic programming machinery inapplicable. We characterize the solution of our problem via a set of novel 'saddle-point Hamilton-Jacobi-Bellman (HJB) conditions'. Our HJB characterization restores the dynamic programming logic in a saddle-point sense and hence serves as a powerful toolkit that can make analysis tractable without compromising sharp prediction. We expect the approach, and the micro-foundation justifying it, to be valuable beyond our setting.

Related Literature. We incorporate ambiguity into an optimal stopping framework à la Wald (1947). Optimal stopping problems in the Bayesian framework with Poisson signals have been analyzed in Peskir and Shiryaev (2006, ch.VI), with economic applications including Che and Mierendorff (2019) and Nikandrova and Pancs (2018). Other applications adopt a drift-diffusion model, e.g. Moscarini and Smith (2001), Fudenberg et al. (2018), while Zhong (2019) studies the problem under flexible choice of information.

The paper is part of a growing literature on optimal stopping problems under ambiguity. Most of the existing literature adopts a recursive approach, thereby precluding the possibility of changes in the worst-case scenario and time inconsistency, e.g. Epstein and Schneider (2007), Riedel (2009), Miao and Wang (2011), Cheng and Riedel (2013). Closest to our paper in this literature is Epstein and Ji (2020), which studies a Wald problem in a drift-diffusion framework with recursive preferences. Under the recursive approach, there is a single prior in the initial set that minimizes the DM's expected payoff at every point in time. Due to this feature, standard dynamic programming methods are applicable and the optimal stopping rule continues to be described by two stopping boundaries, as in the Bayesian benchmark. By contrast, dynamic inconsistency figures prominently in our model, which we analyze by developing a saddle-point version of the dynamic programming technique.

There is also a growing literature looking at the implication of dynamic inconsistency due to ambiguity in different applications. For example, Bose and Daripa (2009), Bose and Renou (2014), Ghosh and Liu (2021) and Auster and Kellner (2022) consider auctions/mechanisms with ambiguity-averse agents, while Kellner and Le Quement (2018) and Beauchêne et al. (2019) study ambiguity and updating in communication and persuasion settings.

Finally there is a literature studying dynamic inconsistent stopping in other behavioral frameworks. For instance, Barberis (2012), Xu and Zhou (2013), Ebert and Strack (2015, 2018), and Henderson et al. (2017) analyze stopping problems under prospect theory. Christensen and Lindensjö (2018; 2020) study stopping problems for a class of time inconsistent models, which can capture endogenous habit formation, non-exponential discounting and mean-variance utility.

2 Model

A DM must choose between two actions, r and ℓ, whose payoffs depend on an unknown state $\omega \in \{R, L\}$. Specifically, the DM's payoffs satisfy $u_r^R > u_\ell^R$ and $u_\ell^L > u_r^L$, where $u_a^\omega \in \mathbb{R}$ denotes her payoff from choosing action $a \in \{r, \ell\}$ in state $\omega \in \{R, L\}$. The DM thus wants to "match" the action with the state. We further assume $u_r^R > u_r^L$ and $u_\ell^R < u_\ell^L$. This means that whether a state is desirable or not depends on the action that the DM takes, a feature that makes ambiguity nontrivial.[1] Let $U_a(p) = pu_a^R + (1-p)u_a^L$ denote the DM's payoff associated with action a when the probability of state R is p.

The DM may delay her action and acquire information. We model the DM's information acquisition problem as a stopping problem in continuous time. At each point in time $t \geq 0$, the DM decides whether to wait for more information or whether to stop and take an immediate action, ℓ or r. If the DM waits (or "experiments"), she incurs a flow cost $c > 0$ per unit of time. Payoffs are realized when the DM stops and takes an action. Information takes the form of R-evidence which is received only in state R with arrival rate $\lambda > 0$. Observing news thus conclusively reveals state R, while the absence of news is a signal favoring state L.

[1] If this assumption does not hold, the ambiguity-averse DM minimizing over a set of probabilities is behaviorally indistinguishable from a Bayesian DM, whose prior is equal to the element of the set that maximizes the likelihood of the "bad" state.

Bayesian benchmark. Considering the benchmark case of a Bayesian DM, let p_t denote the probability that the DM assigns to state R at time t. In the absence of news, the DM's belief drifts "leftward" according to

$$\dot{p}_t = \eta(p_t) := -\lambda p_t(1 - p_t).$$

Intuitively, a DM becomes pessimistic toward R when she looks for R-evidence but is unable to find any.

As is well known, the Bayesian optimal information acquisition rule can be described by two thresholds, p_ℓ^B and p_r^B. If $p_t \le p_\ell^B$ or $p_t \ge p_r^B$, the DM is sufficiently confident in the state and takes the appropriate action without gathering any additional information. If instead p_t belongs to (p_ℓ^B, p_r^B), the DM experiments until either she receives R-evidence or the Bayesian update of p_t reaches the threshold p_ℓ^B.

$$0 \underbrace{\text{———————}}_{\text{action } \ell} p_\ell^B \underbrace{\longleftarrow\longleftarrow\longleftarrow\longleftarrow\longleftarrow\longleftarrow\longleftarrow\longleftarrow}_{\text{experimentation}} p_r^B \underbrace{\text{———————}}_{\text{action } r} 1$$

Let $\phi(p; p', U_\ell(p'))$ denote the value to the Bayesian DM of experimenting until the belief p reaches $p' < p$, assuming that, if she receives R-evidence during the experimentation, she realizes u_r^R, while if she does not receive such evidence until her belief reaches p', she stops and takes action ℓ. The standard HJB conditions characterize DM's value function as a solution to an ODE:

$$c = \lambda p(u_r^R - \phi(p)) + \phi'(p)\eta(p), \forall p > p'. \tag{1}$$

Intuitively, the flow cost of experimentation (LHS) equals the rate of value increase due to possible breakthrough news and belief updating in its absence (RHS). Optimal stopping calls for p' to be chosen optimally, so the value of the stopping problem is

$$\Phi(p) = \max_{p' \in [0,p]} \phi(p; p', U_\ell(p')), \tag{2}$$

and the left stopping boundary p_ℓ^B is the associated maximizer.[2] Finally, since the DM has an option to choose r at any point in time, the DM with belief p enjoys the payoff of

$$\Phi^*(p) = \max\{\Phi(p), U_r(p)\}, \tag{3}$$

[2] One can also characterize the optimal stopping boundary in a dynamic programming framework by invoking value matching and smooth pasting.

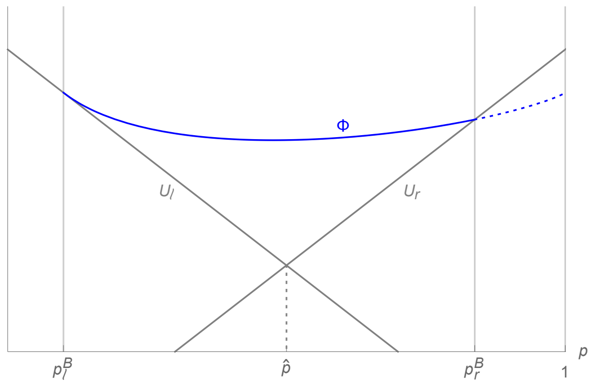

Figure 1: Bayesian optimal stopping rule.

with the DM indifferent between experimentation and action r at the right stopping boundary p_r^B. We will call $\Phi^*(p)$ the **Bayesian value function** throughout. Its characterization is relegated to Appendix B.1.

Figure 1 illustrates graphically the value to the Bayesian DM in a typical situation. The downward sloping line shows the DM's expected payoff from taking action ℓ as a function of the probability p she assigns to state R. Similarly, the upward sloping line represents the DM's expected payoff from taking action r. The crossing point $\hat{p}$ is the belief at which the DM is indifferent between actions ℓ and r. The blue curve depicts Φ, i.e. the value from experimenting with stopping belief p_ℓ^B. The optimal value Φ^* can be seen as the upper envelope of Φ and $U := U_\ell \vee U_r$. Naturally, the experimentation range (p_ℓ^B, p_r^B) depends on the flow cost c; the range shrinks as c increases and disappears when

$$c \geq \bar{c} := \lambda \frac{(u_r^R - u_\ell^R)(u_\ell^L - u_r^L)}{(u_r^R - u_\ell^R) + (u_\ell^L - u_r^L)}. \tag{4}$$

The threshold $\bar{c}$ is the value of c at which $\Phi(\hat{p}) = U(\hat{p})$.

Ambiguity. We assume that the DM faces ambiguity over the likelihood of state R and that her preferences are represented by the maxmin expected utility model (Gilboa and Schmeidler, 1989). According to this model, the DM seeks to maximize her payoff guarantee across a set of priors, described by an interval of probabilities on state R: $P_0 = [\underline{p}_0, \overline{p}_0]$. We assume $0 < \underline{p}_0 \leq \overline{p}_0 < 1$.

It will be useful to consider the DM's optimal information acquisition rule from the perspective of time 0. Specifically, suppose the DM can commit to a dynamic strategy that maximizes her expected payoff against the worst prior in P_0. Finding such a strategy amounts to analyzing a zero-sum game between the DM and adversarial nature choosing $p \in P_0$ to minimize the DM's expected payoff. Its equilibrium—a saddle point—describes the DM's highest value obtained from any information acquisition rule. By the minmax theorem, the value of the saddle point can be obtained by the minimax problem where nature first chooses a prior in P_0 and the DM then chooses her best strategy for the chosen prior (see Osborne and Rubinstein (1994) Proposition 22.2). It follows that the minimax value is simply the lowest Bayesian value Φ^* within P_0, or

$$\min_{p \in P_0} \Phi^*(p).$$

In particular, the commitment value coincides with the Bayesian value at the left-most belief $\underline{p}_0$ of P_0 if the prior beliefs P_0 are sufficiently far to the right, and at the right most belief $\overline{p}_0$, if P_0 is sufficiently far to the left, as depicted in Figure 2. For later purpose, we denote the global minimizer of Φ^* by

$$p_* := \arg \min_{p \in [0,1]} \Phi^*(p).$$

The commitment solution is characterized in Appendix B.2. While the maxmin commitment value is easily obtained, the saddle point strategy on the part of the DM requires a little care. If $p_* < p_r^B$ or $p_r^B \notin P_0$, then the maxmin strategy is indeed the Bayesian optimal strategy for the worst belief in P_0. If $p_* = p_r^B \in P_0$, however, nature's indifference, which is needed for the saddle point requirement, calls for the DM to randomize. If $c < \bar{c}$, then the DM must randomize between r and experimentation. If instead $c \geq \bar{c}$, experimentation is too costly and the DM mixes between actions r and ℓ with probabilities $\hat{\rho}$ and $1 - \hat{\rho}$, respectively, where $\hat{\rho}$ satisfies

$$\hat{\rho} U_r'(p) + (1 - \hat{\rho}) U_\ell'(p) = 0. \tag{5}$$

Updating and preference reversals. A key question is how the DM updates her ambiguous beliefs as information arrives. We assume that the DM uses *Full Bayesian Updating*:[3] she updates each element of P_0 according to Bayes rule and considers the worst case over the set of posterior beliefs obtained in this way. The set of posteriors at time $t > 0$ is thus given by an interval $P_t := [\underline{p}_t, \overline{p}_t]$, where $\underline{p}_t$ is the Bayesian update of the initial lower bound $\underline{p}_0$ and $\overline{p}_t$ is the

[3]This rule is also referred to as *prior-by-prior updating*.

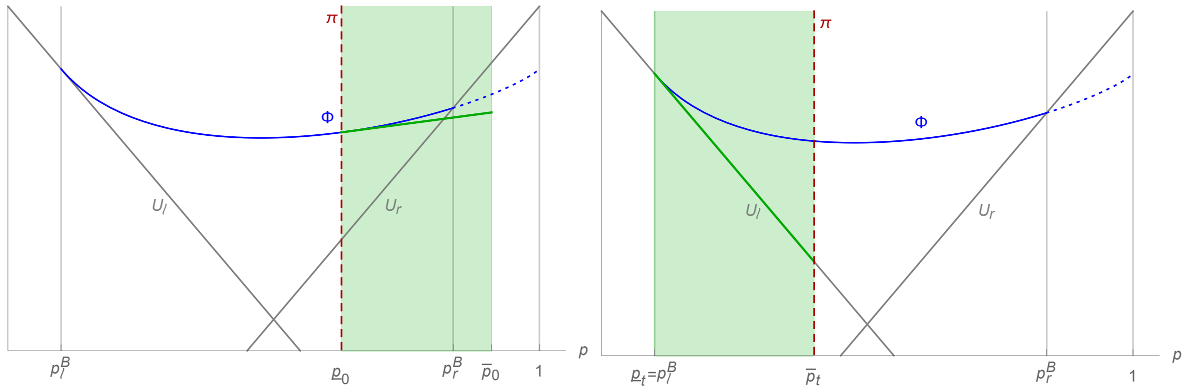

Figure 2: Worst-case scenario (indicated by the red dashed line) at the start of experimentation (left panel) and at the ex-ante optimal stopping time (right panel).

Bayesian update of the initial upper bound $\overline{p}_0$. As time progresses and no breakthrough occurs, the interval P_t drifts to the left. In the process, the interval expands in size as it moves away from the right boundary and shrinks as it approaches the left boundary. Rescaling the belief in terms of a log-likelihood ratio keeps the interval "size" constant. Hence, from now on, we shall refer to

$$\Delta := \ln\left(\frac{\overline{p}_0}{1-\overline{p}_0}\right) - \ln\left(\frac{\underline{p}_0}{1-\underline{p}_0}\right)$$

as *the degree of ambiguity*. It is easy to verify that Δ does not vary as one updates beliefs over time.

As is well known, the Full Bayesian Updating rule does not guarantee dynamic consistency. To see this, suppose the interval P_0 is sufficiently far to the right so that $\underline{p}_0$ minimizes Φ^* over P_0 (see the left panel of Figure 2). If $\underline{p}_0 < p_r^B$, the DM's optimal (commitment) plan at time 0 is to experiment until either she receives R-evidence or the Bayesian update $\underline{p}_t$ of $\underline{p}_0$ reaches p_ℓ^B. Once the intended stopping point t where $\underline{p}_t = p_\ell^B$ is reached, however, the worst-case belief is no longer given by $\underline{p}_t$ but instead by $\overline{p}_t$, as illustrated in the right panel of Figure 2. According to $\overline{p}_t$, the optimal plan (as prescribed by the commitment solution) is to acquire more information, until the Bayesian update of $\overline{p}_t$ reaches the stopping boundary p_ℓ^B. Hence, later selves of the DM will renege on the original plan prescribed by the commitment solution and *experiment excessively* from the perspective of the original self.

A dynamic inconsistency arises as a result of a switch in the worst-case scenario. When the DM experiments, she is concerned about two events: 1) not receiving R-evidence in state L and

2) taking action ℓ in state R. At the start of the experimentation phase, pessimism about the first event dominates and the worst-case belief is given by $\underline{p}_t$, which minimizes the likelihood of R. When the DM gets closer to the intended stopping point, the concern about taking the wrong action looms large and the worst-case belief switches to $\overline{p}_t$, i.e. the belief maximizing the likelihood of R. The new worst-case belief calls for more experimentation, thereby creating a conflict with the preferences of earlier selves. A sophisticated DM understands that later selves deviate from the commitment plan by experimenting for too long and optimizes accordingly. We next analyze the behavior of such a DM.

3 Sophisticated Stopping Rule

We now analyze the dynamic behavior of a sophisticated DM. The dynamic inconsistency problem requires us to analyze the DM's stopping problem as an intrapersonal game in which a current self plays against future selves.

3.1 Intrapersonal game.

At each point in time $t \geq 0$, the DM decides between continuing to gather information and stopping, taking as given the choices of her future selves. If a breakthrough occurs, all posteriors in the updated set are equal to one, so it is optimal for the DM to stop and take action r. Taking this part of the strategy as given, it is without loss of generality to condition the DM's strategy on the set of posteriors rather than time. Indeed, in the absence of R-evidence, there is a one-to-one mapping between calendar time and the set of posteriors. Having fixed Δ as a primitive, we can index the possible sets of posteriors by their upper bound $\overline{p} \in (0,1)$. The right-most belief in the set, $\overline{p}$, is thus our state variable. We denote the set of posteriors at state $\overline{p}$ by

$$P(\overline{p}) := \left\{ p \leq \overline{p} : \ln\left(\frac{\overline{p}}{1-\overline{p}}\right) - \ln\left(\frac{p}{1-p}\right) \leq \Delta \right\},$$

and define $\underline{p}(\overline{p}) := \min P(\overline{p})$ as the lower bound of this set. Clearly, it follows that $\underline{p}_t = \underline{p}(\overline{p}_t)$.

A Markov strategy is a measurable function $\sigma : (0,1) \to [0,\infty) \times [0,1] \times [0,1]$, with $\sigma(\overline{p}) = (\nu(\overline{p}), m(\overline{p}), \rho(\overline{p}))$. The strategy specifies a *stopping rate* $\nu(\overline{p})$, an *instantaneous stopping probability* $m(\overline{p})$, and the *probability of taking action* r *conditional on stopping* $\rho(\overline{p})$. Intuitively, densities in the distribution of stopping points are accommodated by interior stopping rates

$\nu(\overline{p})$, while atoms in the distribution over stopping points are captured by positive values of $m(\overline{p})$. To ensure that a strategy σ induces a well-defined stopping time, we impose the following *admissibility* restrictions on strategies: (1) every connected set $\mathcal{P} \subseteq [0,1]$ with $m(\overline{p}) = 1$ for all $\overline{p} \in \mathcal{P}$ contains its supremum; (2) there is a finite set $M^\sigma \subset (0,1)$ such that $m(\overline{p}) \in (0,1)$ only if $\overline{p} \in M^\sigma$.[4]

Fixing a strategy σ, we denote by $V^\sigma(p, \overline{p})$ the DM's expected payoff at state $\overline{p}$ when the strategy is evaluated at belief p. Nature's goal is to minimize $V^\sigma(p, \overline{p})$ at each state $\overline{p}$. Since $V^\sigma(p, \overline{p})$ is a convex combination of the payoffs that strategy σ yields in state R and in state L with coefficient p, the DM's value $V^\sigma(p, \overline{p})$ is linear in the first argument. Nature's problem thus has a bang-bang solution. Letting $\pi(\overline{p}) \in P(\overline{p})$ denote nature's choice at state $\overline{p}$, we define a solution concept as follows.

Definition 1. A strategy $\sigma = (\nu, m, \rho)$ constitutes a *pseudo equilibrium* if for each state $\overline{p} \in (0,1)$, there exists some $\pi = \pi(\overline{p})$ with

$$\pi(\overline{p}) \in \arg\min_{p \in P(\overline{p})} V^\sigma(p, \overline{p}) \tag{6}$$

such that[5]

$$U(\pi) \leq V^\sigma(\pi, \overline{p}); \tag{7}$$

$$U(\pi) = V^\sigma(\pi, \overline{p}) \text{ if } \nu(\overline{p}) \text{ or } m(\overline{p}) > 0; \tag{8}$$

$$\rho(\overline{p}) \in \arg\max_{\rho \in [0,1]} \rho U_r(\pi) + (1-\rho) U_\ell(\pi). \tag{9}$$

A strategy σ is a pseudo equilibrium if, given nature's choice π, no self of the DM has an incentive to deviate. Due to the continuous time nature of our game, many pseudo equilibria exist. For instance, it is a pseudo equilibrium for the DM to stop at every state and choose the optimal action against the worst-case belief. If some self of the DM were to deviate and refuse to stop, her infinitesimally close future selves would stop immediately. Since experimentation for an instant yields zero probability of a breakthrough, the deviation is never profitable. However,

[4]Condition (1) requires that whenever a strategy specifies instantaneous stopping for an interval of states, then the interval is closed on the right. This ensures that the stopping time is well defined when the belief drifts toward a (left) stopping boundary. Meanwhile, Condition (2) states that the set of states in which the DM randomizes instantaneously between continuation and stopping is finite. In principle, such states can be (countably) infinite, but this extra flexibility proves irrelevant.

[5]Recall that $U = U_\ell \vee U_r$ is the maximum stopping payoff from actions ℓ and r.

the prescribed strategy is implausible and simply an artifact of the continuous time nature of our game.

To rule out such implausible strategies, we focus on the pseudo equilibria in which the DM can control her action for a vanishingly small amount of time. Formally, for any pseudo equilibrium σ we imagine the DM who can commit at each time t her action for period $[t, t+\epsilon)$ against adversarial nature, anticipating that σ will follow after $t+\epsilon$. We then require σ to be obtained as a limit of such ϵ-commitment solutions as $\epsilon \to 0$.[6] Such a vanishing control over future behavior avoids total coordination failure among selves. As justified in Appendix A.1 more precisely, the refinement introduces a saddle-point version of dynamic programming conditions.

To begin, for a value function $V : [0,1]^2 \to \mathbb{R}$, define a functional:

$$\begin{aligned} G(m, \nu, \rho, p, \overline{p}, V, dV) =& m\left[U_\rho(p) - V(p, \overline{p}_-)\right] + (1-m)\left[-c + \nu(U_\rho(p) - V(p, \overline{p}_-)) \right.\\ &\left. + p\lambda(u_r^R - V(p, \overline{p}_-)) + V_p(p, p_-)\eta(p) + V_{\overline{p}}(p, \overline{p}_-)\eta(\overline{p})\right], \end{aligned}$$

where dV is the gradient of the value function, $V(p, \overline{p}_-) := \lim_{\overline{p}' \uparrow \overline{p}} V(p, \overline{p}')$, and $V_x(p, p_-) := \lim_{\overline{p}' \uparrow \overline{p}} \frac{\partial V(p, \overline{p}')}{\partial x}$.[7] Recall $\eta(p) = -\lambda p(1-p)$ denotes the law of motion. The functional G has a familiar interpretation from the dynamic programming literature. Its first term captures a value increase when the DM stops with a point mass m; she collects $U_\rho(p) - V(p, \overline{p}_-)$. The second term (inside the square brackets) captures a value increase when the DM "continues." In this case, at the flow cost c, the DM engages in a flow stopping at rate ν and experimentation yielding a breakthrough at rate λp: her value increases by $U_\rho(p) - V(p, \overline{p}_-)$ in the former case and by $u_r^R - V(p, \overline{p}_-)$ in the latter case. The last two terms track the change of value arising from the updating of the state $\overline{p}$ and the updating of the current belief.

For each $\overline{p} \in (0,1)$, consider a saddle point version of Hamilton-Jacobi-Bellman equations:

$$0 = \max_{m,\nu,\rho} G(m, \nu, \rho, \pi(\overline{p}), \overline{p}, V, dV), \tag{10}$$

$$\sigma(\overline{p}) = (m(\overline{p}), \nu(\overline{p}), \rho(\overline{p})) \in \arg\max_{m,\nu,\rho} G(m, \nu, \rho, \pi(\overline{p}), \overline{p}, V, dV), \tag{11}$$

[6] This refinement is similar in effect to taking a limit of the behavior in discrete times models as the time interval shrinks to zero; in such models the DM's experimentation during a unit period has some non-negligible probability of producing breakthrough. The current refinement accomplishes the same effect within a continuous time framework.

[7] We use the left limits of the value function and its partial derivatives (a) to make sure the conditions are well defined even at point where the derivatives do not exist and (b) since the belief drifts downwards and the value function in the HJB equation reflects the continuation value.

$$\pi(\overline{p}) \in \arg\min_{p \in P(\overline{p})} G(m(\overline{p}), \nu(\overline{p}), \rho(\overline{p}), p, \overline{p}, V, dV). \tag{12}$$

Condition (10) captures the optimality requirement embodied in the value function: the variation from the optimal value for the DM is precisely zero at the equilibrium. Condition (11) requires the DM's flow decision to be optimal against nature's choice π, taking her future strategy as given. This is not required in Definition 1, as is seen in the implausible "always stopping" behavior. Finally, (12) chooses the belief to be worst for the DM against the flow action, taking the continuation strategy into consideration. Effectively, equations (10), (11), and (12) represent the saddle point version of HJB equations.

We define our equilibrium as follows:

Definition 2. A pseudo equilibrium $\sigma = (\nu, m, \rho)$, together with π, is an *intrapersonal equilibrium*, or simply an *equilibrium*, if it satisfies (10), (11), and (12) for each $\overline{p} \in (0, 1)$.

The precise microfoundation for this solution concept based on vanishing commitment power is provided in Appendix A.1. The HJB conditions for dynamic decisions under ambiguity, and the micro-foundation justifying them, are of independent interest that we believe are applicable well beyond the particular setting we consider here. While time inconsistency makes the dynamic programming machinery inapplicable, the HJB characterization restores the dynamic programming logic in a saddle point sense. This characterization serves as a powerful toolkit that can make analysis tractable without compromising sharp prediction, as we argue is the case here.

3.2 Main Result

To state our main result, we introduce an additional threshold for the flow cost c:

$$\underline{c} := \frac{\delta_r}{\delta_r + \delta_\ell} \overline{c},$$

where $\delta_a = |u_a^R - u_a^L|$ denotes the payoff difference across the two states given the DM's action $a \in \{\ell, r\}$. As can be easily seen, $\underline{c}$ is strictly smaller than $\overline{c}$, i.e. the Bayesian threshold defined in Section 2. We will focus on the case in which the Bayesian value function Φ^* attains its minimum at $p_* < p_r^B$ such that $\Phi'(p_*) = 0$, as graphed in Figure 3. This case, labeled **Case 1**, occurs when the value of outright action r is not too high. The opposite case, labeled **Case**

2, in which $p_* = p_r^B$, or $\Phi'(p_*) < 0$ (as graphed in Figure 7) will be treated at the end of this section.

Theorem 1. *In Case 1, the following intrapersonal equilibrium exists.*

(i) *Suppose $c \geq \bar{c}$. The DM does not acquire information. If $\overline{p} \leq \hat{p}$, she takes action ℓ; if $\underline{p}(\overline{p}) \geq \hat{p}$, she takes action r; if $\hat{p} \in (\underline{p}(\overline{p}), \overline{p})$, she randomizes between ℓ and r with probability $\hat{\rho}$.*

(ii) *Suppose $c < \bar{c}$ and $\Phi'(p_*) = 0$. For each $c \in (0, \bar{c})$ there is a threshold $\Delta_c \in (0, +\infty]$ with $\Delta_c = +\infty$ if and only if $c \leq \underline{c}$ such that the intrapersonal equilibrium is described as follows:*

(a) If $\Delta \leq \Delta_c$, there exist cutoffs $0 < \overline{p}_1 < \overline{p}_2 \leq \overline{p}_3 < \overline{p}_4 < 1$ with $\overline{p}_1 = p_\ell^B$ and $\overline{p}_2 = p_*$ such that[8]

$$(m(\overline{p}), \nu(\overline{p}), \rho(\overline{p})) = \begin{cases} (1,0,0) \\ (0,0,\cdot) \\ (0,\nu^*(\overline{p}),0) \\ (0,0,\cdot) \\ (1,0,1) \end{cases} \qquad \pi(\overline{p}) = \begin{cases} \overline{p} & \textit{if } \overline{p} \in [0, \overline{p}_1] \\ \overline{p} & \textit{if } \overline{p} \in (\overline{p}_1, \overline{p}_2] \\ \pi^*(\overline{p}) & \textit{if } \overline{p} \in (\overline{p}_2, \overline{p}_3) \\ \underline{p}(\overline{p}) & \textit{if } \overline{p} \in [\overline{p}_3, \overline{p}_4) \\ \underline{p}(\overline{p}) & \textit{if } \overline{p} \in [\overline{p}_4, 1], \end{cases}$$

where $\nu^*(\overline{p}) > 0$ and $\pi^*(\overline{p}) \in (\underline{p}(\overline{p}), \overline{p})$, for $\overline{p} \in (\overline{p}_2, \overline{p}_3)$, and $\overline{p}_2 < \overline{p}_3$ if and only if $\Phi(p^*) < U_\ell(\underline{p}(p^*))$.

(b) If $\Delta > \Delta_c$, the equilibrium has the same structure as in (a) except that

$$(m(\overline{p}), \nu(\overline{p}), \rho(\overline{p})) = (1, 0, \hat{\rho}) \quad \textit{and} \quad \pi(\overline{p}) = \hat{p},$$

for $\overline{p} \in [\overline{p}_3, \overline{p}_4)$.

We structure the discussion of Theorem 1 according to the level of experimentation costs.

[8] We leave ρ unspecified if $m = \nu = 0$.

High experimentation costs. If $c \geq \overline{c}$, the DM does not experiment, and her equilibrium strategy coincides with the commitment solution described in Proposition B.2. This is a consequence of the facts that 1) given $c \geq \overline{c}$, the DM's commitment solution involves no experimentation, as we have seen in Section 2, and 2) implementing this solution requires no commitment. The DM's value in this case is $V^{\sigma}(\pi(\overline{p}), \overline{p}) = \min_{p \in P(\overline{p})} U(p)$.

Low experimentation costs. Our main focus will lie on the case $c \leq \underline{c}$. Under this restriction, the equilibrium is always described by Theorem 1-(ii)(a) and can be illustrated as follows:

$$0 \underbrace{\text{———————}}_{\text{action } \ell} \overline{p}_1 \underbrace{\longleftarrow\longleftarrow\longleftarrow\longleftarrow}_{\text{pure exp.}} \overline{p}_2 \underbrace{\longleftarrow\longleftarrow\longleftarrow\longleftarrow\longleftarrow}_{\text{mixed stopping } (\ell)} \overline{p}_3 \underbrace{\longleftarrow\longleftarrow\longleftarrow\longleftarrow}_{\text{pure exp.}} \overline{p}_4 \underbrace{\text{———————}}_{\text{action } r} 1$$

When $\overline{p}$ is sufficiently extreme ($\overline{p} \leq \overline{p}_1$ or $\overline{p} \geq \overline{p}_4$), the DM takes an immediate action without gathering additional information. This is qualitatively similar to Bayesian DM, although the exact stopping boundaries are not the same. For the remaining states, however, the characterization is qualitatively different if Δ is sufficiently large. Unlike the Bayesian DM, there is a middle region of states, where the DM randomizes between experimentation and action ℓ at a positive rate. To derive the solution we leverage the fact that in the absence of breakthrough news, the set of beliefs drifts to the left. Hence, we explain the logic of the equilibrium strategy via backward induction, starting from a far left set of beliefs.

• **Region 1:** $\overline{p} \leq \overline{p}_1$. The first threshold $\overline{p}_1$ coincides with the Bayesian stopping boundary p^B_ℓ. If $\overline{p} \leq p^B_\ell$, then stopping followed by action ℓ is optimal for all beliefs in $[\underline{p}(\overline{p}), \overline{p}]$ and thus constitutes the equilibrium strategy for our ambiguous-averse DM. The worst-case belief under this strategy is $\overline{p}$.

• **Region 2:** $\overline{p} \in (\overline{p}_1, \overline{p}_2]$. In Region 2, the right-most belief $\overline{p}$ continues to be the worst-case belief for all $\overline{p} \in (\overline{p}_1, \overline{p}_2]$ and the DM follows the Bayesian optimal stopping rule with respect to $\overline{p}$. Since $\overline{p} < \overline{p}_1 = p^B_\ell$, this strategy entails experimentation until either a breakthrough occurs or the Bayesian update of $\overline{p}$ reaches p^B_ℓ. The right boundary $\overline{p}_2 := p_*$ of this region is the belief at which the Bayesian value function attains its minimum.

For each state $\overline{p}$, we can then evaluate the expected payoff of strategy σ at alternative beliefs

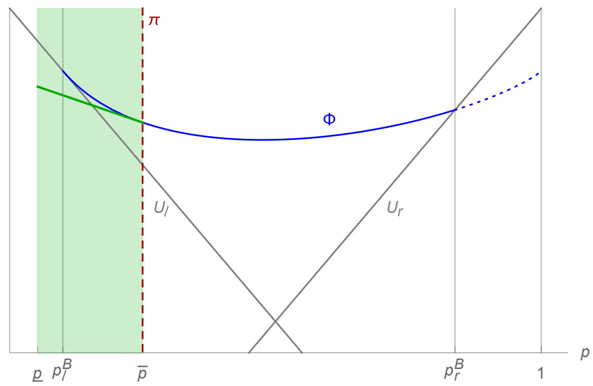

Figure 3: The value segment (in green) and worst-case belief for $\overline{p} \in (\overline{p}_1, \overline{p}_2)$.

p, denoting the value $V^\sigma(p, \overline{p})$ as a function of the belief p. The set of values corresponding to possible alternative beliefs, $\{(p, V^\sigma(p, \overline{p})) : p \in P(\overline{p})\}$, forms a linear segment, which we will call **value segment** throughout (see for instance Figure 3). It is useful to track the movement of the value segment, as time moves backward or forward. Note that the value segment is tangent to the value function Φ at $\overline{p}$, due to the fact that the stopping rule is Bayesian optimal with respect to $\overline{p}$.[9] Tangency at this point implies that the value segment is downward sloping and, hence, that $V^\sigma(p, \overline{p})$ is indeed minimized at $p = \overline{p}$. This means that nature optimally chooses the right most belief, justifying the earlier hypothesis. At $\overline{p} = p_* =: \overline{p}_2$, the value segment becomes flat (see Figure 4), as we assumed $\Phi'(p_*) = 0$. Hence, at $\overline{p} = p_*$ all posteriors in $[\underline{p}(\overline{p}), \overline{p}]$ minimize the DM's expected payoff.

• **Regions 3:** $\overline{p} \in (\overline{p}_2, \overline{p}_3]$**.** Moving further back in time, suppose the state $\overline{p}$ is slightly higher than p_* and the DM contemplates experimenting according to the strategy prescribed in Region 2. The value segment becomes now upward sloping, implying that the worst-case scenario is described by $\underline{p}(\overline{p})$ rather than $\overline{p}$. We then distinguish two cases. If the amount of ambiguity Δ is sufficiently small and, hence, the value segment is sufficiently "short," experimentation is optimal at the worst-case belief $\underline{p}(\overline{p})$ (see Figure 4, left panel).[10] The randomization region, Region 3, is empty here, and we directly transition to Region 4 at $\overline{p} = p_*$. More interestingly, if Δ is sufficiently large such that the value segment crosses the stopping payoff line U_ℓ from

[9] Bayesian optimality of the strategy with respect to $\overline{p}$ clearly implies $V^\sigma(\overline{p}, \overline{p}) = \Phi(\overline{p})$. This, together with linearity of $V^\sigma(p, \overline{p})$ in the first argument and the fact that $\Phi(p) \geq V^\sigma(p, \overline{p})$ holds for all $p \in (p_\ell^B, p_r^B)$, implies that the value segment is tangent to Φ at $\overline{p}$.

[10] Formally, we have $V^\sigma(\underline{p}(p_*), p_*) \geq U_\ell(\underline{p}(p_*))$.

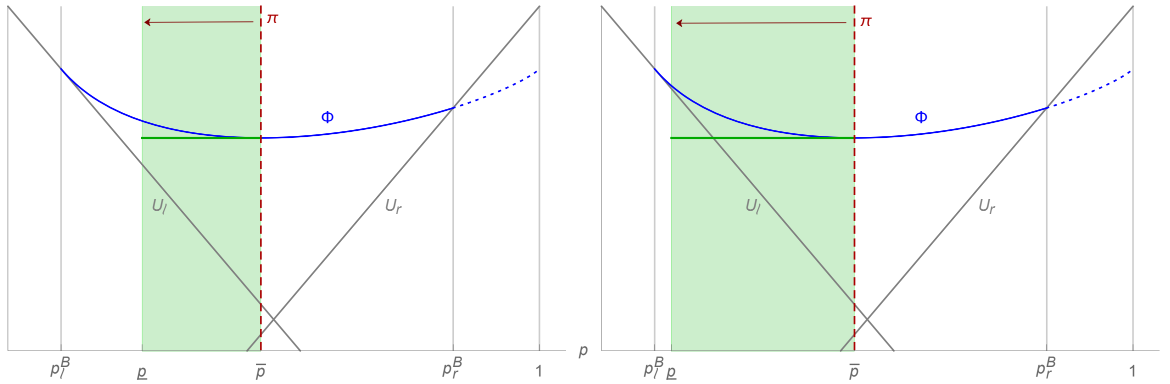

Figure 4: The value segment and worst-case belief for $\overline{p} = \overline{p}_2$ with $V^\sigma(\underline{p}(p_*), p_*) \geq U_\ell(\underline{p}(p_*))$ (left panel) and $V^\sigma(\underline{p}(p_*), p_*) < U_\ell(\underline{p}(p_*))$ (right panel).

below, then at state $\overline{p} = p_*$ experimentation is dominated by stopping followed by ℓ according to the worst-case belief $\underline{p}(p_*)$ (see Figure 5, right panel). Choosing ℓ, however, cannot be an equilibrium either, as nature would now choose $\overline{p}$, which makes it optimal for the DM to experiment. The intrapersonal equilibrium is thus in mixed strategies, where the DM randomizes between experimentation and action ℓ.

The DM's value and the equilibrium stopping rate in the mixed stopping region can be derived via the saddle-point HJB condition. Randomization between experimentation and action ℓ requires that the coefficient of ν in (11) must vanish and thus:

$$V(\pi(\overline{p}), \overline{p}) = U_\ell(\pi(\overline{p})). \tag{13}$$

Graphically, the stopping payoff U_ℓ and the value segment must cross at the worst-case belief $\pi(\overline{p})$, making the DM indifferent between experimentation and action ℓ. Since the required belief $\pi(\overline{p})$ will be in the interior of $[\underline{p}(\overline{p}), \overline{p}]$, for nature to optimally choose it, the value segment must remain flat so that all posteriors in $[\underline{p}(\overline{p}), \overline{p}]$ minimize the DM's expected payoff. We can thus write $V(p, \overline{p}) = \hat{V}(\overline{p}), \forall p$. Substituting this into the DM's indifference condition (13), we obtain nature's choice

$$\pi(\overline{p}) = \frac{u_\ell^L - \hat{V}(\overline{p})}{u_\ell^L - u_\ell^R}. \tag{14}$$

The randomization by the DM in turn implies that the derivative of the objective in (11) with

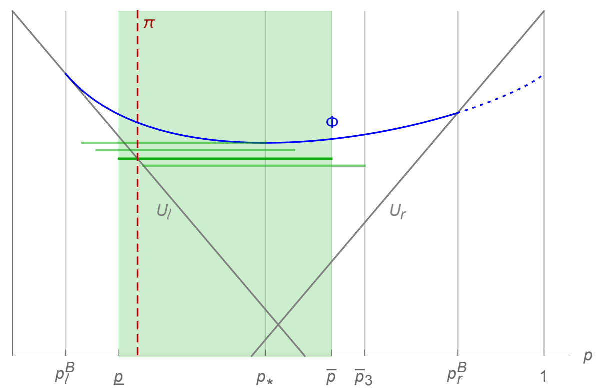

Figure 5: The value segment and worst-case belief for $\overline{p} \in (\overline{p}_2, \overline{p}_3)$.

respect to p must vanish. This fact, together with $V(p, \overline{p}) = \hat{V}(\overline{p}), \forall p$, yields the equilibrium stopping rate

$$\nu(\overline{p}) = \frac{\lambda(u_r^R - \hat{V}(\overline{p}))}{u_\ell^L - u_\ell^R}. \tag{15}$$

Using (14) and (15), the value function $\hat{V}(\overline{p})$ is derived by substituting for π and ν in the HJB condition (10), which gives rise to the following differential equation:

$$\overline{p}(1-\overline{p})\hat{V}'(\overline{p}) = \frac{\left(u_r^R - \hat{V}(\overline{p})\right)\left(u_\ell^L - \hat{V}(\overline{p})\right)}{u_\ell^L - u_\ell^R} - \frac{c}{\lambda}. \tag{$\widehat{\text{ODE}}$}$$

Together with the boundary condition $\hat{V}(\overline{p}_2) = \Phi(\overline{p}_2)$, ($\widehat{\text{ODE}}$) admits a unique solution, describing the DM's value in Region 3.

The reason that the DM stops according to a Poisson process, and thus with vanishing probability (rather than positive probability), is because when $\overline{p} \approx p_*$, the value segment is already arbitrarily close to flat, so it takes a vanishingly small stopping probability to turn the value segment completely flat. The same situation and intuition are repeated as we move backward in time: for a positive range of states, labeled Region 3, the DM repeatedly hedges between experimentation and Poisson stopping at a positive rate $\nu(\overline{p})$.

Intuitively, the repeated hedging by the DM can be seen as a resolution of conflicts between current and future selves. On the one hand, the DM is highly uncertain about the state and thus finds it optimal to acquire information. On the other hand, the DM is aware that the journey

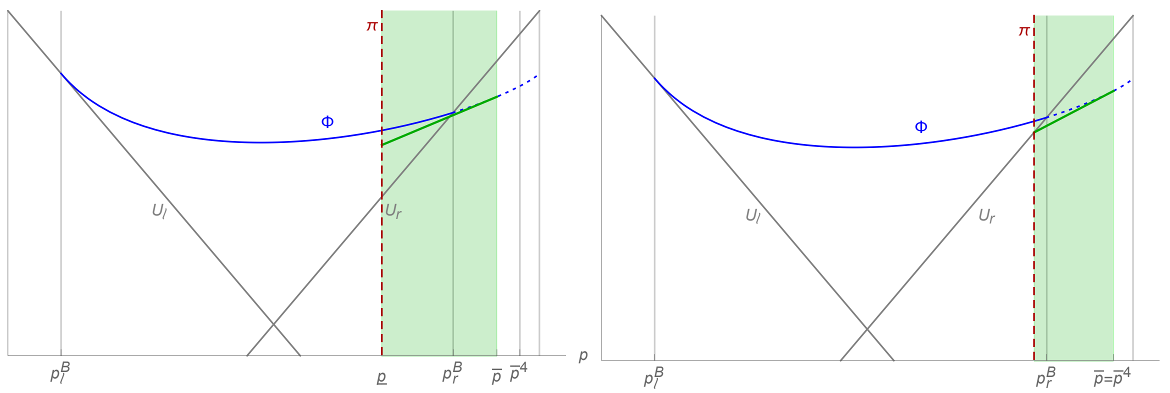

Figure 6: The value segment and worst-case belief for $\overline{p} \in (\overline{p}_3, \overline{p}_4)$ in the case of Δ small.

of experimentation will take on its own life, with the future selves doing too much of it from her current perspective. The randomized stopping resolves the conflict and restores the delicate balance between the multiple selves. While at first glance it may seem odd that the randomized stopping arises in the middle region around p_*, this is no coincidence. As noted earlier, it is precisely where preference reversal and dynamic inconsistencies arise. As will be seen, this seemingly peculiar feature of the equilibrium proves quite robust across different specifications of the learning environment and technologies.

• **Regions 4:** $\overline{p} \in (\overline{p}_3, \overline{p}_4]$**.** As already mentioned, when ambiguity Δ is small, Region 3 is empty, and Regions 2 and 4 collapse to a single experimentation region. Its right boundary, $\overline{p}_4$, occurs at the point where, according to the worst-case belief $\underline{p}(\overline{p})$, action r yields the same expected payoff as experimentation, i.e. $U_r(\underline{p}(\overline{p}_4)) = V^\sigma(\underline{p}(\overline{p}_4), \overline{p}_4)$. Graphically, $\overline{p}_4$ can be seen as the point where the left endpoint of the value segment crosses the stopping payoff U_r (see right panel of Figure 6). It is important to notice that at $\overline{p} = \overline{p}_4$, the value of $\underline{p}(\overline{p})$ lies strictly to the left of the Bayesian stopping boundary p_r^B. This follows from the fact that the DM's future selves follow a rule that is optimal with respect to $\overline{p}$ but not with respect to $\underline{p}(\overline{p})$. From the perspective of $\underline{p}(\overline{p})$, the value of experimentation is thus strictly lower than the Bayesian value $\Phi(\underline{p}(\overline{p}))$. This means that the DM with $\underline{p}(\overline{p})$ is more inclined to stop than the Bayesian DM with the corresponding belief.

Next consider the case in which Δ is sufficiently large so that Region 3 with mixed stopping is nonempty. The right boundary of the mixed stopping region, $\overline{p}_3$, is given by the state $\overline{p}$ at which the value segment separates from U_ℓ and experimentation becomes again optimal for all posteriors in $[\underline{p}(\overline{p}), \overline{p}]$ (see Figure 5). To the right of $\overline{p}_3$, we then have a second region of pure experimentation, Region 4. The value segment becomes upward sloping in this region, so $\underline{p}(\overline{p})$ becomes the worst-case belief. As with the case of small Δ, the right boundary of Region 4 is determined as the point where the left end of the value segment reaches the stopping payoff U_r. Notice, however, that the value segment in the experimentation region remains below the Bayesian value function. Compared to the case of small Δ, the possibility of randomization in future states reduces the value of experimentation further, which pushes the stopping boundary $\overline{p}_4$ towards the middle. Intuitively, the DM here is discouraged from experimentation, and is thus more inclined to stop, not only because it leads to the excessive experimentation in the distant future, but also because it leads to the inefficient stopping in the intermediate future.

We now complete the characterization by addressing several remaining cases and issues. The reader interested in the behavioral implications of the baseline characterization may skip to the next section.

Intermediate experimentation costs. Consider next the case in which $c \in (\underline{c}, \overline{c})$. Theorem 1 shows that for each $c \in (\underline{c}, \overline{c})$ there exists a threshold $\Delta_c \in \mathbb{R}$ such that for $\Delta \leq \Delta^c$ the equilibrium continues to be described by case (a) of Theorem 1, as described above. If instead $\Delta > \Delta_c$, the fourth region features randomization between actions ℓ and r rather than experimentation.

$$0 \underbrace{\rule{4em}{0.4pt}}_{\text{action } \ell} \overline{p}_1 \underbrace{\longleftarrow\!\!\longleftarrow\!\!\longleftarrow\!\!\longleftarrow}_{\text{pure exp.}} \overline{p}_2 \underbrace{\longleftarrow\!\!\longleftarrow\!\!\longleftarrow\!\!\longleftarrow\!\!\longleftarrow}_{\text{mixed stopping } (\ell)} \overline{p}_3 \quad \underbrace{\rule{5em}{0.4pt}}_{\text{mixing between } \ell \text{ and } r} \quad \overline{p}_4 \underbrace{\rule{4em}{0.4pt}}_{\text{action } r} 1$$

The equilibrium behavior for the first three regions is the same as in the case with low experimentation costs. In the third region, however, where the DM randomizes between experimentation and action ℓ, the value segment may drop as low as the point at which the two stopping payoff lines, U_ℓ and U_r, cross. The state at which this occurs constitutes our new boundary $\overline{p}_3$. The DM's value at this point is equal to $\min_p U(p) = U(\hat{p})$, the minimal expected payoff that the DM can guarantee to herself by randomizing between actions ℓ and r with probabilities $1 - \hat{\rho}$ and $\hat{\rho}$, respectively. At states to the right of this point, randomizing between immediate actions with $\hat{\rho}$ strictly dominates experimentation with randomized stopping, giving rise to the fourth region $[\overline{p}_3, \overline{p}_4)$. Its right boundary is the state $\overline{p}$ that satisfies $\underline{p}(\overline{p}) = \hat{p}$. Further to the right, where $\underline{p}(\overline{p}) > \hat{p}$, the DM optimally takes action r. The solution for intermediate costs $c \in (\underline{c}, \overline{c})$ and large ambiguity $\Delta > \Delta_c$ can thus be viewed as a combination of the previous two solutions.

Case 2: $\Phi'(p_*) < 0$**.** The formal characterization of the intrapersonal equilibrium for the case $\Phi'(p_*) < 0$ is relegated to Appendix B.7. Compared with the case $\Phi'(p_*) = 0$, there is only one significant change. At the right boundary of the first experimentation region $(\overline{p}_1, \overline{p}_2)$, the stopping distribution has an atom: the DM takes action r with positive probability $m(\overline{p}_2)$ and experiments with the complementary probability.

$$0\,\underbrace{\rule{2cm}{0.4pt}}_{\text{action } \ell}\,\overline{p}_1\,\underbrace{\longleftarrow\longleftarrow\longleftarrow\longleftarrow}_{\text{pure exp.}}\,\underbrace{\overline{p}_2}_{m>0}\,\underbrace{\longleftarrow\longleftarrow\longleftarrow\longleftarrow}_{\text{mixed stopping } (\ell)}\,\overline{p}_3\,\underbrace{\longleftarrow\longleftarrow\longleftarrow\longleftarrow\longleftarrow}_{\text{pure exp. or mixing } \ell/r}\,\overline{p}_4\,\underbrace{\rule{2cm}{0.4pt}}_{\text{action } r}\,1$$

Moving backwards in time through the experimentation region $(\overline{p}_1, \overline{p}_2)$, the value segment remains downward sloping as the state $\overline{p}$ approaches $\overline{p}_2 = p_*$. The unique worst-case belief in the limit as $\overline{p} \nearrow p_*$ is thus $\overline{p}$. For states to the right of p_*, the Bayesian optimal strategy under $\overline{p}$ is to take action r rather than experimenting. Taking action r in those states does, however, not constitute an intrapersonal equilibrium, as the worst-case belief under this strategy would be $\underline{p}(\overline{p})$ rather than $\overline{p}$. The DM must therefore experiment with positive probability in states to the right of p_*. To make experimentation optimal, the value segment must be weakly upward sloping for states belonging to an interval $(p_*, p_* + \varepsilon)$. This property requires the DM to take action r with strictly positive probability at state $\overline{p} = p_*$. Graphically, the DM's randomization flips the left end of the value segment downwards, so that it is flat at $\overline{p} = p_*$, as illustrated in Figure 7.

Uniqueness. In Appendix B.8, we prove that the intrapersonal equilibrium is unique when either $c < \underline{c}$ or $c \geq \overline{c}$. When $c \in [\underline{c}, \overline{c})$, however, there is an additional class of intrapersonal equilibria. Fix any $q \in [\hat{p}, \underline{p}^{-1}(\hat{p}))$. We then construct a strategy profile $(\tilde{\sigma}, \tilde{\pi})$ indexed by q as follows. For each $\overline{p} \leq q$, $(\tilde{\sigma}, \tilde{\pi}) = (\sigma, \pi)$, the intrapersonal equilibrium defined in Theorem 1; for each $\overline{p} \in (q, \underline{p}^{-1}(\hat{p})]$, $\tilde{\sigma}(\overline{p}) = (\tilde{m}(\overline{p}), \tilde{\nu}(\overline{p}), \rho(\overline{p})) = (1, 0, \hat{\rho})$ and $\tilde{\pi}(\overline{p}) = \hat{p}$; and for each $\overline{p} > \underline{p}^{-1}(\hat{p})$, $\tilde{\sigma}(\overline{p}) = (\tilde{m}(\overline{p}), \tilde{\nu}(\overline{p}), \rho(\overline{p})) = (1, 0, 1)$ and $\tilde{\pi}(\overline{p}) = \underline{p}(\overline{p})$. In words, the DM follows the same strategy as in Theorem 1 for states $\overline{p} \leq q$ but stops immediately for states $\overline{p} > q$, followed by the "hedging" mixed action $\hat{\rho}$ if $\hat{p} \in [\underline{p}(\overline{p}), \overline{p}]$ and action r if $\hat{p} < \underline{p}(\overline{p})$. Since $c < \overline{c}$, this

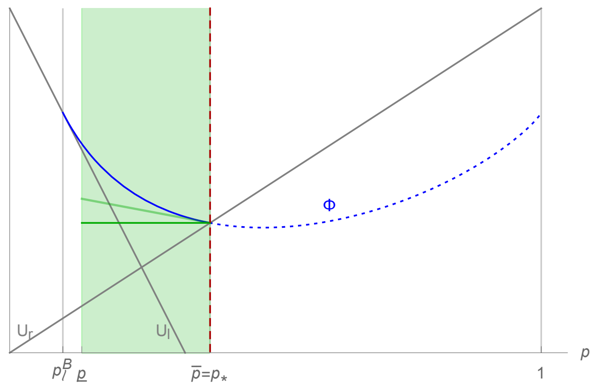

Figure 7: Case $\Phi'(p_*) < 0$. The value segment for $\overline{p} \nearrow p_*$ and $\overline{p} = p_*$.

new strategy is worse for the DM *regardless of the belief* whenever $\overline{p} > q$. In fact, the value function jumps down at the "switching" state $\overline{p} = q$. Nevertheless, the strategy profile—there is a continuum of such indexed by $q \in [\hat{p}, \underline{p}^{-1}(\hat{p}))$—satisfies all requirements of our solution concept, including the HJB conditions. The intuition is that, even though the HJB conditions capture the control over vanishingly proximate future actions, these conditions are not strong enough, since there always exists some $\epsilon < \overline{p} - q$ such that deviating only until the state falls to $\overline{p} - \epsilon$ will not generate a strictly higher payoff.

Nevertheless, we view these equilibria as implausible. First of all, we see that $\tilde{\sigma}$, regardless of q, is Pareto worse than σ for all selves of the DM evaluated at the worst-case beliefs. Hence, $\tilde{\sigma}$ will be eliminated if the multi-selves of the DM can coordinate on the selection of equilibria, as is often invoked in the name of "consistent planning;" see Siniscalchi (2011).[11] Second, we could strengthen the notion of vanishing commitment to eliminate the equilibria. Namely, instead of requiring the equilibrium strategy profile to be a limit of the ϵ-commitment solution *pointwise* for each $\overline{p}$, one could require the limit convergence to be *uniform* across all $\overline{p}$'s. The profile $(\tilde{\sigma}, \tilde{\pi})$ cannot survive this stronger test, since for any $\epsilon > 0$, there will always exist a state $\overline{p} \in (q, q+\epsilon)$, such that the ϵ-commitment solution for the DM at that state will be to deviate to σ and enjoy a discontinuous payoff jump.[12]

[11] More precisely, Siniscalchi (2011) requires the current self to break a tie in favor of her earlier self. See Ebert and Strack (2018) for a related approach.

[12] One could adopt this stronger requirement as a solution concept. We do not take this approach for two reasons. First, the uniform convergence requirement is not easy to check since it requires one to verify the equi-continuity property of the value function. Second, the requirement jeopardizes the existence in the more general Poisson model we discuss in Appendix C.1 and Appendix C.1. The unique intrapersonal equilibrium presented in Theorem 4 involves a discontinuous value function when the cost is intermediate, which does not

Knightian Uncertainty. The extreme case in which $\underline{p} = 0$ and $\overline{p} = 1$, known as Knightian uncertainty, constitutes an important theoretical benchmark. Naturally, the equilibrium characterization for this case is obtained as the limit of the solution we characterized above for $(\underline{p}_0, \overline{p}_0) \to (0, 1)$.

Proposition 1. *Assume $P_0 = [0, 1]$. If $c \geq \underline{c}$, the DM mixes between immediate actions ℓ and r with probability $\hat{\rho}$. Otherwise the DM mixes between experimentation and stopping followed by ℓ at a stationary rate*

$$\tilde{\nu} = \frac{\lambda}{\delta_\ell}(u_r^R - \tilde{u}), \text{ where } \tilde{u} = \frac{u_r^R + u_\ell^L}{2} - \sqrt{\left(\frac{u_r^R - u_\ell^L}{2}\right)^2 + \frac{c}{\lambda}\delta_\ell}.$$

This solution is the limit of the equilibrium strategy characterized in Theorem 1 (Theorem 3 for Case 2) for the case $0 < \underline{p}_0 < \overline{p}_0 < 1$ as $(\underline{p}_0, \overline{p}_0) \to (0, 1)$.

As $(\underline{p}_0, \overline{p}_0) \to (0, 1)$, Region 3 featuring randomized stopping takes over the entire region to the right of p_*. Starting from $\overline{p}_0$ close to one, the state $\overline{p}$ will then remain in Region 3 for an arbitrarily large amount of time. In the extreme case, when $[\underline{p}_0, \overline{p}_0] = [0, 1]$, the DM's set of priors remains unchanged in the absence of breakthrough news. Hence, under any Markovian strategy the DM either stops immediately or stops at a constant rate. The latter is optimal if and only if $c \leq \underline{c}$.

We can interpret the Knightian solution as a form of hedging by the DM to ensure her payoff equalizes between the states. Indeed, it is easy to verify that the stationary stopping rate $\tilde{\nu}$ satisfies the following equality:

$$u_\ell^L - \frac{c}{\tilde{\nu}} = \frac{\lambda}{\tilde{\nu} + \lambda} u_r^R + \frac{\tilde{\nu}}{\tilde{\nu} + \lambda} u_\ell^R - \frac{c}{\tilde{\nu} + \lambda}. \tag{16}$$

The LHS shows the DM's payoff from stopping at a constant rate $\tilde{\nu}$ when the state is L. In state L the DM never observes a breakthrough and thus always ends up taking the correct action. The downside is a relatively high expected cost of experimentation equal to $c/\tilde{\nu}$. The RHS shows the DM's payoff when the state is R. Now, the DM receives R-evidence before she stops experimenting, with positive probability $\lambda/(\tilde{\nu} + \lambda)$. With the complimentary probability, however, the DM takes the wrong action and receives the lower payoff u_ℓ^R. The payoff gross of

survive the stronger refinement.

experimentation costs is then lower in state R than in state L, which is compensated by the DM experimenting for shorter periods (in expectation) and therefore incurring a lower expected cost of experimentation equal to $c/(\tilde{\nu} + \lambda)$.

The role of randomization. Whether randomization helps to hedge against ambiguity depends on the DM's internal view on how the uncertainty unfolds (see Saito (2015) and Ke and Zhang (2020)). We adopt the assumption that the DM views the game against nature as one with simultaneous moves. Hence, in the DM's mind, nature cannot condition her choice on the outcome of the DM's randomization. If the DM takes on an even more pessimistic view, believing that nature can adjust her choice *after* the DM's randomization outcome, any realization of a mixed strategy will be evaluated according to the worst-case belief for this realization, so the DM has no incentives to randomize. Modifying the setting in this way clearly changes the characterization of our solution for the case when ambiguity is large. It does, however, not erase the DM's desire to stop preemptively in anticipation of a long experimentation phase when ambiguity is large, as we show next. Stopping occurs again at intermediate sets of beliefs where the worst-case belief is about to switch. Anticipating this, earlier selves of the DM are happy to experiment until that point. Rather than an interval with randomized stopping, we thus obtain an isolated stopping point. To keep the proof short, we restrict attention to the symmetric case where $u_r^R = u_\ell^L =: \delta$ and $u_\ell^R = u_r^L = 0$.

Proposition 2. *Suppose the DM is restricted to pure strategies. Assume $\Phi'(p_*) = 0$, $c < \bar{c}$ and payoffs are symmetric. If Δ is sufficiently large, then there is a state $\overline{p} \in (0, 1)$ such that the DM takes action ℓ at $\overline{p}$ and experiments on a left and right neighborhood of $\overline{p}$.*

4 Behavioral Implications of Ambiguity

In this section, we perform comparative statics to further explore the behavioral implications of ambiguity.

Prolonged Learning and Indecision. We first investigate the DM's incentive to prolong her experimentation late in the game and how it varies with ambiguity. For this latter purpose, it is more convenient to use the mid-point belief θ in log-likelihood ratio as the state; namely, θ is chosen so that $z(\theta) = \frac{z(\underline{p})+z(\overline{p})}{2}$, where $z(p) := \ln(\frac{p}{1-p})$. We then represent the DM's prior set

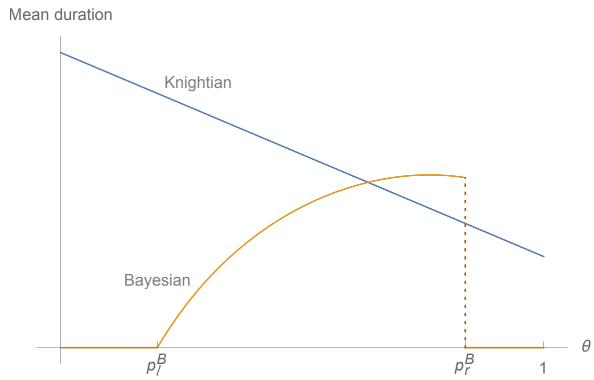

Figure 8: Mean sampling time for the Bayesian case and the case of Knightian uncertainty.

of beliefs as $P(\theta) = [\underline{p}(\theta), \overline{p}(\theta)]$, where $\underline{p}(\theta) < \theta < \overline{p}(\theta)$. One interpretation is that θ is the *true* probability that $\omega = R$. Alternatively, θ may be the "subjective" belief of the corresponding Bayesian DM.

We are interested in studying how DMs with different degrees of ambiguity respond to a change in state θ. Naturally, an increased ambiguity is represented by the shift of the prior set from $P(\theta)$ to $Q(\theta) = [\underline{q}(\theta), \overline{q}(\theta)]$, such that $\underline{q}(\theta) < \underline{p}(\theta)$ and $\overline{p}(\theta) < \overline{q}(\theta)$.

Let $T_P(\theta)$ and $T_Q(\theta)$ respectively denote the expected lengths of experimentation time under ambiguity P and Q. We have the following result.

Proposition 3 (Length of experimentation)**.** *Suppose $c < \overline{c}$ so that experimentation occurs under some θ. There exists $\hat{\theta} \in (0, \overline{\theta}]$, where $\overline{\theta} := \sup\{\theta : T_P(\theta) > 0\}$, such that $T_Q(\theta) \geq T_P(\theta)$ if $\theta < \hat{\theta}$ and $T_Q(\theta) \leq T_P(\theta)$ if $\theta \in [\hat{\theta}, \overline{\theta})$. Further, $T'_Q(\theta) < T'_P(\theta)$ for $\theta < \hat{\theta}$.*

In words, an increased ambiguity leads the DM to prolong her experimentation when the likelihood θ of state R is not too large. For a benchmark, consider the Bayesian DM with $p_0 = \theta$. The length of experimentation for the Bayesian DM is concave and strictly increasing for θ not too large,[13] as illustrated in Figure 8. There are two competing forces. As θ increases, on the one hand, experimentation is more likely to produce conclusive R-evidence; this tends to shorten experimentation. On the other hand, the DM adjusts her learning strategy to lengthen its duration as θ rises. For θ not too large, the latter effect dominates the former effect, so the overall learning time increases as R becomes more likely.

The single crossing property stated in Proposition 3 captures the sense in which ambiguity

[13] See Proposition 1 of Che and Mierendorff (2019).

blunts the responsiveness with which the DM adjusts her strategy to a change in θ. The reduced responsiveness to θ corresponds to the behavioral trait of "indecisiveness." In the extreme case as $Q \to [0,1]$, the learning strategy becomes completely inelastic or unresponsive to the change in the reference belief θ. In this situation, only the first effect plays a role, so that the learning time actually decreases as the likelihood θ of R actually rises (see Figure 8).

Premature Decisiveness. While the Bayesian optimal stopping time is deterministic conditional on not receiving R-evidence, a sufficiently large ambiguity leads to randomized stopping at earlier points in time. Therefore, the ambiguity-averse DM exhibits early "decisiveness" not shown by her Bayesian counterpart. At the same time, an increased ambiguity also leads to prolonged experimentation, as seen in Proposition 3, which increases the likelihood of later stopping. The combined effect is the increased dispersion of the stopping times. To formalize this result, consider again two prior sets P and Q corresponding to two different ambiguity levels with $P \subset Q$. Let $F_P(t)$ denote the probability that the DM stops by time t when the initial set of priors is P.

Proposition 4 (Dispersion of stopping times.)**.** *Assume $c \leq \underline{c}$ and $\Phi'(p_*) \geq 0$. Let P and Q be such that $P \subseteq Q$. There exists a time $\hat{t}$ such that $F_Q(t) \geq F_P(t)$ for all $t < \hat{t}$ and $F_Q(t) \leq F_P(t)$ for all $t > \hat{t}$.*

Proposition 4 shows that the stopping time distribution under the two nested sets P and Q exhibits a single crossing property. This property indicates that the distribution over stopping times under the larger set Q is more dispersed than the one under P. Indeed, whenever the mean stopping time under P is weakly greater than under Q, then F_P second-order stochastically dominates F_Q (when the means are equal, F_Q constitutes a mean-preserving spread of F_P). The single crossing property implies that the range of potential stopping points spreads as ambiguity becomes larger: an expansion of the initial set of priors leads to an earlier first potential stopping point and a later last potential stopping point. The property is illustrated in Figure 9, where we plot the cumulative probability of stopping conditional on state L for the Bayesian DM (orange graph) and the ambiguity-averse DM (blue graph).

Can the two traits coexist? According to our theory, ambiguity-aversion leaves potentially testable behavioral markers—prolonged indecision and impulsive decisiveness. While seemingly contradictory, there is a coherent logic that ties these two traits together. For illustration,

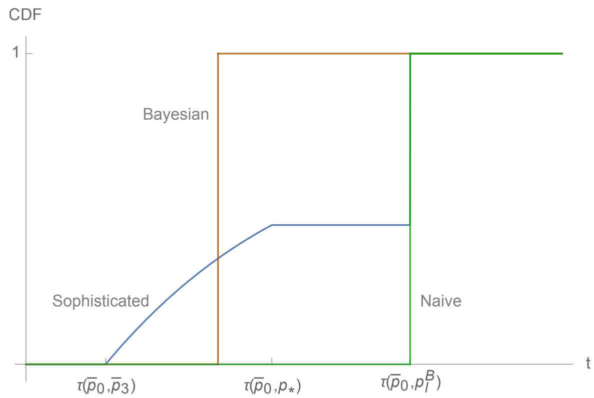

Figure 9: Cumulative distribution of stopping times (conditional on $\omega = L$) for the Bayesian DM and the ambiguity-averse DM when sophisticated (in blue) and naive (in green). In the graph, $\tau(p, q)$ denotes the time it takes for belief p to be updated to q in the absence of R-evidence.

consider again the theorist seeking to prove a new theorem with unknown validity. According to our theory, ambiguity—with regard to the unknown validity of the theorem—makes it difficult for the theorist to dismiss the validity of the theorem even after many unsuccessful attempts to prove it. Hence, ambiguity might lead the DM to a costly process of continued trial that could be seen by the outsider as a "wilde goose chase". Recognizing that such an obsessive pursuit is in the offing—perhaps also reminded by how such endeavors ended in the past—the DM may simply refrain from getting into the situation by stopping early; the theorist may prematurely give up after making a few unsuccessful attempts at proving the theorem.

While the empirical plausibility of the connection between these two traits remains unclear, there is some evidence of their coexistence. Barkley-Levenson and Fox (2016) find indecision and impulsivity to be positively correlated among the respondents of their behavioral survey, and reach an interpretation uncannily close to our theory:

> Our findings suggest ... indecisiveness and impulsivity can be viewed as two sides of the same coin. Both indecisiveness and impulsivity are maladaptive behavioral responses to difficulty engaging with a decision. We surmise that these responses arise from a common desire to avoid negative affect that some individuals experience when making choices. ... they may behave in an impulsive manner, hastily choosing in order to avoid unpleasant deliberation over opportunity costs. (Barkley-Levenson and Fox, 2016)

The role of sophistication. While the feature of prolonged learning as a response to ambiguity does not hinge on the DM's sophistication, premature decisiveness is driven by the DM's ability to correctly forecast her future behavior and respond accordingly. Indeed, a fully naive DM—one who acts according to the best plan given by the current worst-case scenario but fails to carry out the plans of earlier selves—only exhibits the first behavioral trait and will thus always experiment longer than the corresponding Bayesian DM. The sophisticated DM, on the other hand, will engage in preemptive stopping when ambiguity is sufficiently large, so she will tend to stop earlier than her naive counterpart. Note, however, that both DMs share the same final stopping points, so with positive probability both stop at the same time. We illustrate this feature in Figure 9: the CDF for the naive DM first-order stochastically dominates that of the sophisticated DM.

It is also interesting to contrast the sophisticated and naive DMs with a DM who has full commitment power. As we saw in Section 2, the commitment solution is given by the Bayesian optimal stopping rule for the prior in $[\underline{p}_0, \overline{p}_0]$ that minimizes the Bayesian value function Φ^*. Assuming $\overline{p}_0 > p_*$, the graph illustrating the commitment solution is thus qualitatively similar to that of the Bayesian DM in Figure 9. Specifically, a committed DM will stop before a naive one. The comparison with a sophisticated DM is ambiguous, except that the latter's stopping time is more dispersed in the sense established in Proposition 2.

5 Extensions

Our results are robust to different learning environments. Here, we extend the baseline model to consider two canonical ones. Our discussion will be largely informal, with the detailed analysis relegated to Appendix C. We also assume throughout that the stopping payoffs are symmetric: $u^R_r = u^L_\ell = \delta > 0$ and $u^R_\ell = u^L_r = 0$.

5.1 General Poisson Learning

One could consider a general Poisson model introduced by Che and Mierendorff (2019): at each instant, the DM may seek either R-evidence (as before) or L-evidence. One interpretation is that there are two news sources emitting the two types of evidence. The DM may then allocate a share $\alpha \in [0, 1]$ of her attention to the R-evidence news source and a share $1 - \alpha$ to the L-evidence news source, and receive the evidence proportionately at rates $\alpha\lambda$ in state R and at rates $(1-\alpha)\lambda$ in state L. For instance, a theorist may try either to "prove" a theorem (R-evidence), to find a "counter-example" disproving it (L-evidence), or to divide effort between the two endeavors.

As before, the Bayesian optimal strategy is characterized by two stopping boundaries, but as shown by Che and Mierendorff (2019), the optimal learning strategy becomes nontrivial. Figure 10 summarizes the strategy for two different ranges of learning costs.

$$\underset{p=0}{|}\underbrace{\text{————————}}_{\text{action } \ell}\; p_\ell^B \underbrace{\longleftarrow\longleftarrow\longleftarrow\longleftarrow\longleftarrow}_{\alpha=1} \check{p} \underbrace{\longrightarrow\longrightarrow\longrightarrow\longrightarrow\longrightarrow}_{\alpha=0} p_r^B\; \underbrace{\text{————————}}_{\text{action } r}\underset{1}{|}$$

(a) intermediate cost: $c \in [\underline{c}^*, \overline{c})$, where $\underline{c}^* := \frac{\lambda\delta}{1+e^2}$.

$$\underset{p=0}{|}\underbrace{\text{——————}}_{\text{action } \ell} p_\ell^B \underbrace{\longleftarrow\longleftarrow\longleftarrow}_{\alpha=1} p_L^B \underbrace{\longrightarrow\longrightarrow}_{\alpha=0} \underbrace{p^*}_{\alpha=\frac{1}{2}} \underbrace{\longleftarrow\longleftarrow}_{\alpha=1} p_R^B \underbrace{\longrightarrow\longrightarrow\longrightarrow}_{\alpha=0} p_r^B \underbrace{\text{——————}}_{\text{action } r}\underset{1}{|}$$

(b) low cost: $c < \underline{c}^*$.

Figure 10: Bayesian optimal strategy.

The (Bayesian) DM adopts the two types of Poisson signals for different beliefs p. If the belief is close to a stopping boundary or if c is not too low, the DM looks for a *contradictory evidence*—i.e., L-evidence when p is relatively high and R-evidence when p is relatively low. Intuitively, the DM is seeking to "rule out" the less likely state. When such evidence does not arrive, the belief drifts outward, toward a stopping boundary. Upon reaching that boundary belief, the DM stops and chooses the action matching the likely state.

If the cost is low and p is in a middle range, the DM seeks *confirmatory evidence*—i.e., R-evidence if p is relatively large and L-evidence if p is relatively small. When such evidence does not arrive, the DM becomes more uncertain, and her belief drifts inwardly toward $p^* = 1/2$. Once $p^* = 1/2$ is reached, the DM splits her attention between the two news sources with $\alpha = 1/2$. Her belief then never moves, and learning continues indefinitely until either evidence is obtained. Intuitively, at that belief, the DM finds both types of experimentation equally tempting and acts like a "Buridan's donkey," unable to drift away from the belief that causes the dilemma.

Suppose now the DM is ambiguity averse, endowed with an interval of priors. How would the presence of the additional news source affect her behavior? Several patterns of behavior established from the baseline model continue to hold. First, close to a stopping boundary,

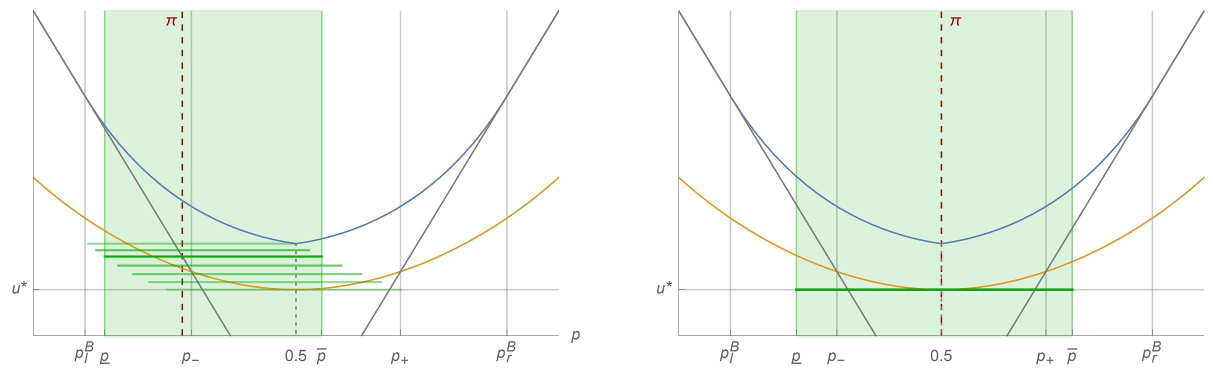

Figure 11: The case of intermediate c and large Δ.

the "inner-most" belief becomes the worst-case belief, and the DM learns—which she does by seeking contradictory evidence much like the Bayesian DM—until that belief reaches the Bayesian stopping boundary. Hence, the DM learns excessively compared with the Bayesian counterpart.

Next, when the cost is not too small (its precise range being specified in Appendix C.2), the DM may also engage in randomized stopping much like in the baseline model. As before, this happens when the ambiguity is sufficiently large so that the set of priors contains $p = 1/2$ (corresponding to the minimizer of the Bayesian value function), and the associated value segment crosses the stopping payoff associated with either action ℓ or r. Specifically, when the prior set is toward the left side, the DM randomizes between R-evidence seeking and stopping with action ℓ (as before); when the prior set is toward the right side, the randomization is between L-evidence seeking and stopping with action r. See the former case depicted in Figure 11 (left panel).

A new feature also emerges. The DM now adopts split-attention learning for a range of prior sets. While this strategy is also employed by the Bayesian DM, it becomes more valuable as a hedging device since it allows the DM to collect both types of evidence. An ambiguity-averse DM may thus use the split-attention strategy even when the Bayesian DM would not. For instance, Figure 11 depicts the case of intermediate costs, where the (Bayesian) value of seeking confirmatory evidence followed by split attention (the orange curve) is strictly below the value of seeking contradictory evidence (the blue curve). In this case, it is never optimal for a Bayesian DM to employ split-attention learning. Yet, as depicted in the right panel of Figure 11, the ambiguity-averse DM adopts this strategy when ambiguity is sufficiently large and the prior set is sufficiently centered.[14] In this situation, split-attention learning supplants the randomized stopping strategy as a superior hedging device.

In sum, the equilibrium exhibits the behavioral patterns observed in the baseline model, such as *excessive learning* and *premature stopping*, but it also features a new strategy, namely split attention. Whenever the DM adopts the latter strategy, she continues learning until she fully resolves the uncertainty. This behavior thus provides another pattern of excessive learning, when it is seen from the perspective of a Bayesian observer with comparable beliefs.

5.2 Incremental Learning

So far, we have focused on Poisson learning. Alternatively, one could consider an incremental learning model in which the signal follows a diffusion process. Suppose the DM receives a signal $X \in \mathbb{R}$ about the state $\omega \in \{L, R\}$ at a flow cost of c per unit time, which follows the process:

$$dX_t = \mu_\omega dt + \sigma dB_t,$$

where $\mu_R > \mu_L$. As usual, it is more convenient to convert X into a belief, particularly into a log-likelihood ratio Z, where

$$dZ_t = \frac{\psi}{\sigma} dX_t - \frac{\psi}{\sigma}\left(\frac{\mu_R + \mu_L}{2}\right) dt = \pm \frac{\psi^2}{2} dt + \psi dB_t,$$

and $\psi := \frac{\mu_R - \mu_L}{\sigma}$ is the signal-to-noise ratio (see Daley and Green (2012, 2020) for example). Then, the Bayesian optimal behavior is characterized by two stopping boundaries (in terms of log-likelihood ratio), z_ℓ^B and z_r^B, such that the DM stops and chooses ℓ if $z \leq z_\ell^B$, she stops and chooses r if $z \geq z_r$, and she experiments if $z \in (z_\ell^B, z_r^B)$. Given the symmetry of payoffs $z_\ell^B < 0 < z_r^B = -z_\ell^B$, where $z = 0$ corresponds to the belief of $p = 1/2$. As before, the stopping boundaries are graphically depicted by tangent points of the Bayesian value function, denoted by Φ^B, at the stopping payoffs (which are nonlinear due to the change of variable): see Figure 12.

Consider now an ambiguity-averse DM whose prior set is represented by an interval $Y(z) := [z - \frac{\Delta}{2}, z + \frac{\Delta}{2}]$. In contrast to before, let us index the "state" by the midpoint z. Upon suitable modification, the intra-personal equilibrium is characterized much like in the baseline model.

[14] More specifically, this happen when the set of priors includes the interval $[p_-, p_+]$ such that $p_- < 1/2 < p_+$, where these parameters are characterized in Appendix C.1.

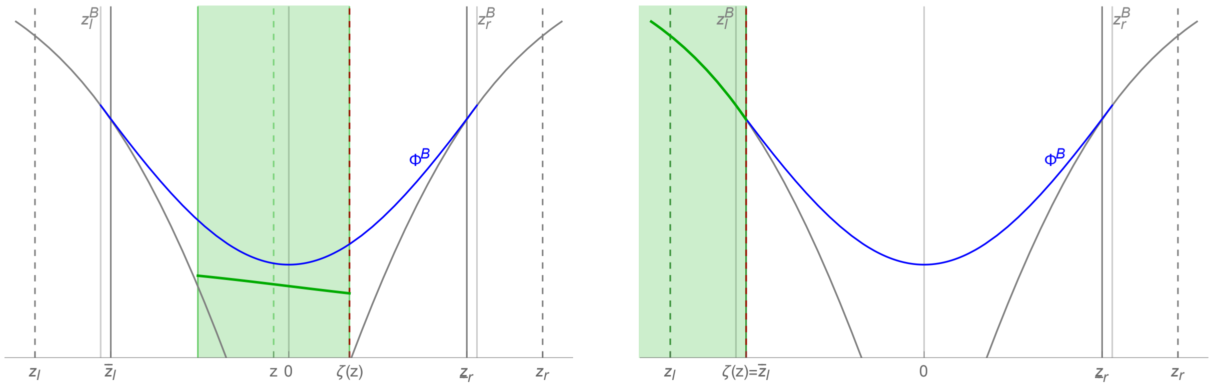

Figure 12: Agent's value when experimenting (left) and when stopping (right), with worst-case belief $\zeta(z)$ and boundaries $\overline{z}_\ell := z_\ell + \Delta/2$ and $\underline{z}_r := z_r - \Delta/2$.

If the ambiguity small, there are again two stopping boundaries $z_\ell < 0 < z_r$ (where $z_r = -z_\ell$ given the symmetry) such that the DM experiments if and only if the state z belongs to the interval (z_ℓ, z_r). As before, the inner-most belief constitutes the worst-case belief. That is, stopping for action ℓ occurs at state z_ℓ when the DM is indifferent between stopping and experimenting *at her right-most belief* $z_\ell + \frac{\Delta}{2}$; and likewise for the stopping for action r. Differently from the Poisson model, however, the stopping belief $z_\ell + \frac{\Delta}{2}$ no longer coincides with the Bayesian stopping boundary z_ℓ^B. In fact, one can show that $z_\ell^B < z_\ell + \frac{\Delta}{2} < 0 < z_r - \frac{\Delta}{2} < z_r^B$.[15] Namely, the DM stops experimenting before stopping becomes Bayesian optimal for all beliefs in the set. To see this, consider the case where the DM's right-most belief is close to z_ℓ^B. If the DM continues experimentation, there is a small chance that the state drifts all the ways to the other side of the right stopping boundary. The DM understands that if this happens, she will stop only when the left-most belief reaches the stopping boundary. From the viewpoint of the right-most belief, there is hence excessive experimentation close to the right stopping boundary, which reduces the value of experimenting. The same applies to the other side. The argument further implies that in the experimentation region, there is no belief in $[z - \frac{\Delta}{2}, z + \frac{\Delta}{2}]$ for which the equilibrium strategy is Bayesian optimal. This means that the value segment for the ambiguity-averse DM is strictly below, and thus never touches, the Bayesian value function, as illustrated in Figure 12.

Graphically, in the small ambiguity case, the value segment is sufficiently short so that its lower end does not fall below the stopping payoff. Otherwise stopping becomes optimal under the worst-case belief, as we illustrate in Figure 13 (left panel). In such situations, the DM

[15] Epstein and Ji (2020) obtain a qualitatively similar result for the case of rectangular sets of priors.

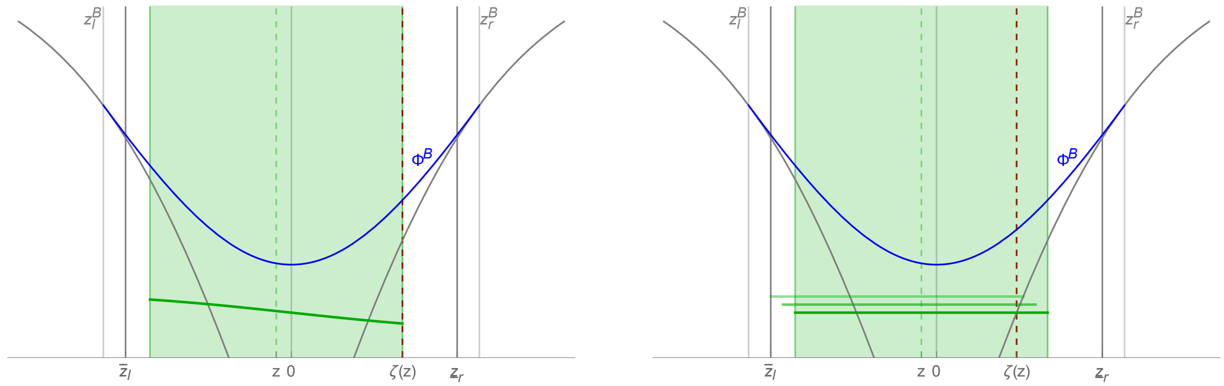

Figure 13: The case of large Δ: on the left, profitable deviation if the DM would experiment; on the right, intrapersonal equilibrium with mixed stopping (followed by r).

again randomizes between experimentation and stopping. Mixed stopping works here as follows: starting from $z = 0$, if z moves slightly to the left, the DM mixes between experimentation and stopping followed by action r; if instead z moves slightly to the right, the DM mixes between experimentation and stopping followed by ℓ. This pattern is analogous to how randomized stopping occurs in the baseline model. Intuitively, mixing with action r when $z < 0$ (resp. ℓ when $z > 0$) hedges the DM against the possibility that the state is in fact R (resp. L) and, hence, that experimentation will take a very long time.

Apart from technical details (discussed in Appendix C.2), the solution method and qualitative features of the solution thus extend to the drift-diffusion model. In particular: 1) the DM has incentives to engage in excessive experimentation close to the stopping boundaries; 2) if ambiguity is large and the flow cost c is small, the DM will exert self control and preempt herself from experimenting excessively by randomly stopping at intermediate sets of beliefs.

6 Concluding Remarks

This paper studies the canonical Wald problem under the assumption that the DM faces uncertainty about the underlying probabilities of the different states. We consider a DM who seeks to maximize her payoff guarantee and updates beliefs prior by prior. Given the assumption that ambiguity concerns the state but not the information technology, switches in the DM's worst-case scenario naturally arise, thereby leading to violations of dynamic consistency. Rather than ruling out such violations, we seek to understand their behavioral implications on information acquisition. The main technical challenge is then to solve the DM's optimization problem in the presence of dynamic preference reversals. We develop a new method, based on saddle-point HJB conditions, and provide a micro-foundation for the approach. The method proves highly tractable and is applicable to dynamic optimization problems well beyond information acquisition, such as real options theory, dynamic portfolio choice, or search models.

An open question is how our solution changes under different updating rules. Prior-by-prior updating entails that the DM updates on the state but not on the probability model. On the other extreme of the spectrum is maximum likelihood updating (Gilboa and Schmeidler, 1993), according to which the DM only retains those priors which explain the data best. In our setting, this would not be very interesting, as the DM, in the absence of a breakthrough, would discard all priors but the left-most after the first instant of time. There are, however, less extreme versions that combine features of both updating rules (e.g. Cheng (2021)). It will be interesting to explore the implications of such rules on the incentives to acquire information in dynamic settings.

We focus on the case where the DM is fully sophisticated and thus correctly forecasts the behavior of her future selves. An open question is how the results extend to the case where the DM is partially naive. For instance, the DM may be paying attention only at certain points in time. This could be modelled by having a Poisson clock run in parallel, which then determines the times at which the DM contemplates the actions of her future selves and optimizes accordingly (otherwise she follows the naively optimal plan). A higher Poisson rate would thus capture a more sophisticated DM. Given such a model, one could ask how the DM's welfare (say with respect to initial preferences) would depend on the level of sophistication. We leave these interesting questions for future research.

A Appendix

A.1 Microfoundation of the saddle-point HJB equations

We now justify our equilibrium concept, or the use of saddle-point HJB equations. The idea is to imagine that at each point in time, the DM can control, or commit to, her actions for a brief period of time, say $\epsilon > 0$. She thus acts dynamically consistently for a brief period of time, while taking as given the strategies of her future selves beyond that time interval. We then require the strategy to be a limit of the ϵ-dynamically consistent strategies as $\epsilon \to 0$.

To make this formal, fix a candidate strategy σ^*. Suppose that, at any state $\overline{p}$ and any time t, the DM can commit to any strategy σ (possibly different from σ^*) during a time interval $[t, t+\epsilon)$, for some $\epsilon > 0$, and the strategy switches to σ^* thereafter. Let $(\sigma_\epsilon(\cdot), \pi_\epsilon(\cdot))$ be the saddle point strategy for such an ϵ horizon problem (followed by σ^*), or an ϵ-commitment solution. We say that a pseudo equilibrium (σ^*, π^*) is *dynamically credible* if for each $\overline{p}$ there is a sequence of ϵ-commitment solutions $(\sigma_\epsilon(\overline{p}), \pi_\epsilon(\overline{p}))$ that converges to (σ^*, π^*) as $\epsilon \to 0$. We further say that a pseudo equilibrium (σ^*, π^*) is *well-behaved* if, for each $\overline{p} \in (0,1)$, $V^{\sigma^*}(\pi^*(\overline{p}), \overline{p})$, $V_p^{\sigma^*}(\pi^*(\overline{p}), \overline{p})$ and $V_p^{\sigma^*}(\pi^*(\overline{p}), \overline{p}_-)$ are well-defined and continuous at $(\pi^*(\overline{p}), \overline{p})$. We now provide the justification based on dynamic consistency.

Write $W_\epsilon^\sigma(p, \overline{p})$ for the DM's value when the DM follows strategy σ for ϵ-period, followed by the candidate strategy σ^*. Then,

$$\begin{aligned}
&W_\epsilon^\sigma(p,\overline{p}) \\
=& m(\overline{p})U_{\rho(\overline{p})}(p) + (1-m(\overline{p}))\left[-p\int_{\overline{p}^\epsilon}^{\overline{p}} e^{-\int_0^{\tau(\overline{p},\overline{p}')}(\lambda+\nu(p^\tau))d\tau}\left(-c+\lambda u_r^R+\nu(\overline{p}')u_{\rho(\overline{p}')}^R\right)\frac{1}{\eta(\overline{p}')}d\overline{p}'\right. \\
&-(1-p)\int_{\overline{p}^\epsilon}^{\overline{p}} e^{-\int_0^{\tau(\overline{p},\overline{p}')}\nu(p^\tau)d\tau}\left(-c+\nu(\overline{p}')u_{\rho(\overline{p}')}^L\right)\frac{1}{\eta(\overline{p}')}d\overline{p}' \\
&\left.+\left(pe^{-\int_0^{\tau(\overline{p},\overline{p}^\epsilon)}(\lambda+\nu(p^\tau))d\tau}+(1-p)e^{-\int_0^{\tau(\overline{p},\overline{p}^\epsilon)}\nu(p^\tau)d\tau}\right)V^{\sigma^*}(p^\epsilon,\overline{p}^\epsilon)\right],
\end{aligned}$$

where for any $q \in [0,1]$, q^δ denotes the δ-ahead update of belief q, and for any $p, p' \in [0,1]$ with $p < p'$

$$\tau(p,p') := \frac{1}{\lambda}\ln\left(\frac{p}{1-p}\frac{1-p'}{p'}\right)$$

denotes the time it takes for the Bayesian update of p to reach p' in the absence of R-evidence. Let $(\sigma_\epsilon(\cdot), \pi_\epsilon(\cdot))$ denote an ϵ-commitment solution. Then, it must satisfy the following saddle point condition: for each $\overline{p}$

$$\sigma_\epsilon(\overline{p}) \in \arg\max_\sigma W_\epsilon^\sigma(\pi_\epsilon(\overline{p}), \overline{p}) \tag{17}$$

$$\pi_\epsilon(\overline{p}) \in \arg\min_{p\in P(\overline{p})} W_\epsilon^{\sigma_\epsilon}(p, \overline{p}). \tag{18}$$

Let $W_\epsilon(\pi_\epsilon, \overline{p}) := W_\epsilon^{\sigma_\epsilon}(\pi_\epsilon, \overline{p})$ be the saddle point value of the ϵ-commitment solution at $\overline{p}$.

The ϵ-commitment saddle point value $W_\epsilon(\pi_\epsilon, \overline{p})$ is characterized as follows: For any $\delta \in (0, \epsilon)$,

$$
\begin{aligned}
&W_\epsilon(\pi_\epsilon, \overline{p}) \\
&= \max_{m,\nu,\rho} m(\overline{p}) U_{\rho(\overline{p})}(\pi_\epsilon) + (1 - m(\overline{p})) \Bigg[-\pi_\epsilon \int_{\overline{p}^\delta}^{\overline{p}} e^{-\int_0^{\tau(\overline{p},\overline{p}')} (\lambda + \nu(p^\tau)) d\tau} \left(-c + \lambda u_r^R + \nu(\overline{p}') u^R_{\rho(\overline{p}')} \right) \frac{1}{\eta(\overline{p}')} d\overline{p}' \\
&\quad - (1 - \pi_\epsilon) \int_{\overline{p}^\delta}^{\overline{p}} e^{-\int_0^{\tau(\overline{p},\overline{p}')} \nu(p^\tau) d\tau} \left(-c + \nu(\overline{p}') u^L_{\rho(\overline{p}')} \right) \frac{1}{\eta(\overline{p}')} d\overline{p}' \\
&\quad + \left(\pi_\epsilon e^{-\int_0^{\tau(\overline{p},\overline{p}^\delta)} (\lambda + \nu(p^\tau)) d\tau} + (1 - \pi_\epsilon) e^{-\int_0^{\tau(\overline{p},\overline{p}^\delta)} \nu(p^\tau) d\tau} \right) W_{\epsilon-\delta}(\pi_\epsilon^\delta, \overline{p}^\delta) \Bigg].
\end{aligned}
$$

Following the dynamic programming principle, the RHS does not depend on any $\delta \in (0, \epsilon)$. Hence, the derivative of the RHS with respect to δ evaluated at $\delta = 0$ must equal zero. So, we obtain the standard HJB equations:[16]

$$
0 = \max_{m,\nu,\rho} G(m, \nu, \rho, \pi_\epsilon, \overline{p}, W_\epsilon, dW_\epsilon), \tag{19}
$$

$$
\sigma_\epsilon(\overline{p}) = (m_\epsilon(\overline{p}), \nu_\epsilon(\overline{p}), \rho_\epsilon(\overline{p})) \in \arg\max_{m,\nu,\rho} G(m, \nu, \rho, \pi_\epsilon, \overline{p}, W_\epsilon, dW_\epsilon). \tag{20}
$$

Likewise, by the saddle point condition, $W_\epsilon(\pi_\epsilon, \overline{p})$ is also equivalently characterized as: for any $\delta \in (0, \epsilon)$,

$$
\begin{aligned}
&W_\epsilon(\pi_\epsilon, \overline{p}) \\
&= \min_{p} m_\epsilon(\overline{p}) U_{\rho_\epsilon(\overline{p})}(p) + (1 - m_\epsilon(\overline{p})) \Bigg[-p \int_{\overline{p}^\delta}^{\overline{p}} e^{-\int_0^{\tau(\overline{p},\overline{p}')} (\lambda + \nu_\epsilon(p^\tau)) d\tau} \left(-c + \lambda u_r^R + \nu_\epsilon(\overline{p}') u^R_{\rho(\overline{p}')} \right) \frac{1}{\eta(\overline{p}')} d\overline{p}' \\
&\quad - (1 - p) \int_{\overline{p}^\delta}^{\overline{p}} e^{-\int_0^{\tau(\overline{p},\overline{p}')} \nu_\epsilon(p^\tau) d\tau} \left(-c + \nu_\epsilon(\overline{p}') u^L_{\rho(\overline{p}')} \right) \frac{1}{\eta(\overline{p}')} d\overline{p}' \\
&\quad + \left(p e^{-\int_0^{\tau(\overline{p},\overline{p}^\delta)} (\lambda + \nu_\epsilon(p^\tau)) d\tau} + (1 - p) e^{-\int_0^{\tau(\overline{p},\overline{p}^\delta)} \nu_\epsilon(p^\tau) d\tau} \right) W^\sigma_{\epsilon-\delta}(p^\delta, \overline{p}^\delta) \Bigg].
\end{aligned}
$$

---

[16]Since $W_{\epsilon-\delta}(\pi_\epsilon^\delta, \overline{p}^\delta) = W_\epsilon(\pi_\epsilon, \overline{p}_-) + \frac{\partial W_\epsilon(\pi_\epsilon, \overline{p}_-)}{\partial p} \eta(\pi_\epsilon) \delta + \frac{\partial W_\epsilon(\pi_\epsilon, \overline{p}_-)}{\partial \overline{p}} \eta(\overline{p}) \delta + o(\delta)$, the derivative of $W_{\epsilon-\delta}(\pi_\epsilon^\delta, \overline{p}^\delta)$ with respect to δ is:

$$
\lim_{\delta \downarrow 0} \frac{W_{\epsilon-\delta}(\pi_\epsilon^\delta, \overline{p}^\delta) - W_\epsilon(\pi_\epsilon, \overline{p}_-)}{\delta} = \frac{\partial W_\epsilon(\pi_\epsilon, \overline{p}_-)}{\partial p} \eta(\pi_\epsilon) + \frac{\partial W_\epsilon(\pi_\epsilon, \overline{p}_-)}{\partial \overline{p}} \eta(\overline{p}).
$$

Following the dynamic programming principle, the RHS does not depend on any $\delta \in (0, \epsilon)$. Hence, the derivative of the RHS with respect to δ evaluated at $\delta = 0$ must equal zero. So, we obtain the standard HJB equations:

$$\pi_\epsilon \in \arg\min_{p \in P(\overline{p})} G(m_\epsilon, \nu_\epsilon, \rho_\epsilon, p, \overline{p}, W_\epsilon, dW_\epsilon). \tag{21}$$

The proof is Appendix Appendix B.3.We begin with a useful observation:

Lemma A.1. *Suppose a well-behaved pseudo equilibrium (σ^*, π^*) is dynamically credible. Then, for each $\overline{p} \in (0,1)$, for any sequence such that $(\sigma_\epsilon(\overline{p}), \pi_\epsilon(\overline{p})) \to (\sigma^*(\overline{p}), \pi^*(\overline{p}))$ as $\epsilon \to 0$, $W_\epsilon(\pi_\epsilon(\overline{p}), \overline{p}_-) \to V^{\sigma^*}(\pi^*(\overline{p}), \overline{p}_-)$, $\frac{\partial W_\epsilon(\pi_\epsilon(\overline{p}), \overline{p}_-)}{\partial p} \to V_p^{\sigma^*}(\pi^*(\overline{p}), \overline{p}_-)$, and $\frac{\partial W_\epsilon(\pi_\epsilon(\overline{p}), \overline{p}_-)}{\partial \overline{p}} \to V_{\overline{p}}^{\sigma^*}(\pi^*(\overline{p}), \overline{p}_-)$.*

The result now follows.

Theorem 2. *If a well-behaved pseudo equilibrium is dynamically credible, then it is an intrapersonal equilibrium. Conversely, if there is a unique intrapersonal equilibrium, then it is dynamically credible.*

Proof. Fix any $\overline{p}$. By the dynamic credibility of (σ^*, π^*), for each $\overline{p} \in (0,1)$, there exists a sequence of ϵ such that an associated ϵ-commitment solution $(\sigma_\epsilon(\overline{p}), \pi_\epsilon(\overline{p}))$ converges to $(\sigma^*(\overline{p}), \pi^*(\overline{p}))$ as $\epsilon \to 0$. Given the continuity assumption, by Lemma A.1, $W_\epsilon(\pi_\epsilon(\overline{p}), \overline{p}_-) \to V^{\sigma^*}(\pi^*(\overline{p}), \overline{p}_-)$, $\frac{\partial W_\epsilon(\pi_\epsilon(\overline{p}), \overline{p}_-)}{\partial p} \to V_p^{\sigma^*}(\pi^*(\overline{p}), \overline{p}_-)$, and $\frac{\partial W_\epsilon(\pi_\epsilon(\overline{p}), \overline{p}_-)}{\partial \overline{p}} \to V_{\overline{p}}^{\sigma^*}(\pi^*(\overline{p}), \overline{p}_-)$.

Next, observe that G is continuous. Hence, by Berge's maximum theorem, the maximized value is continuous and the set of maximizers must be upper hemicontinuous, which in this case should mean a closed graph. Hence, (10) and (11) must follow from (19) and (20) in the limit as $\epsilon \to 0$.

Applying the same argument, (12) follows from (21) in the limit as $\epsilon \to 0$. Again, by Berge's maximum theorem the minimized value is continuous and the set of minimizers must be upper hemicontinuous, which implies a closed graph. Hence, the conditions must hold in the limit as $\epsilon \to 0$.

Conversely, suppose (σ^*, π^*) is intrapersonal equilibrium. Then, it satisfies (10), (11), and (12), for each $\overline{p}$. Hence, for each $\overline{p}$, $V^{\sigma^*}(\pi^*(\overline{p}), \overline{p}), V_p^{\sigma^*}(\pi^*(\overline{p}), \overline{p})$ and $V_p^{\sigma^*}(\pi^*(\overline{p}), \overline{p}_-)$ are well-defined and continuous at $(\pi^*(\overline{p}), \overline{p})$. Take a sequence (subsequence if necessary) of $\epsilon > 0$ such that $(\sigma_\epsilon, \pi_\epsilon, W_\epsilon(\pi_\epsilon(\overline{p}), \overline{p}), \partial W_\epsilon(\pi_\epsilon(\overline{p}), \overline{p})/\partial p, \partial W_\epsilon(\pi_\epsilon(\overline{p}), \overline{p})/\partial \overline{p})$ all converge to some limit, say $(\sigma_0, \pi_0, W_0(\pi_0(\overline{p}), \overline{p}), \tilde{W}_{0,p}(\pi_0(\overline{p}), \overline{p}), \tilde{W}_{0,\overline{p}}(\pi_0(\overline{p}), \overline{p}))$. The reason that such a sequence can be found is as follows. First, $(W_\epsilon(\pi_\epsilon(\overline{p}),\overline{p}), \partial W_\epsilon(\pi_\epsilon(\overline{p}),\overline{p})/\partial p)$ is bounded within a compact set $\left[\min\{u_\ell^R, u_r^L\}, \max\{u_r^R, u_\ell^L\}\right] \times [u_\ell^R - u_\ell^L, u_r^R - u_r^L]$. This together with (20) means that $\nu_\epsilon(\overline{p})$ can be bounded within $[0, \overline{\nu}]$ for some large $\overline{\nu}$. The boundedness of $\nu_\epsilon(\overline{p})$ in turn means that $\partial W_\epsilon(\pi_\epsilon(\overline{p}),\overline{p})/\partial\overline{p}$ can be bounded within an compact set. This proves that the converging sequence can be found for each $\overline{p}$. Next, since $\partial^2 W_\epsilon(\pi_\epsilon(\overline{p}),\overline{p})/\partial p \partial\overline{p} = 0$ for all ϵ (which follows from the fact that $W_\epsilon(\cdot,\overline{p})$ is linear), we have $\tilde{W}_{0,p}(\pi_0(\overline{p}),\overline{p}) = \partial W_0(\pi_0(\overline{p}),\overline{p})/\partial p$ and $\tilde{W}_{0,\overline{p}}(\pi_0(\overline{p}),\overline{p}) = \partial W_0(\pi_0(\overline{p}),\overline{p})/\partial\overline{p}$. It is immediate that the limiting strategy profile (σ_0, π_0) is a pseudo equilibrium. Also using the same argument as above, we conclude that the limit strategy profile (σ_0, π_0), together with the value function W_0, satisfies (10), (11), and (12) for each $\overline{p}$. Therefore, (σ_0, π_0) is also an intrapersonal equilibrium. The uniqueness of the equilibrium then means that $(\sigma_0, \pi_0) = (\sigma^*, \pi^*)$, so for each $\overline{p}$, $(\sigma^*(\overline{p}), \pi_0^*(\overline{p}))$ is a limit of some ϵ-commitment solution $(\sigma_\epsilon(\overline{p}), \pi_\epsilon(\overline{p}))$ as $\epsilon \to 0$. Hence, (σ^*, π^*) is dynamically credible. □

A.2 Proof of Theorem 1

For our purpose, we first establish a preliminary result about the Bayesian value function. Recall the solution ϕ to the ODE (1). Then, the following lemma holds.

Lemma A.2. *For any $p > p_\ell^B$, $\phi'(p; p, U_\ell(p)) - U_\ell'(p) > 0$ whenever $\phi(p; p, U_\ell(p)) = U_\ell(p)$.*

Proof. Define $\delta^\omega := |\delta_r^\omega - \delta_\ell^\omega|, \omega = L, R$. It follows from (1) that

$$\begin{aligned}
\lambda p(1-p)(\phi'(p; p, U_\ell(p)) - U_\ell'(p)) &= \lambda p(u_r^R - \phi(p; p, U_\ell(p))) - c + \lambda(1-p)(u_\ell^L - U_\ell(p)) \\
&= \lambda\left(p u_r^R + (1-p)u_\ell^L - U_\ell(p)\right) - c \\
&= \lambda p \delta^R - c,
\end{aligned}$$

where the second equality uses the fact that $\phi(p; p, U_\ell(p))) = U_\ell(p)$. Since the last line is strictly increasing in p and equals 0 when $p = p_\ell^B$, the above claim is proven. □

For the proof, for each region of $\overline{p}$'s, we will specify the strategy profile (σ, π) and the associated value $V^\sigma(p, \overline{p})$ that are not fully specified in Theorem 1. We will then verify the value function together with the strategy profile (σ, π) satisfy the equilibrium conditions.

For ease of presentation, we shall mainly focus on Case 1. Further, we will assume that either $c < \underline{c}$ or $c \in (\underline{c}, \overline{c}]$ but $\Delta < \Delta_c$, where Δ_c will be specified later (in Region 3 in Appendix A.2.1).

We call this case **main case**. We will first treat the main case in Appendix A.2.1. The remaining cases will be treated in Appendix A.2.2 and Appendix B.7.

A.2.1 Main Case

Region 1: $\overline{p} \in [0, \overline{p}_1]$, where $\overline{p}_1 := p_\ell^B$.

• **Computation of equilibrium value.** Recall that the strategy σ calls for an immediate choice of ℓ for all state $\overline{p} \le p_\ell^B = \overline{p}_1$. It thus immediately follows that the value associated with that strategy is:

$$V^\sigma(p, \overline{p}) := U_\ell(p), \text{ for all } \overline{p} \in [0, \overline{p}_1].$$

• **Verification of equilibrium conditions.** With $V^\sigma(p, \overline{p}) = U_\ell(p)$, $m(\overline{p}) = 1$ and $\rho(\overline{p}) = 0$ imply that $\pi(\overline{p}) = \overline{p}$ is a minimizer in (12). Since $U'_\ell(p) < 0$ by assumption, $\pi(\overline{p}) = \overline{p}$ is the unique minimizer in (6). Substituting $V^\sigma(p, \overline{p}) = U_\ell(p)$, $V_p^\sigma(p, \overline{p}) = U'_\ell(p)$, $V_{\overline{p}}^\sigma(p, \overline{p}) = 0$, and $\pi(\overline{p}) = \overline{p}$, (10) simplifies to

$$\max_{m,\nu,\rho} m\left[U_\rho(\overline{p}) - U_\ell(\overline{p})\right] + (1-m)\left[-c + \nu(U_\rho(\overline{p}) - U_\ell(\overline{p})) + \overline{p}\lambda(u_r^R - U_\ell(\overline{p})) + U'_\ell(\overline{p})\eta(\overline{p})\right] = 0.$$

With $m = 1$ and $\rho = 0$, the LHS is zero. Moreover, for any ρ, ν, simple algebra shows that the coefficient of $(1 - m)$ is non-positive for all $\overline{p} \le \overline{p}_1$. Hence $\sigma = (1, 0, 0)$ is a maximizer and both (10) and (11) are satisfied for $V^\sigma(p, \overline{p}) = U_\ell(p)$. Setting $m(\overline{p}) = 1$ and $\rho(\overline{p}) = 0$, the objective in (12) is independent of p and $\pi(\overline{p}) = \overline{p}$ is a minimizer. Hence we have shown that for the posited strategy profile (σ, π) satisfy the equilibrium conditions, (6)–(12) for all $\overline{p} \in [0, \overline{p}_1]$.

Region 2: $\overline{p} \in (\overline{p}_1, \overline{p}_2]$, where $\overline{p}_2 := p_*$.

• **Computation of equilibrium value.** Recall that the strategy σ calls on the DM to experiment until $\overline{p}$ drifts to $\overline{p}_1 = p_\ell^B$. Fix any state $\overline{p}$ in this region. To compute the associated value $V^\sigma(\cdot, \overline{p})$, it is useful to ask: *for which belief $p \in P(\overline{p})$ is the continuation strategy $\sigma_{\overline{p}}$ Bayesian optimal?* The answer is $p = \overline{p}$, since the strategy is exactly what a Bayesian DM with belief $\overline{p}$ will do—experimenting until $\overline{p}$ reaches the Bayesian optimal stopping belief $p_\ell^B = \overline{p}_1$. Consequently, we must have

$$V^\sigma(\overline{p}, \overline{p}) = \Phi(\overline{p}).$$

Since the strategy needs not be optimal for any DM with belief $p \neq \overline{p}$ (including p outside $P(\overline{p})$), we must have

$$V^\sigma(p,\overline{p}) \leq \Phi(p), \forall p \in [0,1].$$

Finally, since $V^\sigma(p,\overline{p})$—the valuation of a fixed action path—must be linear in p, the preceding observations must imply:

$$V^\sigma(p,\overline{p}) := \Phi(\overline{p}) + (p-\overline{p})\Phi'(\overline{p}), \text{ for all } \overline{p} \in (\overline{p}_1, \overline{p}_2]. \tag{22}$$

• **Verification of equilibrium conditions.** Since $\Phi'(\overline{p}) \leq 0$ for all $\overline{p} \in (\overline{p}_1, \overline{p}_2]$, $V^\sigma_p(p,\overline{p}) \leq 0$ and therefore $\pi(\overline{p}) = \overline{p}$ is a minimizer in (6). With $\pi(\overline{p}) = \overline{p}$ we have

$$V^\sigma(\pi(\overline{p}),\overline{p}) = \Phi(\overline{p}), \quad V^\sigma_p(\pi(\overline{p}),\overline{p}) = \Phi'(\overline{p}), \quad V^\sigma_{\overline{p}}(\pi(\overline{p}),\overline{p}) = 0.$$

Therefore (10) simplifies to

$$\max_{m,\nu,\rho} m\left[U_\rho(\overline{p}) - \Phi(\overline{p})\right] + (1-m)\left[-c + \nu(U_\rho(\overline{p}) - \Phi(\overline{p})) + \overline{p}\lambda(u_r^R - \Phi(\overline{p})) + \Phi'(\overline{p})\eta(\overline{p})\right] = 0.$$

Since $\Phi(\overline{p}) \geq U(\overline{p})$ for any $\overline{p}$, $\nu = 0$ is optimal. Moreover, $m = 0$ is optimal since we have from (1)

$$-c + \overline{p}\lambda(u_r^R - \Phi(\overline{p})) + \Phi'(\overline{p})\eta(\overline{p}) = 0$$

for the Bayesian value function. Therefore (10) and (11) are satisfied for all $\overline{p} \in (\overline{p}_1, \overline{p}_2]$. Condition (12) holds since $m(\overline{p}) = \nu(\overline{p}) = 0$ implies that the objective in (12) is given by

$$\begin{aligned} &-c + p\lambda\left(u_r^R - (\Phi(\overline{p}) + (p-\overline{p})\Phi'(\overline{p}))\right) - \lambda p(1-p)\Phi'(\overline{p}) - \lambda\overline{p}(1-\overline{p})(p-\overline{p})\Phi''(\overline{p}) \\ = &-c + p\lambda\left(u_r^R - (\Phi(\overline{p}) + (1-\overline{p})\Phi'(\overline{p}))\right) - \lambda\overline{p}(1-\overline{p})(p-\overline{p})\Phi''(\overline{p}). \end{aligned}$$

Differentiating this with respect to p yields

$$\lambda\left(u_r^R - (\Phi(\overline{p}) + (1-\overline{p})\Phi'(\overline{p}))\right) - \lambda\overline{p}(1-\overline{p})\Phi''(\overline{p}).$$

Differentiating the ODE for the Bayesian Value function (see (1)) with respect to $\overline{p}$ we obtain that this is expression is equal to zero. Hence $\pi(\overline{p})$ is a minimizer in (12).

To summarize, we have shown that for the posited profile (σ, π), together with the value function V^σ, satisfy (6)–(12) for all $\overline{p} \in (\overline{p}_1, \overline{p}_2]$.

Region 3: $\overline{p} \in (\overline{p}_2, \overline{p}_3]$**.**

This region exists only when the ambiguity is sufficiently large so that $V^\sigma(\underline{p}(\overline{p}_2), \overline{p}_2) < U_\ell(\underline{p}(\overline{p}_2))$. If this inequality is reversed, then we set $\overline{p}_3 = \overline{p}_2$, and Region 3 is empty. Assuming the inequality, we will specify the upper bound $\overline{p}_3$, and the critical cost $\underline{c}$ and ambiguity level Δ_c referred to in the statement of the theorem. Finally, we will specify $\nu(\overline{p})$ and $\pi(\overline{p})$ fully for $\overline{p}$ in this region.

• **Specification of ν and π and equilibrium value.** The strategy σ for this region involves randomization between experimentation and stopping for ℓ, with the latter done at a Poisson rate $\nu(\overline{p})$. Meanwhile, nature chooses belief $\pi(\overline{p}) \in P(\overline{p})$. Here, we specify (ν, π) precisely, together with the value $V^\sigma(p, \overline{p})$ associated with the strategy. We first construct a value function $V(p, \overline{p})$ that satisfies the equilibrium conditions, (6)-(12), given the candidate strategy σ. We will then establish that the constructed function $V(p, \overline{p})$ indeed coincides with the value of σ.

To begin, fix any $\overline{p} > \overline{p}_2$. First, the fact that the DM randomizes between experimentation and action ℓ means that the coefficient of ν in (11) must vanish, which implies (13). For the required belief $\pi(\overline{p})$ to be nature's choice satisfying (6), it is sufficient, and will be seen also necessary, to have $V_p(p, \overline{p}) = 0$. Namely,

$$V(p, \overline{p}) = \hat{V}(\overline{p}), \forall p. \tag{23}$$

Substituting this into (13), we get (14). The randomization by DM in turn implies that the derivative of the objective in (11) with respect to p must vanish. This fact, together with (23) yields (15).

It now remains to specify the function $\hat{V}(\overline{p})$. To this end, we use (14),(15) and (23) to simplify (10) and obtain the differential equation ($\widehat{\text{ODE}}$). Together with the boundary condition that $\hat{V}(\overline{p}_2) = \Phi(\overline{p}_2)$, ($\widehat{\text{ODE}}$) admits a unique solution:[17]

$$\hat{V}(\overline{p}) := \frac{C}{\left(\frac{1-\overline{p}}{\overline{p}}\right)^{\frac{u_2-u_1}{\delta_\ell}} + C} u_1 + \frac{\left(\frac{1-\overline{p}}{\overline{p}}\right)^{\frac{u_2-u_1}{\delta_\ell}}}{\left(\frac{1-\overline{p}}{\overline{p}}\right)^{\frac{u_2-u_1}{\delta_\ell}} + C} u_2, \tag{24}$$

[17] The derivation of the solution appears in Appendix B.5.

where

$$C := \frac{u_2 - \Phi(\overline{p}_2)}{\Phi(\overline{p}_2) - u_1} \left(\frac{1 - \overline{p}_2}{\overline{p}_2} \right)^{\frac{u_2 - u_1}{\delta_\ell}}, \tag{25}$$

and u_1 and u_2 are lower and higher roots of a quadratic equation; namely,

$$u_{1,2} := \frac{u_r^R + u_\ell^L}{2} \pm \sqrt{\left(\frac{u_r^R - u_\ell^L}{2} \right)^2 + \frac{c}{\lambda} \delta_\ell}. \tag{26}$$

It is instructive to visualize how nature's feasible "set of values" evolves on the equilibrium path as $\overline{p}$ drifts down in this region. Our construction so far indicates that at each state the feasible set forms a flat segment $\{(p, \hat{V}(\overline{p})) : p \in [\underline{p}(\overline{p}), \overline{p}]\}$, and the belief at which the segment crosses ℓ-payoff function $U_\ell(p)$ is precisely nature's choice $\pi(\overline{p})$. As $\overline{p}$ falls in this region, the segment shifts left. The starting state of Region 3, or its right most boundary, $\overline{p}_3$, is the state such that its associated value segment just meets or "touches" the ℓ-payoff function $U_\ell(p)$ at the former's left-most end. Formally, $\overline{p}_3$ is defined by:

$$\hat{V}(\overline{p}_3) = U_\ell(\underline{p}(\overline{p}_3)). \tag{27}$$

Let $\hat{u} := \min_{p \in [0,1]} U(p)$. The following observations are useful to establish:

Lemma A.3. **(i)** *$\hat{V}(\cdot)$ is strictly decreasing, and $\hat{V}(\overline{p}) \in (u_1, u_2)$ for all $\overline{p} > \overline{p}_2$.*

(ii) *If $V^\sigma(\underline{p}(\overline{p}_2), \overline{p}_2) < U_\ell(\underline{p}(\overline{p}_2))$ (i.e., the condition for Region 3 to exist holds) and $u_1 > u_\ell^R$, then there exists a unique solution $\overline{p}_3 \in (\overline{p}_2, 1)$.*

(iii) *If $c \leq \underline{c}$, then $u_1 \geq \hat{u}$, so $V^\sigma(\cdot, \overline{p}_2) > \hat{u}$ for all $\overline{p} \geq \overline{p}_2$. If $c \in [\underline{c}, \overline{c})$, there exists $\Delta_c > 0$ such that $V^\sigma(\cdot, \overline{p}_3) = \hat{V}(\overline{p}_3) \geq \hat{u}$ if and only if $\Delta < \Delta_c$.*

Proof. The proof can be found in Appendix B.4 below. □

Lemma A.3-(i) implies that the value segment shifts up (as well as shifts left) as $\overline{p}$ drifts down.[18] Next, Lemma A.3-(ii) implies that $\overline{p}_3$ is well defined and above $\overline{p}_2$ whenever the region exists. Last, Lemma A.3-(iii) ensures that the value segment remains above $\hat{u}$ for all $\overline{p} \in (\overline{p}_2, \overline{p}_3]$,

[18]One can also see that V^σ is continuous, including at the boundary. This follows from the fact that $\hat{V}(\overline{p}_2) = \Phi(\overline{p}_2)$, and that $\overline{p}_2 = p_*$ and $\Phi'(p_*) = 0$.

if either $c \leq \underline{c}$ or if $c \in [\underline{c}, \overline{c})$ but $\Delta < \Delta_c$, for some $\Delta_c > 0$. Our verification below will use this fact, assuming the sufficient condition.

So far, our value function $V(p, \overline{p})$ is constructed from HJB equations along with (6). We now claim the value function indeed represents the value of the candidate strategy σ:

Lemma A.4. *For all $\overline{p} \in (\overline{p}_2, \overline{p}_3)$, $V(p, \overline{p}) = V^\sigma(p, \overline{p})$.*

Proof. See Appendix B.6. □

• **Verification of equilibrium conditions.** To verify (10)–(11), note that the definition of $\pi(\overline{p})$ in (14) implies that $U_\ell(\pi(\overline{p})) - V^\sigma(\pi(\overline{p}), \overline{p}) = 0$. We have already observed from Lemma A.3-(iii) that $V^\sigma(\cdot, \overline{p}) \geq \hat{u}$ for all $\overline{p} \in (\overline{p}_2, \overline{p}_3]$. This further implies that $\pi(\overline{p}) \leq \hat{p}$ for $\overline{p} \in (\overline{p}_2, \overline{p}_3]$. Therefore $\rho(\overline{p}) = 0$ is a maximizer in (10)–(11). By (13), $\nu(\overline{p})$ as defined in (15) is a maximizer as well. It remains to show that $m(\overline{p}) = 0$ is a maximizer and the RHS in (10)–(11) is equal to zero, both of which would follow if the coefficient of $(1-m)$ in the objective in (10)–(11) is equal to zero. We can use (13) and (23) to simplify the coefficient of $(1-m)$ to:

$$-c + \pi(\overline{p})\lambda(u_r^R - \hat{V}(\overline{p})) - \lambda\overline{p}(1-\overline{p})\hat{V}(\overline{p}), \tag{28}$$

which vanishes precisely because $\hat{V}$ solves $(\widehat{\mathrm{ODE}})$. We thus conclude that $m(\overline{p}) = 0$ is a minimizer in (11). We have thus verified (24) and (11).

To verify (12) note that, with $m(\overline{p}) = \rho(\overline{p}) = 0$ and $\nu(\overline{p})$ defined in (15), the derivative of the objective in (12) with respect to p simplifies to:

$$\nu(\overline{p})\left(u_\ell^R - u_\ell^L\right) + \lambda(u_r^R - V(\pi(\overline{p}), \overline{p})) = \nu(\overline{p})\left(u_\ell^R - u_\ell^L\right) + \lambda\left(u_r^R - \hat{V}(\overline{p})\right) = 0,$$

where we have used $V_p(\pi(\overline{p}), \overline{p}) = V_{p\overline{p}}(\pi(\overline{p}), \overline{p}) = 0$. Hence, we conclude that $\pi(\overline{p})$ satisfies (12). It also satisfies (6) since $V^\sigma(p, \overline{p})$ is constant in p.

Region 4: $\overline{p} \in (\overline{p}_3, \overline{p}_4]$.

Here, $\overline{p}_4$ will be defined as part of the analysis.

Computation of equilibrium value. Recall that the strategy σ calls on the DM to experiment fully at each state $\overline{p} > \overline{p}_3$ until state $\overline{p}_3$ is reached. To compute the associated value

$V^\sigma(p, \overline{p})$, it is useful to consider a (hypothetical) Bayesian DM who would find such a strategy optimal. To this end, we represent the DM's problem as a stopping problem where the stopping payoff is the continuation value $v_{**} := \hat{V}(\overline{p}_3)$:[19]

> **Auxiliary Problem:** Imagine a hypothetical Bayesian DM with any belief $p \geq \underline{p}(\overline{p}_3)$ who at each instant may experiment or stops. She may experiment until either breakthrough occurs or her belief reaches $\underline{p}(\overline{p}_3)$. If she stops at any point, she collects the payoff $\hat{V}(\overline{p}_3)$ (independent of p). The optimal value of this stopping problem is:
>
> $$\Psi(p) = \max_{p' \in [\underline{p}(\overline{p}_3), p]} \phi(p; p', v_{**}). \tag{29}$$
>
> Let p_{**} denote the optimal stopping belief. We note that $\Psi(p_{**}) = v_{**}$ and $\Psi'(p_{**}) \geq 0$, with equality holding whenever $p_{**} > \underline{p}(\overline{p}_3)$ (a consequence of smooth pasting). We can easily see that $p_{**} \in [\underline{p}(\overline{p}_3), \overline{p}_3)$.[20]

Fix any state $\overline{p} > \overline{p}_3$. We ask: *for what belief p is the continuation strategy $\sigma_{\overline{p}}$ optimal over Region 3*? Recall $\sigma_{\overline{p}}$ prescribes: "experiment for the duration of $\tau(\overline{p}, \overline{p}_3)$ and, absent breakthrough by the end of the experimentation, stop and collect $\hat{V}(\overline{p}_3)$," where $\tau(p, p')$ denotes the time it takes for a belief to drift from p to p'. Since, by definition, p_{**} is the optimal stopping belief, the answer to the above question is precisely the belief, $q(\overline{p})$, such that

$$\tau(\overline{p}, \overline{p}_3) = \tau(q(\overline{p}), p_{**}). \tag{30}$$

In words, it is the belief that would be updated to p_{**} after the prescribed duration $\tau(\overline{p}, \overline{p}_3)$ of experimentation.[21] Since p_{**} is the optimal stopping belief in the above Auxiliary Problem, a DM with belief $q(\overline{p})$ finds the prescribed strategy optimal at state $\overline{p}$ and thus will realize the

[19] Recall that if Region 3 does not exist, then $\overline{p}_3 = \overline{p}_2 = p_*$, so $\hat{V}(\overline{p}_3) = \hat{V}(\overline{p}_2) = \Phi(\overline{p}_2)$, as defined before. All subsequent results hold since $V_p(p, \overline{p}_2) = \Phi'(\overline{p}_2) = \Phi'(p_*) = 0$.

[20] To see that $p_{**} < \overline{p}_3$, it suffices to show that $p_{**} < p_*$ since $p_* \leq \overline{p}_3$. To show $p_{**} < p_*$, suppose otherwise. Then,

$$\lambda p_{**}(1 - p_{**})\Psi'(p_{**}) = \lambda p_{**}(u_r^R - \Psi(p_{**})) - c = \lambda p_{**}(u_r^R - v_{**}) - c > \lambda p_*(u_r^R - \Phi(p_*)) - c = \Psi'(p_*) \geq 0,$$

where the first and second last equalities follow from the fact that both Ψ and Φ solve (1), the strict inequality from $v_{**} < \Phi(p)$ for all p, and the last inequality holds since $\overline{p}_3 \geq p_*$.

[21] Obviously, $q(\overline{p}_3) = p_{**}$.

value of $\Psi(q(\overline{p}))$. Consequently,

$$V^\sigma(q(\overline{p}),\overline{p}) = \Psi(q(\overline{p})).$$

What about a DM with belief $p \neq q(\overline{p})$? Her value is weakly below $\Psi(q(p))$ and as noted before linear in p. This pins down the value function for belief p at state $\overline{p}$:

$$V^\sigma(p,\overline{p}) := \Psi(q(\overline{p})) + (p - q(\overline{p}))\Psi'(q(\overline{p})). \tag{31}$$

We define $\overline{p}_4$ to be such that

$$V^\sigma(\underline{p}(\overline{p}_4),\overline{p}_4) = U_r(\underline{p}(\overline{p}_4)). \tag{32}$$

• **Verification of equilibrium conditions.** We first claim that $\pi(\overline{p}) = \underline{p}(\overline{p})$ is a worst-case belief satisfying (6). This follows from the fact, for any $p \in P(\overline{p}), \overline{p} \geq \overline{p}_3$,

$$V_p^\sigma(p,\overline{p}) = \Psi'(q(p)) \geq 0,$$

where the equality follows from (31) and the inequality from the convexity of $\Psi(\cdot)$ and $\Psi'(p_{**}) \geq 0$.

We next verify (11). First, we show that $\nu(\overline{p}) = 0$. To this end, we first observe:

Lemma A.5. *For all $\overline{p} > \overline{p}_3$, $V^\sigma(\underline{p}(\overline{p}),\overline{p}) \geq U_\ell(\underline{p}(\overline{p}))$.*

Proof. Note $V^\sigma(\underline{p}(\overline{p}_3),\overline{p}_3) = \hat{V}(\overline{p}_3) \geq U_\ell(\underline{p}(\overline{p}_3))$.[22] Next, note that $\underline{p}(\overline{p}_3) > \underline{p}_r^B$ and $V^\sigma(\underline{p}(\overline{p}_3),\overline{p}_3) = \phi(\underline{p}(\overline{p}_3); \underline{p}(\overline{p}_3), U_\ell(\underline{p}(\overline{p}_3)))$. Since, by Lemma A.2, $\phi(p; p, U_\ell(p))$ can only cross $U_\ell(p)$ from below, we have $\left.\frac{dV^\sigma(\underline{p}(\overline{p}),\overline{p})}{d\overline{p}}\right|_{\overline{p}=\overline{p}_3} \geq U_\ell'(\underline{p}(\overline{p}_3))$. Observe further $V^\sigma(\underline{p}(\overline{p}),\overline{p}) = \phi(\underline{p}(\overline{p}); \underline{p}(\overline{p}_3), U_\ell(\underline{p}(\overline{p}_3)))$ is convex in $\underline{p}(\overline{p})$ whereas $U_\ell(\underline{p}(\overline{p}))$ is linear in $\underline{p}(\overline{p})$. Combining the two facts leads to the desired conclusion. □

Next, we observe that $U_\ell(\underline{p}(\overline{p}_3)) = \hat{V}(\overline{p}_3) > \hat{u}$. This means that

$$V^\sigma(\underline{p}(\overline{p}_3),\overline{p}_3) = U_\ell(\underline{p}(\overline{p}_3)) > U_r(\underline{p}(\overline{p}_3)).$$

By definition of $\overline{p}_4$, we have $V^\sigma(\underline{p}(\overline{p}),\overline{p}) > U_r(\underline{p}(\overline{p}))$ for all $\overline{p} \in (\overline{p}_3,\overline{p}_4)$. Combining this with Lemma A.5, we conclude that $\nu(\overline{p}) = 0$.

[22] If $V^\sigma(\underline{p}(\overline{p}_2),\overline{p}_2) < U_\ell(\underline{p}(\overline{p}_2))$, then the inequality holds with equality by definition. If $V^\sigma(\underline{p}(\overline{p}_2),\overline{p}_2) \geq U_\ell(\underline{p}(\overline{p}_2))$, then the inequality follows from the fact that $\overline{p}_3 = \overline{p}_2$.

Next, we prove $m(\overline{p}) = 0$. Substituting $\nu(\overline{p}) = 0$ and $\pi(\overline{p}) = \underline{p}(\overline{p})$, the objective in (11) becomes

$$m\left[U_\rho(\underline{p}) - V(\underline{p}, \overline{p}_-)\right] + (1-m)\left[-c + \lambda \underline{p}(u_r^R - V(\underline{p}, \overline{p}_-)) + V_p(\underline{p}, \overline{p}_-)\eta(p) + V_{\overline{p}}(\underline{p}, \overline{p}_-)\eta(\overline{p})\right],$$

where we suppress the argument of $\underline{p}(\overline{p})$ for notational ease. Note that the coefficient of m is negative (by the same argument as for $\nu = 0$). The coefficient of $(1-m)$ can be written

$$\begin{aligned}
&-c + \lambda \underline{p}(u_r^R - V(\underline{p}, \overline{p}_-)) + V_p(\underline{p}, \overline{p}_-)\eta(\underline{p}) + V_{\overline{p}}(\underline{p}, \overline{p}_-)\eta(\overline{p}) \\
=& -c + \lambda \underline{p}\left(u_r^R - \Psi(q(\overline{p})) - (\underline{p} - q(\overline{p}))\Psi'(q(\overline{p}))\right) \\
&+ \eta(\underline{p})\Psi'(q(\overline{p})) + \eta(\overline{p})\left[(\underline{p} - q(\overline{p}))\Psi''(q(\overline{p}))q'(\overline{p})\right] \\
=& -c + \lambda q(\overline{p})\left(u_r^R - \Psi(q(\overline{p}))\right) + \eta(q(\overline{p}))\Psi'(q(\overline{p})) \\
&+ \lambda(\underline{p} - q(\overline{p}))\left[u_r^R - \Psi(q(\overline{p})) - (1-q(\overline{p}))\Psi'(q(\overline{p})) - q(\overline{p})(1-q(\overline{p}))\Psi''(q(\overline{p}))\right] \\
=&0,
\end{aligned}$$

where we have used (31) for the first equality and $q'(\overline{p}) = \frac{\eta(q(\overline{p}))}{\eta(\overline{p})}$ for the second equality.[23] The last equality follows since Ψ satisfies (1) and its derivative vanishes. We have shown that the coefficient of m in the objective in (10) and (11) is negative and the coefficient of $(1-m)$ is zero. Therefore, $m = 0$ is a maximizer and (10) holds.

It remains to verify (12). Substituting $m(\overline{p}) = \nu(\overline{p}) = 0$, the objective becomes the same as the coefficient of $1-m$ above, except that $\underline{p}$ is replaced by p. Differentiating the expression with respect to p yields:

$$\lambda\left(u_r^R - \Psi(q(\overline{p})) - (1-q(\overline{p}))\Psi'(q(\overline{p}))\right) - \lambda q(\overline{p})(1-q(\overline{p}))\Psi''(q(\overline{p})) = 0,$$

where the equality follows from the fact that Ψ satisfies (1), so its derivative, which coincides with the LHS of the above equation, must vanish.

[23]This holds since $\tau(q(\overline{p}), q(\overline{p}_3)) = \tau(\overline{p}, \overline{p}_3)$ implies that $\ln\frac{q(\overline{p})}{1-q(\overline{p})} - \ln\frac{p_*}{1-p_*} = \ln\frac{\overline{p}}{1-\overline{p}} - \ln\frac{\overline{p}_3}{1-\overline{p}_3}$. Hence the difference in log-likelihood ratios of $q(\overline{p})$ and $\overline{p}$ is constant

$$\ln\frac{q(\overline{p})}{1-q(\overline{p})} - \ln\frac{\overline{p}}{1-\overline{p}} = K,$$

for some K. Differentiating this with respect to $\overline{p}$, we obtain $q'(\overline{p}) = \frac{\eta(q(\overline{p}))}{\eta(\overline{p})}$.

We have thus shown that the objective in (12) is independent of p and hence the requirement that $p = \underline{p}(\overline{p})$ is a minimizer in (12) is satisfied.

Region 5: $\overline{p} \in [\overline{p}_4, 1]$**.**

• **Computation of equilibrium value.** Recall that the strategy σ calls for an immediate choice of r for all states $\overline{p} \in [\overline{p}_4, 1]$. It thus immediately follows that the value associated with that strategy is:

$$V^\sigma(p, \overline{p}) := U_r(p), \text{ for all } \overline{p} \in [\overline{p}_4, 1].$$

• **Verification of equilibrium conditions.** Since $U_r(\cdot)$ is increasing, $\pi(\overline{p}) = \underline{p}(\overline{p})$ satisfies (6); given $m(\overline{p}) = \rho(\overline{p}) = 1$, the coefficient of p in the objective of (12) vanishes, so $\pi(\overline{p}) = \underline{p}(\overline{p})$ is also a minimizer in (12).

Substituting $\pi(\overline{p}) = \underline{p}(\overline{p}) = \underline{p}$ and $V^\sigma(p, \overline{p}) = U_r(\overline{p})$ in (10)–(11) we get the following expression for $\overline{p} > \overline{p}_4$

$$\begin{aligned}
&m\left[U_\rho(\underline{p}) - U_r(\underline{p})\right] + (1-m)\left[-c + \lambda\underline{p}\left(u_r^R - U_r(\underline{p})\right) - \lambda\underline{p}(1-\underline{p})U_r'(\underline{p})\right] \\
=&m\left[U_\rho(\underline{p}) - U_r(\underline{p})\right] + (1-m)\left[-c + \lambda\underline{p}\left(U_r(1) - U_r(\underline{p}) - (1-\underline{p})U_r'(\underline{p})\right)\right] \\
=&m\left[U_\rho(\underline{p}) - U_r(\underline{p})\right] + (1-m)(-c),
\end{aligned}$$

where the last line follows since $U_r(\overline{p})$ is linear. Recall that $U_\ell(\underline{p}(\overline{p}_4)) \leq V^\sigma(\underline{p}(\overline{p}_4), \overline{p}_4) = U_r(\underline{p}(\overline{p}_4))$, so we note that $U_\ell(\underline{p}) \leq U_r(\underline{p})$ for $\underline{p} \leq \overline{p}_4$. Therefore, $m(\overline{p}) = \rho(\overline{p}) = 1$ satifes (11). Substituting these, we also have (10).

A.2.2 Case 1 with $c \geq \underline{c}$ and $\Delta > \Delta_c$

In this case, Region 4 as well as its boundaries $\overline{p}_3$ and $\overline{p}_4$ need to be modified. Theorem 1 specifies

$$(m(\overline{p}), \nu(\overline{p}), \rho(\overline{p})) = \left(1, 0, \tfrac{\delta_\ell}{\delta_r + \delta_\ell}\right), \tag{33}$$

and $\pi(\overline{p}) = \hat{p}$, for $\overline{p} \in [\overline{p}_3, \overline{p}_4)$, where $\overline{p}_3$ is now set at $\overline{p}_3'$, which satisfies $\hat{V}(\overline{p}_3') = \hat{u}$. We note that this new $\overline{p}_3'$ is smaller than the original $\overline{p}_3$ and is still larger than $\overline{p}_2$.[24] $\overline{p}_4$ is now set at $\overline{p}_4'$, which

[24] Recall $\hat{V}(\cdot)$ is strictly decreasing. Since $\hat{V}(\overline{p}_3) < \hat{u}$ when $c \in (\underline{c}, \overline{c})$ and $\Delta > \Delta_c$, we have $\overline{p}_3' < \overline{p}_3$. We also have $\overline{p}_3' > \overline{p}_2$ since $\hat{V}(\overline{p}_2) = \Phi(\overline{p}_2) > \max\{U_\ell(\overline{p}_2), U_r(\overline{p}_2)\} \geq \hat{u}$.

uniquely satisfies $\underline{p}(\overline{p}_4') = \hat{p}$. One can see that $\overline{p}_4' > \overline{p}_3'$. Given $\sigma(\overline{p})$ for this region, the value of the strategy is given by $V^\sigma(p, \overline{p}) = \hat{u}$.

It is straightforward to verify that $V^\sigma(p, \overline{p}) = \hat{u}$ together with $\pi(\overline{p}) = \hat{p}$ and (33) satisfy (6)–(12) for $\overline{p} \in (\overline{p}_3, \overline{p}_4)$. Indeed, inserting $\pi(\overline{p}) = \hat{p}$ and $V^\sigma(p, \overline{p}) = \hat{u}$ as well as $V^\sigma_p(p, \overline{p}) = 0$ and $V^\sigma_{\overline{p}}(p, \overline{p}) = 0$, the objective in (10) becomes $-(1-m)c$. Hence $m = 1$ is optimal and (10) holds. Since the objective is independent of ν and ρ, (11) also holds. At $\overline{p} = \overline{p}_3$, we have $V^\sigma(p, \overline{p}_-) = \hat{V}^\sigma(\overline{p}_3) = \hat{u}$ and $V^\sigma_p(p, \overline{p}_-) = 0$, and $V^\sigma_{\overline{p}}(p, \overline{p}_-) < 0$, but we can employ the same argument as in Region 3 to show that the coefficients of m and $(1-m)$, in the objective is zero and hence $m = 1$ is optimal.

Finally, we verify (12). Inserting (33), and $V^\sigma(p, \overline{p}_-) = V^\sigma(p, \overline{p}) = \hat{u}$ in the objective in (12), we see that the objective is equal to zero and hence $\pi(\overline{p}) = \hat{p}$ is a minimizer. Finally, $V^\sigma_p(p, \overline{p}) = 0$ implies that (6) holds.

References

Arrow, K. J., Blackwell, D., and Girshick, M. A. (1949). Bayes and minimax solutions of sequential decision problems. *Econometrica*, pages 213–244. 2

Auster, S. and Kellner, C. (2022). Robust bidding and revenue in descending price auctions. *Journal of Economic Theory*, page 105072. 5

Barberis, N. (2012). A model of casino gambling. *Management Science*, 58(1):35–51. 5

Barkley-Levenson, E. E. and Fox, C. R. (2016). The surprising relationship between indecisiveness and impulsivity. *Personality and Individual Differences*, pages 1–6. 27

Beauchêne, D., Li, J., and Li, M. (2019). Ambiguous persuasion. *Journal of Economic Theory*, 179:312–365. 5

Bose, S. and Daripa, A. (2009). A dynamic mechanism and surplus extraction under ambiguity. *Journal of Economic theory*, 144(5):2084–2114. 5

Bose, S. and Renou, L. (2014). Mechanism design with ambiguous communication devices. *Econometrica*, 82(5):1853–1872. 5

Che, Y.-K. and Mierendorff, K. (2019). Optimal dynamic allocation of attention. *American Economic Review*, 109:2993–3029. 2, 4, 25, 28, 29, 22, 23

Cheng, X. (2021). Relative maximum likelihood updating of ambiguous beliefs. *Journal of Mathematical Economics*, page 102587. 34

Cheng, X. and Riedel, F. (2013). Optimal stopping under ambiguity in continuous time. *Mathematics and Financial Economics*, 7(1):29–68. 4

Christensen, S. and Lindensjö, K. (2018). On finding equilibrium stopping times for time-inconsistent markovian problems. *SIAM Journal on Control and Optimization*, 56(6):4228–4255. 5

Christensen, S. and Lindensjö, K. (2020). On time-inconsistent stopping problems and mixed strategy stopping times. *Stochastic Processes and their Applications*, 130(5):2886–2917. 5

Daley, B. and Green, B. (2012). Waiting for news in the market for lemons. *Econometrica*, 80(4):1433–1504. 31

Daley, B. and Green, B. (2020). Bargaining and news. *American Economic Review*, 110(2):428–74. 31

Ebert, S. and Strack, P. (2015). Until the bitter end: On prospect theory in a dynamic context. *American Economic Review*, 105(4):1618–33. 5

Ebert, S. and Strack, P. (2018). Never, ever getting started: On prospect theory without commitment. *Available at SSRN 2765550*. 5, 22

Epstein, L. G. and Ji, S. (2020). Optimal learning under robustness and time-consistency. *Operations Research*. 4, 32

Epstein, L. G. and Schneider, M. (2003). Recursive multiple-priors. *Journal of Economic Theory*, 113(1):1–31. 2

Epstein, L. G. and Schneider, M. (2007). Learning under ambiguity. *The Review of Economic Studies*, 74(4):1275–1303. 4

Fudenberg, D., Strack, P., and Strzalecki, T. (2018). Speed, accuracy, and the optimal timing of choices. *American Economic Review*, 108(12):3651–84. 2, 4

Ghosh, G. and Liu, H. (2021). Sequential auctions with ambiguity. *Journal of Economic Theory*, 197:105324. 5

Gilboa, I. and Schmeidler, D. (1989). Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics*, 18(2):141–153. 2, 7

Gilboa, I. and Schmeidler, D. (1993). Updating ambiguous beliefs. *Journal of Economic Theory*, 59(1):33–49. 34

Henderson, V., Hobson, D., and Tse, A. (2017). Randomized strategies and prospect theory in a dynamic context. *Journal of Economic Theory*, 168:287–300. 5

Ke, S. and Zhang, Q. (2020). Randomization and ambiguity aversion. *Econometrica*, 88(3):1159–1195. 24

Kellner, C. and Le Quement, M. T. (2018). Endogenous ambiguity in cheap talk. *Journal of Economic Theory*, 173:1–17. 5

Miao, J. and Wang, N. (2011). Risk, uncertainty, and option exercise. *Journal of Economic Dynamics and Control*, 35(4):442–461. 4

Moscarini, G. and Smith, L. (2001). The optimal level of experimentation. *Econometrica*, 69(6):1629–1644. 2, 4

Nikandrova, A. and Pancs, R. (2018). Dynamic project selection. *Theoretical Economics*, 13(1):115–143. 4

Osborne, M. J. and Rubinstein, A. (1994). *A Course in Game Theory*. MIT Press. 8

Peskir, G. and Shiryaev, A. (2006). *Optimal stopping and free-boundary problems*. Springer. 4

Riedel, F. (2009). Optimal stopping with multiple priors. *Econometrica*, 77(3):857–908. 4

Saito, K. (2015). Preferences for flexibility and randomization under uncertainty. *American Economic Review*, 105(3):1246–71. 24

Siniscalchi, M. (2011). Dynamic choice under ambiguity. *Theoretical Economics*, 6(3):379–421. 22

Wald, A. (1947). Foundations of a general theory of sequential decision functions. *Econometrica*, pages 279–313. 2, 4

Xu, Z. Q. and Zhou, X. Y. (2013). Optimal stopping under probability distortion. *The Annals of Applied Probability*, 23(1):251–282. 5

Zhong, W. (2019). Optimal dynamic information acquisition. *Available at SSRN 3024275*. 4