Learning and Self-confirming Long-Run Biases*

P. Battigalli,\textsuperscript{a} A. Francetich,\textsuperscript{b} G. Lanzani,\textsuperscript{c} M. Marinacci\textsuperscript{a}

\textsuperscript{a}Department of Decision Sciences and IGIER - Università Bocconi
\textsuperscript{b}Department of Economics - University of Washington Bothell
\textsuperscript{c}Department of Economics - Massachusetts Institute of Technology

14th October 2018

Abstract

We consider an ambiguity averse, sophisticated decision maker facing a recurrent decision problem where information is generated endogenously. In this context, we study self-confirming strategies as the outcomes of a process of active experimentation. We provide inter alia a learning foundation for self-confirming equilibrium with model uncertainty (Battigalli et al., 2015), and we analyze the impact of changes in ambiguity attitudes on convergence to self-confirming equilibria. We identify conditions under which the set of self-confirming equilibrium actions is invariant to changes in ambiguity attitudes. Yet, ambiguity aversion affects the dynamics: We argue that ambiguity aversion tends to stiffen experimentation, increasing the likelihood that decision maker gets stuck into suboptimal “certainty traps.”

1 Introduction

We study the dynamic behavior of a decision maker (DM) who faces a recurrent decision problem in which the actions he selects depend on the information endogenously gathered through his past behavior as, for example, in multiarmed bandit problems (cf. Gittins 1989). We can diagram the flow of actions and information as follows:

*We thank seminar audiences at Chicago, Tel Aviv, UWB, the 2018 NASMES, Emiliano Catonini, Roberto Corrao, Nicodemo De Vito, Ignacio Esponda, Drew Fudenberg, Anh Nguyen, Elia Sartori, Ran Spiegler, and Lorenzo Stanca for useful comments and suggestions. Special thanks go to Simone Cerreia-Vioglio and Fabio Maccheroni with whom this project started. Financial support from the European Research Council (advanced grant 324219), Guido Cazzavillan Scholarship, and the AXA Research Fund is gratefully acknowledged.
We consider an ambiguity averse, finitely patient DM who is uncertain about the data generating process followed by Nature. In this setting, there are three critical elements of our analysis. First, there is an objective stochastic model governing the state evolution. Second, the uncertainty of the DM about the data generating process is represented through a probability measure, a belief, over the possible stochastic models describing the evolution of the states. Given this belief, he evaluates the possible actions according to the smooth ambiguity criterion of Klibanoff et al. (2005). We assume that the DM updates beliefs about models according to Bayes rule. Third, the DM uses a rational strategy given his belief.

It is essential to understand the meaning of the term “rational” in our setting. An uncertainty averse DM may have dynamically-inconsistent preferences (cf. Example 5). While we allow for reversal of preferences, we assume that the DM is sophisticated in the sense that he formulates a dynamically-consistent strategy. Dynamically-consistent strategies are strategies that satisfy the one-shot-deviation property (OSDP): There is no instance where the DM has an incentive to choose an action different from the one prescribed by the given strategy. In a finite-horizon model, this results from folding-back planning; however, we focus on infinite-horizon models to study the limit properties of behavior and beliefs, and to exploit the ensuing stationarity of the dynamic decision problem.

We study how self-confirming strategies arise from an active experimentation process, providing a novel convergence result under a consistency assumption. More precisely, we show that the stochastic process of beliefs and actions converges with probability one to a random limit action-belief pair. This random limit pair satisfies almost surely the following conditions: The limit action maximizes the one-period value given the limit belief, and the limit belief assigns probability one to the set of stochastic models that are observationally equivalent to the actual data generating process given the limit action.

Therefore, even if the DM cares about the future, the limit action-belief pair must be a self-confirming equilibrium of the repeated game played by the DM against nature. Since the belief may only partially identify the true data generating process (nature’s “strategy”), such limit behavior may be very different from the “Nash” (or “rational expectations”) equilibrium, in which the DM plays the objective best reply.\footnote{Our definition of self-confirming equilibrium (also called “conjectural equilibrium”) is broader than the one of Fudenberg and Levine (1993, 1998). See the discussion in Battigalli et al. (2015).}

Since we assume that the state process is exogenous, that is, the DM’s actions cannot influence the probabilities of states in future periods, our framework cannot model long-run interactions where the states are the co-players’ stage-game choices. However, our exogeneity assumption is justified in the game-theoretic setting of large population games. Our setup can represent the point of view of a DM who plays a game recurrently with other agents independently drawn from large populations. The DM recognizes to be unable to influence the evolution of the environment with his actions. The probability models describe the distribution of actions in the co-players’ populations. Experimentation is valuable to the DM since a better understanding of the correct distribution of strategies in co-players’ populations may allow him to select a better strategy in the following periods. In particular, our setting
is consistent with a steady-state learning environment à la Fudenberg and Levine (1993), where individual agents learn through their life, but the population’s statistics are constant.

Under this interpretation, we provide a learning foundation for self-confirming equilibrium with model uncertainty (Battigalli et al. 2015, henceforth BCMM). Specifically, the random limit pair corresponds to the “smooth” self-confirming equilibrium concept of BCMM since the limit action is a myopic best response, and the evidence generated by the limit action and the steady-state distribution of opponents’ strategies confirm the limit belief. Since we show that self-confirming equilibrium emerges as the long-run outcome of an active experimentation and learning process, the comparative statics result of BCMM implies that higher ambiguity aversion reduces the predictability of long-run behavior. At the same time, such limit behavior is more stable under higher ambiguity aversion, since (possibly) ambiguous deviations from the unambiguous tested action can only become less attractive as ambiguity aversion increases.

We also provide conditions under which the BCMM theorem holds as an invariance result: The set of self-confirming equilibrium actions does not depend on ambiguity attitudes. However, ambiguity aversion affects the dynamics of the choice of actions: We argue that ambiguity aversion tends to stifle experimentation, increasing the likelihood that decision maker gets stuck into suboptimal “certainty traps.” The intuition is that the DM observes ex-post the consequences of frequently chosen actions; hence he learns to be approximately certain about the risks (probabilities of consequences) implied by tested actions, whereas he remains uncertain about the risks implied by deviations. Ambiguity aversion biases him toward tested actions — “exploitation” rather than “exploration.” As a consequence, ambiguity aversion makes convergence to objectively optimal behavior (the best reply to the correct model) less likely.

Related Literature Our setup is strictly related to the literature on active learning (or “stochastic control”), and in particular to the seminal work by Easley and Kiefer (1988, henceforth EK). We refer to the working paper version of this work for a formal connection between the two setups. Our paper departs from their analysis in several fundamental aspects. First, we allow for non-neutral ambiguity attitudes and dynamically inconsistent preferences. Second, EK requires the DM to assign positive subjective probability to the correct data generating process, whereas we only assume that at least one model observationally equivalent to the actual one (given the adopted strategy) lies in the support of the DM’s subjective belief. Last but not least, we posit an exogenous process for states of Nature; our DM cannot affect Nature, only what he observes from Nature. Instead, the DM in EK controls Nature directly. Separating feedback from Nature entails additional notation when describing our setup, but being able to define certain probability measures without reference to actions makes it simpler to state and prove our limit results for strategies and beliefs.

2Available as IGIER w.p. 588.
3Moreover, the working paper version of our paper considers the more general case of non i.i.d. data generating processes.
Our definition of self-confirming equilibrium is closely related to the notion of subjective equilibrium (Kalai and Lehrer 1995, henceforth KL). Besides minor details, there are two key differences. While KL only considers Dirac beliefs over probability models, we allow for arbitrary beliefs over probability models. Focusing on Dirac beliefs is without loss of generality under subjective expected utility maximization, but not under non-neutral ambiguity attitudes (see Sections 2.1 and 4). On the other hand, since in KL the set of interacting players is fixed once and for all, their analysis is better suited to capture non-steady-state strategic interaction.

The results on ambiguity aversion and experimentation are consistent with the findings in Li (2017) and Anderson (2011). Li (2017) characterizes the optimal experimentation strategy under ambiguity aversion in an independent K-armed bandit problem. Aside from focusing on K-armed bandit problems, the key difference with our paper is that Li (2017) models ambiguity aversion following the two-stage multiple-prior model of Marinacci (2002), while we employ the smooth ambiguity criterion of Klibanoff et al. (2005). As a result, the comparative-statics analysis in Li (2017) considers the impact of increases in perceived ambiguity, while ours studies the effect of increases in ambiguity attitude. Moreover, Li (2017) uses a recursive version of the maxmin expected-utility criterion and is thus able to employ standard dynamic programming techniques. Such a recursive representation is precluded in our setting. Anderson (2011) derives the predictions of his model under the implicit assumption that the decision maker can ex-ante commit to any strategy. However, the empirical evidence he presents is consistent with the theoretical predictions of our model.

Since we do not assume that the DM assigns positive probability to the correct data generating process, we relate our work to Esponda and Pouzo (2016) in the literature on agents with misspecified models. They show that, even if the beliefs of myopic players in a strategic game do not assign positive probability to models that are observationally equivalent to the correct one, with positive probability they will converge to models that minimize the Kullback-Leibler divergence from the correct one. In our case, said divergence is zero due to our consistency condition (see Definition 1). Furthermore, our convergence occurs with probability one, because we do not consider a strategic framework. However, we generalize in two dimensions: We allow for ambiguity aversion and patience.

Outline  The paper is structured as follows. Section 2 presents the static and dynamic decision framework, as well as some preliminary notation. Section 3 describes the endogenous information process, while section 4 describes the DM’s payoffs. Section 5 presents self-confirming equilibria. Section 6 presents our results on convergence to SCE, while Section 7 presents our comparative dynamics results on ambiguity attitudes. Finally, Section 8 briefly relates our analysis to the literature on learning in games and concludes. We refer to the working paper version for the computations in the examples. All proofs are in the Appendix.

\footnote{In their Online Appendix, Esponda and Pouzo (2016) extend part of their analysis to the non-myopic case. Of course, unlike us, they can rely on standard dynamic programming arguments because they assume ambiguity neutrality.}


2 Framework

2.1 Static environment

Let $S$ be a finite space of states of nature and let $M \subset \mathbb{R}$ be an outcome space. We consider a control setup where a finite set $A$ of actions (or controls) is available to the DM, and actions and states translate into outcomes through a feedback function $f : A \times S \to M$. The triple $(A, S, f)$ is the basic structure of the decision problem.

Given a probability measure $\theta$ on $S$, an action $a$ induces a pushforward measure over outcomes via the function $F : A \times \Delta(S) \to \Delta(M)$

$$(a, \theta) \mapsto F(a, \theta)(\cdot) = \theta \circ f_a^{-1}(\cdot) = \sum_{s \in f_a^{-1}(\cdot)} \theta(s)$$

where $f_a := f(a, \cdot)$ is the section of $f$ at $a$. We assume that the DM is an Expected Utility Maximizer for these (objective) lotteries.

Assumption 1 - Expected Utility on Lotteries. Given an objective probability measure $\theta$ on $S$, there exists an $u : A \times M \to \mathbb{R}$ such that the DM prefers the (objective) distribution over outcomes induced by $a'$ to the one induced by $a''$ if and only if

$$\sum_{m \in M} u(a', m) F(a', \theta)(m) \geq \sum_{m \in M} u(a'', m) F(a'', \theta)(m).$$

Given this, we define the expected payoff

$$R(a, \theta) := \mathbb{E}_{F(a, \theta)}[u]$$

where $u_a := u(a, \cdot)$ is the section of $u$ at $a$. It is often convenient to use the notation $R(a, \theta) = \sum_{s \in S} r(a, s) \theta(s)$, where $r : A \times S \to \mathbb{R}$ is the payoff (or reward) function $r := u_a \circ f_a$.

Let $\Theta \subseteq \Delta(S)$ be a collection of probability measures on $S$ that represents the structural information available to DM, that is, the set of models that are physically possible. We identify $\Theta$ with a subset of the simplex of dimension $|S| - 1$ and endow it with the relative Borel $\sigma$-algebra $\mathcal{B}(\Theta)$. If the probability model $\theta$ is unknown, namely, if there is model uncertainty (cf. Marinacci 2015), the DM ranks actions according to the smooth ambiguity criterion of Klibanoff et al. (2005):

$$V(a, \mu) := \phi^{-1} \left( \int_\Theta \phi(R(a, \theta)) \mu(d\theta) \right),$$

where $\mu$ is a prior probability measure on $(\Theta, \mathcal{B}(\Theta))$, and $\phi$ is a strictly increasing and continuous real-valued function that describes attitudes towards ambiguity. In particular, a

---

5 The working paper considers the more general case of an arbitrary space of consequences.

6 We endow all finite sets with the discrete topology.

7 For example, if $S = \{b, g\}$ is the set of possible colors of the ball drawn from an urn of 90 balls that are either black or green, then $\Theta = \{\theta \in \Delta(S) : \theta(b) = i/90 = 1 - \theta(g), i \in \{0, ..., 90\}\}$.

8 See Cerreia-Vioglio et al. (2013a) for an axiomatization.
concave $\phi$ captures ambiguity aversion, while a linear $\phi$ (e.g., the identity) corresponds to the classical subjective expected utility criterion (Cerreia-Vioglio et al. 2013b):

$$\hat{V}(a, \mu) = \int_{\Theta} R(a, \theta) \mu(d\theta) = \sum_{s \in S} r(a, s) \theta_\mu(s) = R(a, \theta_\mu),$$

where $\theta_\mu \in \Theta$ is the predictive probability given by $\theta_\mu(E) := \int_{\Theta} \theta(E) \mu(d\theta)$ for all $E \subseteq S$. Finally, note that:

(i) When the support of $\mu$, $\text{supp} \mu$, is a singleton $\{\theta\}$, criterion (1) reduces to the expected payoff criterion $R(a, \theta)$;

(ii) The limit case of criterion (1) as ambiguity aversion increases is a version of the maximin criterion $\min_{\theta \in \text{supp} \mu} R(a, \theta)$ of Gilboa and Schmeidler (1989).

The static decision problem can be summarized by

$$\Gamma = (A, S, M, \Theta, f, u, \phi, \mu).$$

We conclude the description of the static environment by noting that, if we endow $\Delta(\Theta)$ with the topology of weak convergence of measures, the value function $\hat{V}$ is continuous in beliefs.

Lemma 1. For every $a \in A$, the functional $\hat{V}(a, \cdot)$ is continuous.

### 2.2 Dynamic environment

**Notation** For every set $Z$, we let $Z^t = \prod_{\tau=1}^{t} Z$ and $Z^\infty = \prod_{t=1}^{\infty} Z$. We endow the space $Z^\infty$ with the Borel $\sigma$-algebra, $\mathcal{B}(Z^\infty)$, corresponding to the product topology on $Z^\infty$; when $Z$ is finite, this is the same as the $\sigma$-algebra generated by the elementary cylinders $\{z_1\} \times \cdots \times \{z_t\} \times Z \times \cdots$ (see, e.g., Proposition 1.3 in Folland 2013). We denote by $z^t = (z_1, ..., z_t) \in Z^t$ both the histories and the elementary cylinders that they identify through the map

$$(z_1, ..., z_t) \mapsto \{z_1\} \times \cdots \times \{z_t\} \times Z \times \cdots.$$  

We denote by $z^\infty = (z_1, ..., z_t, ...)$ a generic element of $Z^\infty$.

**Environment** Given $S$, let $(S^\infty, \mathcal{B}(S^\infty))$ be the measurable space on which a coordinate state process $(s_1, s_2, ...)$ is defined, with $s_t : S^\infty \to S$ for each $t$. We will use the less demanding notation $s^\infty$ for the state process describing the exogenous uncertainty in the decision problem. Its realizations are denoted by $s^\infty \in S^\infty$. To ease notation, we set $s^t = (s_1, ..., s_t)$.

---

5To map our decision criterion into theirs, let their space of consequences be $A \times M$ and identify each action $a$ with the act $g(a) = (a, f(a, s))$.

10Unless otherwise stated, it is understood that $t$ is an element of $\mathbb{N}$, the set of natural numbers. We use interchangeably the terms ‘time’ and ‘period’ to refer to $t$.

11We use boldface letters for random variables and normal letters for realizations.
For a generic stochastic process \((z_1, z_2, \ldots)\) defined on \((S^\infty, \mathcal{B}(S^\infty))\) we denote by \(\sigma (z^t)\) the \(\sigma\)-algebra generated by the random variables \(z_1, \ldots, z_t\), namely, by the process up to time \(t\).

Finally, for every \(\theta \in \Theta\), we define \(p_\theta \in \Delta (S^\infty)\) as the unique i.i.d. extension on \(\mathcal{B}(S^\infty)\) of the measure given on all elementary cylinders by

\[
p_\theta(s^t \times S \times \ldots) = \prod_{\tau=1}^{t} \theta(s_{\tau})
\]

for every \(t\) and every \(s^t\) in \(S^t\).

**Actions and outcomes** We describe the DM’s choices as a sequence \((a_t) \in A^\infty\) that consists of an action \(a_t\) for each time \(t\). At each such \(t\), there is a time-independent feedback function \(f : A \times S \to M\); \(f(a_t, s_t)\) is the outcome that the DM receives *ex-post* (i.e., after the decision) at time \(t\) if he chooses action \(a_t\) and state \(s_t\) obtains.

**Information feedback** In a dynamic setting, the outcome the DM observes provides feedback about past states. This feedback is a source of “endogenous” (choice dependent) information. Its relevance is peculiar to the dynamic setting and will play a key role in the paper. By selecting action \(a_t \in A\) at time \(t\), the DM observes ex-post the outcome \(m_t = f(a_t, s_t)\) if state \(s_t\) realizes. A DM who selects action \(a_t\) and ex-post receives the outcome \(m_t\) knows that the true state \(s_t\) belongs to the set \(\{ s \in S : f(a_t, s) = m_t \} = f_{a_t}^{-1}(m_t)\).

In general, ex-post information about the state is typically endogenous; that is, the partition

\[
\{ f^{-1}_a (m) : m \in M \} \subseteq 2^S
\]

of the state space \(S\) induced by the outcomes may depend on the choice of action \(a\). If the DM receives the same information about states regardless of his action, namely, if

\[
\forall a, a' \in A, \{ f^{-1}_a (m) : m \in M \} = \{ f^{-1}_{a'} (m) : m \in M \},
\]

we say that feedback satisfies *own-action independence*. In particular, there is perfect feedback when the DM ex-post observes the realized state \(s_t\); that is, if \(f_a\) is injective for each \(a \in A\).

Actions and outcomes are remembered: At each period \(t > 1\), the *ex-ante* endogenous information — that is, the endogenous information gathered prior to the period-\(t\) decision — is given by the history of outcomes \(m^{t-1} = (m_1, \ldots, m_{t-1})\) received in the previous periods as a result of the history of actions \(a^{t-1} = (a_1, \ldots, a_{t-1})\) and states \(s^{t-1} = (s_1, \ldots, s_{t-1})\).

**Example 1** (Prelude). Consider an urn that contains black (\(B\)), green (\(G\)), and yellow (\(Y\)) balls. At each time \(t\), the DM is asked to bet \(1\) euro on the color of the ball that will be drawn

---

\[^{12}\text{We refer to the working paper version of this paper for a more general setting that separates outcomes and feedback.}\]

\[^{13}\text{We distinguish three points in time within each period: the ex-ante time (before the decision), the decision time, and the ex-post time (after the decision). Any information available ex-post at period } t \text{ is also available ex-ante at } t + 1.\]
from the urn; therefore the possible bets are blue ($b$), green ($g$), and yellow ($y$). Suppose that, ex-ante, as in the classical Ellsberg’s paradox, the DM is told that one-third of the balls are black (and that the only possible colors are $B$, $G$, and $Y$). That is, the set of posited models is $\Theta = \{ \theta \in \Delta (\{B,G,Y\}) : \theta(B) = 1/3 \}$. Ex post, after the draw, he only learns the result of his bet, namely, whether or not he wins 1 euro. Here, $S = \{B,G,Y\}$, $A = \{b,g,y\}$, and $M = \{0,1\}$. The feedback function is described in the following table:

<table>
<thead>
<tr>
<th>$f$</th>
<th>$B$</th>
<th>$Y$</th>
<th>$G$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$b$</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$y$</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>$g$</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Therefore, we have:

\[
\begin{align*}
    f_b^{-1}(1) &= \{B\}, & f_b^{-1}(0) &= \{Y,G\}, \\
    f_y^{-1}(1) &= \{Y\}, & f_y^{-1}(0) &= \{B,G\}, \\
    f_g^{-1}(1) &= \{G\}, & f_g^{-1}(0) &= \{B,Y\}.
\end{align*}
\]

Note that own-action independence is violated: Ex post, betting on $b$ yields the partition $\{\{B\}, \{Y,G\}\}$ of $S$, while the bets on $y$ and $g$ respectively yield the partitions $\{\{Y\}, \{B,G\}\}$ and $\{\{G\}, \{B,Y\}\}$.

**Example 2** (Two-Arm Bandit). There are two urns, $I$ and $II$, with black and green balls. The DM chooses an urn, say $k$, and wins 1 euro if the ball drawn from urn $k$ is green ($G_k$, good outcome from urn $k$) and zero if it is black ($B_k$, bad outcome from urn $k$). The outcome for the chosen urn is observed ex-post. Here, $S = \{B_1B_{II}, B_1G_{II}, G_1B_{II}, G_1G_{II}\}$, $A = \{I, II\}$, and $M = \{0,1\}$. The following table describes the feedback function

<table>
<thead>
<tr>
<th>$f$</th>
<th>$B_1B_{II}$</th>
<th>$B_1G_{II}$</th>
<th>$G_1B_{II}$</th>
<th>$G_1G_{II}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$I$</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$II$</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Therefore:

\[
\begin{align*}
    f_I^{-1}(1) &= \{G_1B_{II}, G_1G_{II}\}, & f_I^{-1}(0) &= \{B_1B_{II}, B_1G_{II}\}, \\
    f_{II}^{-1}(1) &= \{B_1G_{II}, G_1G_{II}\}, & f_{II}^{-1}(0) &= \{B_1B_{II}, G_1B_{II}\}.
\end{align*}
\]

Own-action independence of feedback does not hold.

### 2.3 Strategies and information

**Strategies** At each period $t$, the overall ex-ante information available to the DM is given by the histories of actions and outcomes, $a^{t-1}$ and $m^{t-1}$. The ex-ante information history $h_t$ at time $t$ is given by:

\[
h_1 = (a^0, m^0); \quad \forall t > 1, h_t = (a^{t-1}, m^{t-1}) = (h_{t-1}, a_{t-1}, m_{t-1}),
\]
where \((a^0, m^0)\) represents null data. Hence, the ex-ante information history space \(H_{t+1}\) at the beginning of period \(t + 1\), determined by information about previous periods, is
\[
H_{t+1} = \{ (a^t, m^t) \in A^t \times M^t : \exists s^t \in S^t, \forall k \in \{1, \ldots, t\}, m_k = f(a_k, s_k) \}.
\]

By definition, \(H_1 = \{ (a^0, m^0) \}\).

Strategies specify an action for each possible information history. Thus, they are modelled as sequences \(\alpha = (\alpha_t)\) of functions, with \(\alpha_t : H_t \rightarrow A\) for each \(t\). Since \(H_1 = \{ (a^0, m^0) \}\) is a singleton, the first term \(\alpha_1\) prescribes a non-contingent action.

Information and strategies  A state process \(s^\infty\) and a strategy \(\alpha = (\alpha_t)\) recursively induce an action process \((a^t_\alpha)\), an outcome process \((m^t_\alpha)\), and an information process \(h^\alpha = (h^t_\alpha)\), as follows:

(i) \(a^0_1 = \alpha_1 (a^0, m^0)\) and \(a^t_\alpha = \alpha_t (h^t_\alpha)\) for each \(t > 1\);

(ii) \(m^t_\alpha = f(a^t_\alpha, s_t)\) for each \(t\);

(iii) \(h^0_\alpha = (a^0, m^0)\) and \(h^t_{\alpha,t-1} = (h^0_\alpha, a^t_\alpha, m^t_\alpha)\) for each \(t\).

In words, at each period \(t\), an action \(a_t\) is selected according to the time-\(t\) strategy \(\alpha_t\) based on the information history \(h_t = (h_{t-1}, a_{t-1}, m_{t-1})\). In turn, its execution generates an outcome \(m_t\) that the DM may consider in subsequent periods. Note that \(\alpha_1\) prescribes only one action, \(\alpha_1(s^0, m^0)\), which, together with realization \(s_1\) of \(s_1\), initializes the recursion by sending outcome \(m_1\).

The sequence of \(\sigma\)-algebras \((\sigma(h^t_\alpha))\) on \(S^\infty\) generated by the information process \((h^t_\alpha)\) is a filtration that describes the information structure generated and used by strategy \(\alpha\). Since feedback will typically not be perfect, this filtration is coarser than the one generated by the state process \(s\); that is, \(\sigma(h^t_\alpha) \subseteq \sigma(s^{t-1})\) for each \(t > 1\). For this reason, without loss of generality, we can regard \(h^t_\alpha\) as well as \(a^t_\alpha\) and \(m^t_{\alpha,t-1}\), as functions defined on \(S^{t-1}\)\(^{14}\).

Each finite history \(h_t = (a^{t-1}, m^{t-1})\) corresponds to cylinder
\[
I(h_t) = \{ s^\infty \in S^\infty : \forall \tau \in \{1, \ldots, t-1\}, f(a_\tau, s_\tau) = m_\tau \}.
\]

This is the information about the realized sequence of states revealed by \(h_t\).

Since states are not directly observed, we can focus on processes \((a^t_\alpha)\), \((m^t_\alpha)\), and \((h^t_\alpha)\), keeping the underlying parameterized probability space \((S^\infty, p_\alpha)\) in the background. We write events in terms of the processes observable by DM. In particular,
\[
[h^t_{\alpha,t+1} = (a^t, m^t)] = \begin{cases} I(a^t, m^t), & \text{if } \forall \tau \in \{1, \ldots, t-1\}, \alpha_\tau(a^{\tau-1}, m^{\tau-1}) = a_\tau, \\ \emptyset, & \text{otherwise.} \end{cases}
\]

\(^{14}\)Recall that \(\sigma(s^0)\) is the trivial \(\sigma\)-algebra.
Example 3 (Act I). Assume that only bets on either black or yellow are possible, not on green. As a result, we now have $A = \{b, y\}$ and the table in the Prelude becomes:

<table>
<thead>
<tr>
<th></th>
<th>B</th>
<th>Y</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>b</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>y</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

Throughout we will consider two strategies, denoted by $\alpha^{NE}$ (No Experimentation) and $\alpha^E$ (Experimentation). Strategy $\alpha^{NE}$ dictates betting on black forever; strategy $\alpha^E$ dictates experimenting with yellow in period 1, and, from period 2 onwards, the action prescribed depends on the result of this experimentation: If a success is observed in period 1, $y$ is chosen forever; otherwise $b$ is chosen forever. Formally:

**Strategy $\alpha^{NE}$**: For each $h_t = (a^{t-1}, m^{t-1})$,

$$
\alpha^{NE}_t (h_t) = \begin{cases} 
  b & \text{if } t = 1, \\
  y & \text{if } t > 1, \text{ and } (y, 1) \in \{(a_1, m_1), \ldots, (a_{t-1}, m_{t-1})\}, \\
  b & \text{if } t > 1, \text{ and } (y, 1) \notin \{(a_1, m_1), \ldots, (a_{t-1}, m_{t-1})\}.
\end{cases}
$$

(Of course, to assess deviations, the strategy must specify actions to be taken at histories that the strategy itself excludes, such as what to do after having bet on yellow.)

By always betting on black, the DM cannot observe the relative frequencies of $Y$ and $G$. In particular, for each period $t$ and state history $s^{t-1}$,

$$a^{\alpha^{NE}}_t (s^{t-1}) = b,$$

$$m^{\alpha^{NE}}_t (s^t) = \begin{cases} 
  1 & \text{if } s_t = B, \\
  0 & \text{if } s_t \in \{Y, G\},
\end{cases}$$

$$h_t^{\alpha^{NE}} (s^t) = \begin{cases} 
  (h_t^{\alpha^{NE}} (s^{t-1}), b, 1) & \text{if } s_t = B, \\
  (h_t^{\alpha^{NE}} (s^{t-1}), b, 0) & \text{if } s_t \in \{Y, G\}.
\end{cases}$$

**Strategy $\alpha^E$**: For each $h_t = (a^{t-1}, m^{t-1})$,

$$\alpha^E_t (h_t) = \begin{cases} 
  y & \text{if } t = 1, \\
  y & \text{if } t > 1, \text{ and } (y, 1) \in \{(a_1, m_1), \ldots, (a_{t-1}, m_{t-1})\}, \\
  b & \text{if } t > 1, \text{ and } (y, 1) \notin \{(a_1, m_1), \ldots, (a_{t-1}, m_{t-1})\}.
\end{cases}
$$

The only difference between this strategy and $\alpha^{NE}$ is the action chosen in the first period. Next we describe the induced processes of actions and outcomes:

$$a_1^{\alpha^E} = y,$$

$$m_1^{\alpha^E} (s_1) = \begin{cases} 
  1 & \text{if } s_1 = Y, \\
  0 & \text{if } s_1 \in \{B, G\},
\end{cases}$$
\[ h_2^E(s_1) = \begin{cases} (y, 1) & \text{if } s_1 = Y, \\ (y, 0) & \text{if } s_1 \in \{B, G\}, \end{cases} \]

and, for each \( t > 1 \) and \( s^{t-1} \),

\[ a_t^E(s^{t-1}) = \begin{cases} y & \text{if } s_1 = Y, \\ b & \text{else}, \end{cases} \]

\[ m_t^E(s_t) = \begin{cases} 1 & \text{if } s_1 = Y \text{ and } s_t = Y, \\ 1 & \text{if } s_1 \in \{B, G\} \text{ and } s_t = B, \\ 0 & \text{else}, \end{cases} \]

\[ h_{t+1}^E(s') = \begin{cases} (h_t^E(s^{t-1}), y, 1) & \text{if } s_1 = Y \text{ and } s_t = Y, \\ (h_t^E(s^{t-1}), b, 1) & \text{if } s_1 \in \{B, G\} \text{ and } s_t = B, \\ (h_t^E(s^{t-1}), y, 0) & \text{if } s_1 = Y \text{ and } s_t \in \{B, G\}, \\ (h_t^E(s^{t-1}), b, 0) & \text{if } s_1 \in \{B, G\} \text{ and } s_t \in \{Y, G\}. \end{cases} \]

\[ \widehat{ } \]

3 Models and Learning

3.1 Distributions and updating

Predictive and posterior probabilities A measure \( \mu : \mathcal{B}(\Theta) \rightarrow [0, 1] \) with finite support is called a \textit{prior probability}. The set of finitely-supported Borel probability measures on \( \Theta \) is denoted by \( \Delta(\Theta) \). A prior induces a \textit{predictive distribution} \( p_\mu \in \Delta(S^\infty) \) defined by \( p_\mu(B) = \int_\Theta p_\theta(B) \mu(d\theta) \) for all \( B \in \mathcal{B}(S^\infty) \). Moreover, for each \( t \), we denote by \( \mu(\cdot \mid h_t) \) the \textit{posterior} of \( \mu \) given information history \( h_t \). Formally, for each \( \theta \in \Theta \),

\[ \mu(\theta \mid h_t) = \frac{p_\theta(I(h_t))}{p_\mu(I(h_t))}, \]

provided that \( p_\mu(I(h_t)) > 0 \). In view of this, the \textit{conditional predictive distribution} \( p_\mu(\cdot \mid h_t) \) is such that, for each \( B \in \mathcal{B}(S^\infty) \),

\[ p_\mu(B \mid h_t) = \int_\Theta p_\theta(B \mid h_t) \mu(d\theta \mid h_t). \]

One may be surprised by our use of Bayesian updating for an ambiguity sensitive DM. Bayesian updating has a revealed-preference foundation in SEU axioms for preferences over strategies. However, if we do not allow for commitment, we cannot rely on a revealed-preference approach to justify Bayesian updating in this setting: Strategies cannot be chosen, only actions can be chosen. This impossibility of commitment to a particular strategy is critical in a context where preferences are allowed to be dynamically inconsistent.

\[ ^{15} \text{We abbreviate } (\cdot \mid h_t^\omega = h_t) \text{ in } (\cdot \mid h_t). \]
By sticking to Bayesian updating, we can maintain in this dynamic setting the separation between ambiguity attitudes and the perception of ambiguity of the static KMM decision criterion that is, instead, lost if we consider dynamically consistent rules for updating beliefs. Our approach complements the analysis of the dynamic choices of a KMM decision maker by Hanany et al. (2018). In their work, they maintain dynamic consistency of the preferences of the DM but consider a different updating rule for beliefs.

**Observationally equivalent models** Given a strategy \( \alpha \) and any \( p_\theta \) for \( \theta \in \Theta \), denote by \( p_\theta^\alpha : \sigma ( h^\alpha ) \to [0,1] \) its restriction to the sigma algebra \( \sigma ( h^\alpha ) \) generated by the \( \alpha \)-observable events. Fix a true model \( \bar{\theta} \); the \( \sigma ( h^\alpha ) \)-measurable correspondence \( \Theta^\alpha_\mu ( \bar{\theta} ) : S^{t-1} \to 2^\Theta \) represents the collection of models that are deemed possible and that, conditional on \( h^\alpha_t ( s^{t-1} ) \), are **observationally equivalent** to the true model \( \bar{\theta} \) under \( \alpha \) and prior \( \mu \). Formally:

\[
\Theta^\alpha_\mu ( \bar{\theta} ) ( s^{t-1} ) \equiv \{ \theta \in \text{supp} \mu ( | h^\alpha_t ( s^{t-1} ) | ) , p^\alpha_\theta ( | h^\alpha_t ( s^{t-1} ) | ) = p^\alpha_\bar{\theta} ( | h^\alpha_t ( s^{t-1} ) | ) \}. 
\]

Note that, for some \( s^{t-1} \), the set \( \Theta^\alpha_\mu ( \bar{\theta} ) ( s^{t-1} ) \) may be empty if \( \bar{\theta} \notin \text{supp} \mu \).

The next lemma establishes a monotonicity property of this correspondence. We introduce the following abuse of notation/terminology: When a property holds \( p_\theta \text{-almost surely} \), we will simply say that said property holds \( \theta \text{-almost surely} \).

**Lemma 2.** For every \( \bar{\theta} \in \Theta \) and every period \( t \), \( \Theta^\alpha_\mu ( \bar{\theta} ) \subseteq \Theta^\alpha_{\mu+1} ( \bar{\theta} ) \ \bar{\theta} \text{-almost surely.} \)

The intuition behind the lemma is as follows. The set \( \Theta^\alpha_\mu ( \bar{\theta} ) ( s^{t-1} ) \) may contain models that disagree with \( \bar{\theta} \) on the relative probabilities of past events (up to \( t-1 \)), but that agree with \( \bar{\theta} \) on the relative probabilities of future events (from \( t \)). Almost surely, every model that agrees with \( \bar{\theta} \) on future events conditional on information up to \( t-1 \) also agrees on future events conditional on information up to \( t \). Act II of our running Example will show that the inclusion is not an equality.

It follows from the lemma that, \( \bar{\theta} \text{-a.s.,} \)

\[
\Theta^\alpha_1 ( \bar{\theta} ) \equiv \{ \theta \in \text{supp} \mu : p^\alpha_\theta = p^\alpha_{\bar{\theta}} \} \subseteq \Theta^\alpha_\mu ( \bar{\theta} )
\]

for every \( t \). The set \( \Theta^\alpha_1 ( \bar{\theta} ) \) represents the irreducible model uncertainty that, when \( \bar{\theta} \) is the true model, the DM faces if he plays \( \alpha \) and holds belief \( \mu \). When \( \Theta^\alpha_1 ( \bar{\theta} ) = \text{supp} \mu \), such uncertainty and strategy do not allow any learning, as all the models that the DM initially deems possible are \( \alpha \)-observationally equivalent to the true model. The opposite is true when \( \Theta^\alpha_1 ( \bar{\theta} ) = \{ \bar{\theta} \} \) since in this case the DM will assign probability arbitrarily close to 1 to the true model as he accumulates observations.

---

16 See Hanany and Klibanoff (2009), and Maccheroni et al. (2006).

17 It is actually enough to require \( p^\alpha_\theta ( E | h^\alpha_t ( s^{t-1} ) ) = p^\alpha_{\bar{\theta}} ( E | h^\alpha_t ( s^{t-1} ) ) \) for all \( E \in \cup_{t \geq 1} \sigma ( h^\alpha_t ) \). That is, observational equivalence is determined by the \( \alpha \)-observable events.

18 In this work, we use the term “belief” to denote the probability assessment over (stochastic) models. Using the terminology of Marinacci (2015), this belief represents how the DM addresses epistemic uncertainty, whereas models capture the (perceived) physical uncertainty.
In what follows, we will often study properties of a triple \((\alpha, \mu, \tilde{\theta})\), where \(\alpha\) is the strategy carried out by the DM, \(\mu\) is his prior belief over models at period 0, and the probability measure \(p_{\tilde{\theta}}\) on \((S^\infty, \mathcal{B}(S^\infty))\) is the “correct” model of the data generating process. Given this, we study the triple \((\alpha, \mu, \tilde{\theta})\) to understand what happens to a DM who follows strategy \(\alpha\) when his prior is \(\mu\) and the data generating process is \(\tilde{\theta}\). Therefore, we have a particular interest in statements that hold \(\tilde{\theta}\)-a.s., that is, that are almost surely true for the correct model. The notion of observationally equivalent models motivates the following definition.

**Definition 1.** A triple \((\alpha, \mu, \tilde{\theta})\) is consistent at time \(t\) if \(\Theta^\alpha_{t, \mu} (\tilde{\theta}) \neq \emptyset \tilde{\theta}\)-a.s.

In words, the triple \((\alpha, \mu, \tilde{\theta})\) is consistent at time \(t\) if, conditional on the available information \(h^\alpha_t\), at least one model deemed possible is \(\alpha\)-observationally equivalent to the true model\(^{19}\)

Let 
\[
\sigma_{\geq t}(h^\alpha_{(s^t-1)}) = \{ E \subseteq S^\infty : I(h^\alpha_{(s^t-1)}) \times E \in \sigma(h^\alpha) \}
\]
denote the sigma-algebra of \(\alpha\)-observable events from date \(t\) onwards given \(s^t-1\). Then:

\[
\Theta^\alpha_{t, \mu} (\tilde{\theta}) (s^{t-1}) = \{ \theta \in \text{supp} \mu (\cdot | h^\alpha_{(s^{t-1})}) : \forall E \in \sigma_{\geq t} (h^\alpha_{(s^{t-1})}) , p^\alpha_{\theta} (E) = p^\alpha_{\tilde{\theta}} (E) \} .
\]

Hence, \((\alpha, \mu, \tilde{\theta})\) is consistent at \(t\) if, for \(\tilde{\theta}\)-almost every \(s^{t-1}\), there exists some \(\theta \in \text{supp} \mu (\cdot | h^\alpha_{(s^{t-1})})\) such that \(p^\alpha_{\theta} (E) = p^\alpha_{\tilde{\theta}} (E)\) for all \(E \in \sigma_{\geq t} (h^\alpha_{(s^{t-1})})\). Of course, an obvious sufficient condition for consistency is that \(\mu(\tilde{\theta}) > 0\).

Given Lemma\(^2\) for a triple \((\alpha, \mu, \tilde{\theta})\) it is easier to meet the condition for consistency as \(t\) gets larger. We denote by \(T = T(\alpha, \mu, \tilde{\theta})\) the smallest \(t\) for which the triple is consistent if such \(t\) exists; in this case, we say that the triple is **consistent from period** \(T\). We begin by showing that, under our consistency assumption, beliefs converge almost surely.

**Lemma 3.** Let \((\alpha, \mu, \tilde{\theta})\) be consistent from some period \(T\). Then, the process \(\mu (\cdot | h^\alpha_t)\) converges \(\tilde{\theta}\)-a.s.

For the next result, fix \(s^\infty \in S^\infty\) and denote by \(\alpha^\infty(s^\infty)\) the set of actions that are played infinitely often under \(\alpha\) along the path \(s^\infty\).

**Proposition 1.** Let \((\alpha, \mu, \tilde{\theta})\) be consistent from some period \(T\). Then, \(a \in \alpha^\infty(s^\infty)\) implies
\[
\lim_{t \to \infty} \mu(\{ \theta \in \Theta : F(a, \theta) = F(a, \tilde{\theta}) \} | h^\alpha_t (s^\infty)) = 1 \quad \text{for } \tilde{\theta}\text{-almost every } s^\infty. \tag{3}
\]

In words, a triple \((\alpha, \mu, \tilde{\theta})\) that is consistent from some period \(T\) allows the DM to learn in the long run the \(\alpha\)-observable implications of the true model \(\tilde{\theta}\).

Correspondence \(\theta \mapsto \Theta^\alpha_{T(\alpha, \mu, \tilde{\theta}), \mu} (\theta)\) can be viewed as the (long-run) model-identification map determined by \(\alpha\) and \(\mu\). We have **perfect model identification** when \(\Theta^\alpha_{T(\alpha, \mu, \tilde{\theta}), \mu} (\theta) = \{ \theta \}\)

\(^{19}\)The word “consistent” may remind the reader of the consistency criterion imposed in Arrow and Green (1973). However, theirs is an “existence of equilibrium condition” requiring that, given any DM’s action and true model, there exists a subjective model conceivable by the DM that is observationally equivalent to the actual one.
for each \( \theta \in \Theta \); in this case, the DM who holds belief \( \mu \) and plays strategy \( \alpha \) learns the true model in the long run. Otherwise, we have partial model identification: Even in the long run, the DM is only able to asymptotically identify a collection of possible models.

Perfect model identification occurs, for instance, under perfect feedback: If past states are observable, the true model is asymptotically identified. This is the classical result of Doob (1949); it is enough to recall that, under perfect feedback,

\[
\alpha = \beta = 1.
\]

**Corollary 1.** Let \( (\alpha, \mu, \tilde{\theta}) \) be consistent from period \( T = 1 \). Under perfect feedback,

\[
\mu (\tilde{\theta} | \cdot) \to 1 - \tilde{\theta} \text{ a.s.}
\]

**Example 4 (Act II).** Suppose that the DM:

1. knows that \( 1/3 \) of the balls are black (and so all his models \( \theta \) are such that \( \theta (B) = 1/3 \));
2. has a 3-point prior \( \mu \) with \( \text{supp} \mu = \{ \theta^Y, \theta^\text{uni}, \theta^G \} \) and believes it is equally likely that the true model is either \( \theta^Y \) (with \( \theta^Y (Y) = 2/3 \)), the uniform model \( \theta^\text{uni} \), or \( \theta^G \) (with \( \theta^G (G) = 2/3 \)):

<table>
<thead>
<tr>
<th>Marginals</th>
<th>( B )</th>
<th>( Y )</th>
<th>( G )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \theta^Y )</td>
<td>( \frac{1}{3} )</td>
<td>( \frac{2}{3} )</td>
<td>0</td>
</tr>
<tr>
<td>( \theta^\text{uni} )</td>
<td>( \frac{1}{3} )</td>
<td>( \frac{1}{3} )</td>
<td>( \frac{1}{3} )</td>
</tr>
<tr>
<td>( \theta^G )</td>
<td>( \frac{1}{3} )</td>
<td>0</td>
<td>( \frac{2}{3} )</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Prior</th>
<th>( \tilde{\theta}^Y )</th>
<th>( \tilde{\theta}^\text{uni} )</th>
<th>( \tilde{\theta}^G )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \mu )</td>
<td>( \frac{1}{3} )</td>
<td>( \frac{1}{3} )</td>
<td>( \frac{1}{3} )</td>
</tr>
</tbody>
</table>

By requiring to always bet on the color with known proportion, strategy \( \alpha^{NE} \) does not allow the DM to learn anything. Formally,

\[
\mu (\cdot | h^{\alpha^{NE}}_t) = \mu (\cdot).
\]

Here, \( T(\alpha^{NE}, \mu, \tilde{\theta}) = 1 \) and \( \Theta^{\alpha^{NE}, \mu} (\tilde{\theta}) = \text{supp} \mu \); strategy \( \alpha^{NE} \) only allows partial identification. For strategy \( \alpha^{E} \), if \( \tilde{\theta} \in \text{supp} \mu \), \( T(\alpha^{E}, \mu, \tilde{\theta}) = 2 \). To see why this is the case, note that \( \text{supp} \mu (\cdot | (y, 1)) = \{ \theta^Y, \theta^\text{uni} \} \), \( \text{supp} \mu (\cdot | (y, 0)) = \{ \theta^Y, \theta^\text{uni}, \theta^G \} \), and

\[
\Theta^{\alpha^{E}, \mu} (\tilde{\theta}) (s_1) = \begin{cases} \{ \tilde{\theta} \} & \text{if } h^{E}_2 (s_1) = (y, 1), \\ \{ \theta^Y, \theta^\text{uni}, \theta^G \} & \text{if } h^{E}_2 (s_1) = (y, 0). \end{cases}
\]

Note that this example shows that the inclusion in Lemma 2 can be strict. If \( \tilde{\theta} = \theta^Y \),

\[
\Theta^{\alpha^{E}, \mu} (\theta^Y) = \{ \theta^Y \} \subset \{ \theta^Y, \theta^\text{uni}, \theta^G \} = \Theta^{\alpha^{E}, \mu} (\theta^Y) (y).
\]

By Proposition 1,

\[
\mu (\cdot | h^{\alpha^{E}}_t) \to \begin{cases} \delta_{\tilde{\theta}} & \text{if } h^{E}_2 (s_1) = (y, 1), \\ \mu (\cdot | (y, 0)) & \text{if } h^{E}_2 (s_1) = (y, 0). \end{cases}
\]

where \( \delta_{\tilde{\theta}} \) denotes the Dirac measure on \( \tilde{\theta} \). If experimentation succeeds, the true model is asymptotically learned. Otherwise, if \( h_2 = (y, 0) \), posterior beliefs attain their limit value as early as the second period, and the DM remains in the dark. ▲


4 Value

If there is no model uncertainty, we posit that the DM ranks alternative strategies accordingly to the standard Discounted Expected Utility criterion.

**Assumption 1' - Discounted Expected Utility on Lotteries.** Given an objective probability measure over state $\theta$ and a history $h_t$ with $p_\theta(I(h_t)) > 0$, there are $r : A \times S \to \mathbb{R}$ and $\delta \in [0,1)$ such that the DM prefers strategy $\alpha$ to strategy $\beta$ if and only if

$$\sum_{\tau=t}^{\infty} \delta^{\tau-t} \sum_{h_r \in H_r} R(\alpha(h_\tau), \theta) p_\theta(h_\tau|h_t) \geq \sum_{\tau=t}^{\infty} \delta^{\tau-t} \sum_{h_r \in H_r} R(\beta(h_\tau), \theta) p_\theta(h_\tau|h_t).$$

Under model uncertainty, we assume smooth ambiguity preferences. If $h_t$ is observed, the DM ranks strategy $\alpha$ given prior $\mu$ according to the present value of the continuation stream of utility certainty equivalents:

$$V(\alpha, \mu | h_t) := \sum_{\tau=t}^{\infty} \delta^{\tau-t} \phi^{-1}\left(\int_{\Theta} \phi\left(\sum_{h_r \in H_r} R(\alpha(h_\tau), \theta) p_\theta(h_\tau|h_t)\right) \mu(d\theta | h_t)\right).$$

This criterion ranks instant payoffs according to the smooth ambiguity model and then aggregates over time their (utility) certainty equivalents through discounting. Therefore, (utility) smoothing over time is irrelevant. Indeed, when a DM evaluates two continuation streams of utility certainty equivalents, he is interested only in their discounted sum, not on their variability over time.

In particular, we obtain:

(i) $V(\alpha, \mu | h_t) = \sum_{\tau=t}^{\infty} \delta^{\tau-t} \sum_{h_r \in H_r} R(\alpha(h_\tau), \theta_{\mu(h_t)}(h_r|h_t)) p_{\theta_{\mu(h_t)}}(h_r|h_t)$ when $\phi$ is linear;

(ii) $V(\alpha, \mu | h_t) = \sum_{\tau=t}^{\infty} \delta^{\tau-t} \sum_{h_r \in H_r} R(\alpha(h_\tau), \theta) p_\theta(h_r|h_t)$ when $\text{supp} \mu = \{\theta\}$.

The following simple Lemma will be useful.

**Lemma 4.** For every strategy $\alpha$, prior $\mu$, and history $h_t$ such that $p_\mu(I(h_t)) > 0$, the functional $V(\alpha, \cdot | h_t)$ is continuous.

With this lemma, we obtain the following additional corollary to Proposition 1.

**Corollary 2.** Let $(\alpha, \mu, \theta)$ be consistent from period $T$. Then, $\bar{\theta}$-a.s.,

$$|V(\alpha, \mu | h^0_t) - V(\alpha, \delta_{\bar{\theta}} | h^0_t)| \to 0$$

This corollary tells us that, in a consistent triple, the strategy becomes unambiguous on path in terms of value. Of course, this result does not imply that $\mu(\cdot | h^0_t) \to \delta_{\bar{\theta}}$ but only that the present value of the strategy that is used converges to the true value. In particular, even in the limit, alternative strategies may entail unknown consequences. Therefore, this

---

20 We abbreviate $(h^0_{r-1} = h_{r-1}) | h_t)$ in $(h_{r-1}|h_t)$.
corollary support the assumption made in BCMM that only equilibrium strategies have to be unambiguous.

Note that, except for the benchmark case of ambiguity neutrality, this time-additive value function does not admit a recursive formulation. This is related to the well-known dynamic inconsistency of decision makers with non-neutral attitudes toward ambiguity. For this reason, we are not allowed to use many of the standard dynamic programming results. We provide an example of these inconsistencies in our setting.

**Example 5** (Dynamic inconsistency). Consider a modified version of our running example. There are only two periods. Only bets on either black or yellow are possible, not on green. However, by paying a small cost $\varepsilon$, it is also possible to bet on black and observe the color of the selected ball (action $bo$). Finally, we normalize payoffs as $u(a,m) = m_1$ and we describe $f$ with the following table (* means “no direct observation of the color”):

<table>
<thead>
<tr>
<th></th>
<th>B</th>
<th>Y</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>$b$</td>
<td>1, *</td>
<td>0, *</td>
<td>0, *</td>
</tr>
<tr>
<td>$y$</td>
<td>0, *</td>
<td>1, *</td>
<td>0, *</td>
</tr>
<tr>
<td>$bo$</td>
<td>$1 - \varepsilon, B$</td>
<td>$0 - \varepsilon, Y$</td>
<td>$0 - \varepsilon, G$</td>
</tr>
</tbody>
</table>

Suppose that the DM:

1. knows that 1/3 of the balls are black (and so all her models $\theta$ are such that $\theta(B) = 1/3$);
2. believes it is equally likely that the true model is either $\overline{\theta}^Y$ or $\overline{\theta}^G$:

<table>
<thead>
<tr>
<th>Marginals</th>
<th>$B$</th>
<th>$Y$</th>
<th>$G$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\overline{\theta}^Y$</td>
<td>$\frac{1}{3}$</td>
<td>$\frac{5}{12}$</td>
<td>$\frac{1}{4}$</td>
</tr>
<tr>
<td>$\overline{\theta}^G$</td>
<td>$\frac{1}{3}$</td>
<td>$\frac{1}{3}$</td>
<td>$\frac{5}{12}$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Prior</th>
<th>$\overline{\theta}^Y$</th>
<th>$\overline{\theta}^G$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu$</td>
<td>$\frac{1}{2}$</td>
<td>$\frac{1}{2}$</td>
</tr>
</tbody>
</table>

Let $\phi(u) = -e^{-10u}$. Then, the ex-ante optimal strategy if $\varepsilon$ is sufficiently small is:

$\beta$: “Bet on black observing the color at $t = 1$. For $t = 2$, given yellow in the first period, bet on yellow, otherwise bet on black.”

The ex-ante value of strategy $\beta$ is

$$V (\beta, \mu, | (a^0, m^0)) = 0.3 - \varepsilon + \delta 0.3364.$$

However, $\beta$ does not satisfy the one-shot deviation property. Indeed, after having observed yellow, the DM prefers to bet on black. The posterior belief after having chosen $bo$ and having

\footnote{Note that this is not a proper strategy since it does not assign an action to every information history. In particular, it does not assign an action to personal histories ruled out by the strategy itself. However, the specification of the actions selected at those information histories is irrelevant in determining ex-ante optimality.}

\footnote{We refer to the working paper version for the computations used to obtain formulas in the examples.}
observed yellow is:

<table>
<thead>
<tr>
<th>Posterior</th>
<th>$\tilde{\theta}_Y$</th>
<th>$\tilde{\theta}_G$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu(\cdot</td>
<td>(bo, Y))$</td>
<td>$\frac{5}{8}$</td>
</tr>
</tbody>
</table>

Hence,

$$V(\beta, \mu(\cdot | (bo, Y)) | (bo, Y)) \approx 0.3207 < \frac{1}{3},$$

where $\frac{1}{3}$ is the value of always betting on black.

This is a typical example of dynamically-inconsistent preferences. At period 0, for sufficiently small $\varepsilon$, the DM would want to commit to conditioning his behavior on the observed draw. In particular, he would like to choose $y$ if the draw in the first period is $Y$, that is, after history $(bo, Y)$. Indeed, even if betting on yellow leads to ambiguous consequences, the DM is confident that with high probability, if $\tilde{G}$ is the true model, $Y$ will not be the first-period draw. Therefore, even under model $\tilde{G}$, this strategy presents a moderate expected value. However, after having observed $(bo, Y)$, even if the posterior probability of $\tilde{G}$ is lower, the DM considers the consequences of choosing action $y$ too ambiguous. Indeed, the expected value under model $\tilde{G}, 1/4$, is quite small. Therefore, since the DM is highly ambiguity averse, he will select $b$ (or $bo$).

Moreover, it can be shown that the strategy “always bet on black” has a lower ex-ante value $(1 + \delta)/3$, but satisfies the one-shot deviation property. A sophisticated DM will not pay the cost $\varepsilon$ anticipating that he will not condition his behavior on the observed outcome, even if this conditioning is ex-ante optimal. ▲

The conditional value $V(\alpha, \mu | h^\beta_t)$ is a $\sigma(h^\beta_t)$-measurable function of the state process. We have the following (increasingly) special cases: for every $(\alpha, \mu, t),$

(i) under own-action independence of feedback, for every strategy $\beta$ such that coincides with $\alpha$ from period $t$, $V(\alpha, \mu | h^\alpha_t) = V(\beta, \mu | h^\beta_t)$ almost surely, because updated beliefs depend only on the realized states not on the chosen actions;

(ii) under perfect feedback, there is a one-to-one correspondence between information histories $h_t$ and past histories of states $s^{t-1}$, and so $V(\alpha, \mu | h^\alpha_t)$ can be written as $V(\alpha, \mu | s^{t-1})$.

4.1 Stationary Strategies

A strategy $\alpha$ is stationary if, given the prior $\mu$, it depends on history only through the induced posterior belief; that is,

$$\forall t, t' \in \mathbb{N}, \forall h_t \in H_t, \forall h'_{t'} \in H_{t'}, \mu(\cdot | h_t) = \mu(\cdot | h'_{t'}) \Rightarrow \alpha(h_t) = \alpha(h'_{t'}).$$

Note that stationarity is a property of the pair $(\alpha, \mu)$ of strategy and prior.

The following result shows that when the DM uses a stationary strategy, the value function depends on the history only through beliefs. Hence, without loss of generality, we can write $V(\alpha, \mu)$ to denote the evaluation of a stationary strategy $\alpha$ under beliefs $\mu$. 

17
Lemma 5. For all $\alpha$ and $\mu$ satisfying stationarity, if $h_t$ and $h'_t$ are two histories such that $p_\mu(I(h_t))$ and $p_\mu(I(h'_t))$ are strictly positive, then $\mu(\cdot | h_t) = \mu(\cdot | h'_t)$ implies $V(\alpha, \mu|h_t) = V(\alpha, \mu|h'_t)$.

When the preferences of the DM are dynamically consistent, Hinderer (1970) proves that the restriction to stationary strategies is without loss of generality in an i.i.d. environment, that is, the DM cannot achieve a strictly higher value by using a non-stationary strategy.

We cannot adopt this approach “as is” because it relies on a notion of global optimality that under dynamic inconsistency preferences may violate incentive compatibility. However, Proposition 2 in the next section provides an analogous result for the notion of rationality used in this paper. Given this, in the rest of the analysis we will consider only stationary strategies. Therefore, with an abuse of notation, we will regard $\alpha$ as a function of beliefs over probability models.

5 Self-Confirming Equilibrium

5.1 Steady-state analysis

We introduce the notion of self-confirming equilibrium (SCE) in our framework. The key feature of SCE is that the recommended action is a myopic best reply to confirmed beliefs.

Definition 2. A triple $(a^*, \mu^*, \tilde{\theta}) \in A \times \Delta(\Theta) \times \Theta$ is an SCE if:

(i) $\mu^* \left( \{ \theta \in \Theta : F(a^*, \theta) = F(a^*, \tilde{\theta}) \} \right) = 1$;

(ii) $a^* \in \arg \max_{a \in A} \phi^{-1} \left( \int_{\Theta} \phi(R(a, \theta)) \mu^*(d\theta) \right)$.

We say that $a^*$ is an SCE action given $\tilde{\theta}$ if it is part of an SCE $(a^*, \mu^*, \tilde{\theta})$, and that $a^*$ is a Nash equilibrium action given $\tilde{\theta}$ if $a^* \in \arg \max_{a \in A} R(a, \tilde{\theta})$.

The second condition says that $a^*$ is a (myopic, or one-period) best response to $\mu^*$ given the ambiguity attitudes determined by $\phi$. The first condition is a self-confirming property adapted to the static framework: The distribution of outcomes that the DM “observes” in the long run if he always plays $a^*$ is precisely what he expects it to be. In this sense, $a^*$ is unambiguous for $\mu^*$; since payoffs are observable, the self-confirming property implies that the expected distribution of payoffs coincides with the one implied by the true model $\tilde{\theta}$.

Remark 1. If $(a^*, \mu^*, \tilde{\theta})$ is a SCE, then $R(a^*, \theta) = E_{F(a^*, \theta)}[u_{a^*}]$ is constant over supp$\mu^*$.

---

23Building on the results of Dynkin (1965).
24We refer the interested reader to the working paper version for more general results for non-stationary strategies in a non i.i.d. setting.
5.2 Rational learning dynamics

While the DM faces a recursive choice problem, the notion of SCE characterizes behavior and beliefs after the latter have “converged.” In other words, the data provided by the equilibrium strategy does not lead to any further updating because the models that the DM deems possible in an SCE cannot be distinguished from each other or from the true model.

In dynamic settings, we may be interested not only in behavior after beliefs have become “stationary,” but also in behavior as the DM is learning from the data. To this end, we introduce the following notation. For any information history $h_t = (a_1, m_1, ..., a_{t-1}, m_{t-1})$ and action $a$, let $\alpha/(h_t, a)$ be the strategy that behaves as specified by $h_t$ at information histories that precede $h_t$ (namely, at the empty sequence $(a^0, m^0)$ and each $h_{\tau} = (a_1, m_1, ..., a_{\tau-1}, m_{\tau-1})$ for $\tau < t$), selects action $a$ at information history $h_t$, and coincides with $\alpha$ thereafter.

**Definition 3.** A pair $(\alpha, \mu)$ is rational if it satisfies stationarity and, for every period $t$, information history $h_t$, and action $a$,

$$p_\mu(I(h_t)) > 0 \Rightarrow V(\alpha, \mu \mid h_t) \geq V(\alpha/(h_t, a), \mu \mid h_t).$$

Given stationarity, this condition is the one-shot deviation property, which says that — for every information history $h_t$ that the DM deems reachable with positive probability — action $\alpha(\mu(\cdot \mid h_t))$ maximizes the continuation value conditional on $h_t$ given that he will follow $\alpha$ in the future. The motivation is the following: Strategy $\alpha$ is a plan formulated by a sophisticated DM who understands his sequential incentives. In each period $t$, the DM only controls the action in that period, and therefore we require that he maximizes his value with respect to what he can control, given the predicted behavior of his “future selves,” i.e., his continuation strategy. If the time horizon is finite, then this condition is equivalent to folding-back planning. When the DM is ambiguity neutral — that is, when $\phi$ is positively affine —, the one-shot deviation principle implies that strategy $\alpha$ is subjectively optimal given $\mu$.

**Proposition 2.** For every prior $\mu$ there exists a strategy $\alpha$ such that $(\alpha, \mu)$ is rational.

A strategy $\alpha$ is said to be rational given $\mu$ if the pair $(\alpha, \mu)$ is rational. We illustrate the concept of rationality in our running example.

**Example 6 (Act III).** If we normalize payoffs as $u(a, 0) = 0$ and $u(a, 1) = 1$, we have $r = f$. Outcomes are thus the bets’ payoffs. Moreover, we assume that $\phi(u) = -e^{-\lambda u}$, so that higher (absolute) ambiguity aversion corresponds to higher $\lambda$ (see Klibanoff et al. 2005).

Suppose that the DM features the prior $\mu$ presented in Act II. We consider the strategies $\alpha^{NE}$ and $\alpha^{E}$ presented there. The former strategy involves no experimentation as it recommends always betting on black, the color with the known proportion. Thus, the value of this strategy is independent of histories and beliefs, and it is given by

$$V(\alpha^{NE}, \mu|h_t) = \frac{1/3}{1 - \delta}.$$  

\footnote{This holds only for histories allowed by the strategy, namely on path.}
The second strategy recommends betting on $y$ at $t = 1$ and then switching to $b$ permanently if and only if this first bet is unsuccessful. If the DM chooses $y$, the outcomes are informative about the distribution, and he updates his beliefs. Recall by Act II that the DM has a uniform 3-point prior $\mu$ with $\text{supp} \mu = \{\theta^Y, \theta^u, \theta^G\}$. If we denote $\mu(\cdot|h_t) := \left(\mu(\theta^u|h_t), \mu(\theta^Y|h_t), \mu(\theta^G|h_t)\right)$, the posterior is

$$
\mu(\cdot | y, 1) = \left(\frac{1}{3}, \frac{2}{3}, 0\right)
$$

if the outcome is $Y$, and

$$
\mu(\cdot | y, 0) = \left(\frac{1}{3}, \frac{1}{6}, \frac{1}{2}\right)
$$

otherwise.

After the first period, strategy $\alpha^E$ recommends a fixed action. Thus, the continuation problem is stationary, with beliefs as states. For any history $h_t$ ($t > 1$) that induces belief $\mu(\cdot|h_t)$, the continuation-value function after a success in period 1 is:

$$
V(\alpha^E, \mu|h_t) = \frac{\phi^{-1} \left( \mu \left( \theta^u|h_t \right) \phi(1/3) + \mu \left( \theta^Y|h_t \right) \phi(2/3) \right)}{1 - \delta}.
$$

For the initial history, we have:

$$
V(\alpha^E, \mu) = \phi^{-1} \left( \frac{1}{3} \phi(1/3) + \frac{1}{3} \phi(2/3) + \frac{1}{3} \phi(0) \right) + \frac{\delta}{1 - \delta} \phi^{-1} \left( \frac{1}{3} \phi \left( \frac{1}{3} \right) + \frac{1}{3} \phi \left( \frac{5}{9} \right) + \frac{1}{3} \phi \left( \frac{1}{3} \right) \right).
$$

Two forces affect the option value of experimentation: ambiguity aversion (the higher the value of $\lambda$, the lower the value of experimentation) and patience (the higher the value of $\delta$, the higher the value of experimentation). Given this, strategy $\alpha^{NE}$ is preferred if either $\delta = 0$ or $\lambda$ is high enough given $\delta > 0$; if so, the pair $(\alpha^{NE}, \mu)$ is rational. As for strategy $\alpha^E$, if $\delta$ is sufficiently high and $\lambda$ is low enough, e.g., $\lambda = 1$ and $\delta = 0.39$, strategy $\alpha^E$ satisfies the one-shot deviation property at $(a^0, m^0)$. However, because of experimentation, we need to consider two different contingencies.

1. If experimentation is successful (i.e., $s_1 = Y$), the DM learns that model $\theta^G$ is false and updates his belief from $(1/3, 1/3, 1/3)$ to $(1/3, 2/3, 0)$. Moreover, at every subsequent period, the updating rule implies that the posterior will be of the form $(1 - k, k, 0)$, with $k \in (0, 1)$. At this point, the strategy recommends sticking to $y$. It can be checked that this recommendation is better than trying out $b$ once before switching to $y$ thereupon, that is, it satisfies the one-shot deviation property: For all $\delta \in (0, 1)$ and all $\lambda > 0$,

$$
V(\alpha^E, (1 - k, k, 0)) = \frac{\phi^{-1} \left( (1 - k) \phi(\theta^u(Y)) + k \phi(\theta^Y(Y)) \right)}{1 - \delta}
$$

$$
= \frac{\phi^{-1} \left( (1 - k) \phi(1/3) + k \phi(2/3) \right)}{1 - \delta}
$$

$$
> \frac{1}{3} + \delta \phi^{-1} \left( (1 - k) \phi(1/3) + k \phi(2/3) \right)
$$

$$
= V(\alpha^E/b, ((1 - k), k, 0)).
$$
2. If experimentation is unsuccessful (i.e., \( s_1 \in \{ B, G \} \)), the posterior lowers the weight of model \( \theta^Y \) relative to models \( \theta^\text{uni} \) and \( \theta^G \), so that \( p_\mu (Y \mid (y, 0)) < p_\mu (B \mid (y, 0)) = 1/3 \). Thereupon, strategy \( \alpha^E \) recommends switching (and sticking) to black, so that the continuation value is the same as that under \( \alpha^\text{NE} \). Moreover, since betting on black does not lead to any further updating, it is enough to check the inequality with second-period beliefs. For sufficiently small \( \delta \), or for sufficiently high \( \lambda \),

\[
V \left( \alpha^Y, \left( \frac{1}{3}, \frac{1}{6}, \frac{1}{2} \right) \right) = \frac{1}{3} \frac{1}{1-\delta} > V \left( \alpha^E \mid y, \left( \frac{1}{3}, \frac{1}{6}, \frac{1}{2} \right) \right).
\]

In particular, this inequality holds with \( \lambda = 1 \) and \( \delta = 0.39 \), and we have already argued that \( (\alpha^E, \mu) \) satisfies the one-shot deviation property at the initial history; therefore, \( (\alpha^E, \mu) \) is rational. ▲

6 Convergence to SCE

We are interested in studying the limit behavior of rational DMs. In particular, we investigate the conditions that imply convergence to SCE. Building on Lemma 3, we provide a learning foundation to the concept proposed by BCMM. We say that the stochastic process of actions and beliefs \((a_t, \mu (\cdot | h_t))\) converges to an SCE given \( \tilde{\theta} \) if, for \( \tilde{\theta} \)-almost every \( s^\infty \), beliefs converge to a limit \( \mu^\alpha_{s^\infty} \), and there exists a finite time \( t \) such that \((a^\alpha_{\tau}(s^{\tau-1}), \mu^\alpha_{s^{\tau}}, \tilde{\theta})\) forms an SCE for all \( \tau \geq t \). Note that the tail sequence of actions \((a^\alpha_{\tau}(s^{\tau-1}))_{\tau \geq t}\) is not required to be constant.

**Proposition 3.** Let \((\alpha, \mu)\) be rational. If \((\alpha, \mu, \tilde{\theta})\) is consistent from some period \( T \), the stochastic process of actions and beliefs \((a^\alpha_{s}, \mu (\cdot | h^\alpha_{s}))\) converges to a SCE given \( \tilde{\theta} \).

The intuition is as follows. Since the action set \( A \) is finite, for every \( s^\infty \) there is a corresponding time \( \tilde{T}_s^\infty \) after which every action chosen by \( \alpha \) is played infinitely often. Under the stated assumptions, beliefs converge almost surely to a random limit \( \mu^\alpha_{s^\infty} \). Thus, each action chosen by \( \alpha \) in the long run must be a myopic best reply to the limit belief. This holds because the updated beliefs converge and the value of experimentation is vanishing for actions played infinitely often. Proposition [3] implies that, for \( \tilde{\theta} \)-almost every \( s^\infty \), the limit belief \( \mu^\alpha_{s^\infty} \) assigns probability 1 to the models that induce the same probabilities over consequences of \( \tilde{\theta} \) given the actions played in the long run. Therefore, for every action \( a^* \) chosen by \( \alpha \) in the long run, \((a^*, \mu^\alpha_{s^\infty}, \tilde{\theta})\) must be an SCE.

Note that the realized sequence of actions \((a^\alpha_{s}(s^{\tau-1}))\) converges if there is a unique myopic best reply to the limit belief \( \mu^\alpha_{s^\infty} \), but such uniqueness is not guaranteed. If the myopic best reply is unique, say \( a^* \), the action sequence \((a^\alpha_{s}(s^{\tau-1}))\) is eventually constant at \( a^* \) and \((a^*, \mu^\alpha_{s^\infty}, \tilde{\theta})\) is an SCE. Moreover, after a finite time, the agent chooses an action that maximizes the one-period value given the current beliefs (and not only limit one); that is, exploration (experimentation) becomes irrelevant, all that matters is exploitation. Formally:
Proposition 4. Let \((\alpha, \mu)\) be rational and let \((a_t^\alpha, \mu(\cdot|h_t^\alpha))\) converge to an SCE on path \(s^\infty\) given \(\bar{\theta}\). If
\[
\arg\max_{a \in A} \phi^{-1}\left(\int_{\Theta} \phi(R(a, \theta)) \mu^\alpha_{s^\infty}(d\theta)\right) = \{a^*\}
\]
for some \(a^* \in A\), then there exist \(T\) such that, for every \(t > T\),
\[
a_t^\alpha(s_{t-1}) = a^* \in \arg\max_{a \in A} \phi^{-1}\left(\int_{\Theta} \phi(R(a, \theta)) \mu(d\theta|h_t^\alpha(s_{t-1}))\right).
\]

It is important to stress that this convergence neither implies that the limit belief is the Dirac measure supported by the correct model, nor that the limit action is the objective myopic best reply. However, the limit pairs of beliefs and actions almost surely satisfy the standard properties of stochastic limits in the (expected utility) stochastic control limit literature. Indeed, the realization \((a^*, \mu^\alpha_{s^\infty})\) is such that we have:

- **(Confirmed Beliefs):** \(\mu^\alpha_{s^\infty}\) assigns probability 1 to the models that are observationally equivalent to the true \(\bar{\theta}\) given \(a^*\), (see, Proposition 1);

- **(Subjective Myopic Best Reply):** Even if the discount factor is strictly positive, the agent maximizes his one-period value. That is, exploitation prevails over exploration.

In contrast, in a Nash equilibrium beliefs are correct (i.e., \(\mu = \delta_{\bar{\theta}}\)) and the action played is an Objective Myopic Best Reply. A sufficient condition for convergence to an SCE that is also a Nash equilibrium is to have own-action independence.

**Corollary 3.** Suppose that there is own-action independence. If \((\alpha, \mu, \bar{\theta})\) is consistent from some period \(T \geq 1\), and \((\alpha, \mu)\) is rational, then the stochastic action process \((a_t^\alpha)\) converges to a Nash equilibrium action given \(\bar{\theta}\).

Note that own-action independence guarantees convergence to a Nash equilibrium under observable payoffs, a maintained assumption in this work. If we relax this hypothesis, the stronger condition of perfect feedback is needed.

Our running example illustrates how the true data generating process may be unidentified in the limit.

**Example 7 (Act IV).** Consider the strategy \(\alpha^E\) of the previous acts. Again, recall by Act II that the DM has a uniform 3-point prior \(\mu\) with \(\text{supp} \mu = \{\theta^Y, \theta^{uni}, \theta^G\}\). In Act II, we have shown that \((\alpha^E, \mu, \bar{\theta})\) is consistent from period 2, whereas, in Act III, we have proved that with parameters \(\lambda = 1\) and \(\delta = 0.39\), \((\alpha^E, \mu)\) is rational. We can show how our convergence result obtains in this specific case. Suppose that \(\bar{\theta} = \theta^Y\). From Lemma 3, we have that beliefs converge. In particular:

\[
\mu^E_{s^\infty} = \left(\mu^{\alpha E}_{s^\infty}(\theta^{uni}), \mu^{E}_{s^\infty}(\theta^Y), \mu^{E}_{s^\infty}(\theta^G)\right) = \begin{cases} 
(1/3, 1/6, 1/2) & \text{if } s_1 \in \{B, G\}, \\
(0, 1, 0) & \text{if } s_1 = Y.
\end{cases}
\]
Indeed, if experimentation is unsuccessful, the posterior of $\mu$ lowers the weight of model $\theta^Y$ relative to models $\theta^\text{uni}$ and $\theta^G$; thereupon, strategy $\alpha^E$ recommends switching (and sticking) to black, and so there is no additional updating. On the other hand, if the experimentation is successful, strategy $\alpha^E$ prescribes sticking to yellow thereupon, and then the correct model $\theta^Y$ is asymptotically identified.

If $s_1 \in \{B,G\}$, for every $t > 1$,
$$a_t^{\alpha^E}(s^{t-1}) = b$$
for every $t > 1$, and $(b, (1/3, 1/6, 1/2), \theta^Y)$ is the SCE that obtains in the limit. Note that in this case the DM will end up choosing an *objectively sub-optimal* action.

If $s_1 = Y$,
$$a_t^{\alpha^E}(s^{t-1}) = y,$$
for every $t > 1$, and $(y, (0, 1, 0), \theta^Y)$ is the SCE that obtains in the limit. Indeed, it is immediate to see that these actions maximize one-period value for limit beliefs and that the distribution of probabilities over outcomes confirms them.

Finally, consider strategy $\alpha^{NE}$. In Act III, we have argued that $(\alpha^{NE}, \mu)$ is rational if the DM is sufficiently ambiguity averse. In this case, regardless of the correct marginal $\hat{\theta} \in \Theta = \{\theta^Y, \theta^\text{uni}, \theta^G\}$, we have almost sure convergence to an SCE from period 1. Indeed, the DM sticks to black from the first period onwards, and black is the myopic best reply to the confirmed prior $\mu = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$. However, note that if the correct model is $\theta^Y$, betting on black is objectively sub-optimal.

### 7 Comparative dynamics for ambiguity aversion

#### 7.1 Certainty traps

Act III of our running example suggests that as ambiguity aversion increases, experimentation becomes less attractive. Actions that induce the same probabilities of payoffs under all the models that the DM deems possible are appealing under ambiguity aversion, but generate no new information about the underlying probability model. To obtain evidence on the correct model, the DM has to choose an action that will potentially induce a different probability measure over payoffs under the different models he deems possible — that is, he has to experiment. An ambiguity averse DM is inclined to avoid such ambiguous actions. For notational simplicity, in this section we will assume that the predictive probability $p_\mu$ induced by prior $\mu$, is absolutely continuous with respect to $p_0$.

Fix an arbitrary belief $\nu$, we say that action $a$ is $\nu$-unambiguous if $F(a, \theta) = F(a, \theta')$ for every $\theta, \theta' \in \text{supp}\, \nu$. In words, an action $a$ is unambiguous given the DM’s beliefs if all models entertained by the DM assign the same outcome probabilities given $a$. Otherwise, we

---

26 Since we are interested in the dynamic effect of ambiguity on convergence to SCE, this assumption is without loss of generality. Indeed, convergence to SCE implies that the posterior will be absolutely continuous with respect to the correct model from some period $t$ onwards. Therefore, we may be shift the time index and relabel period $t$ as period 0, obtaining the same results.
say that $a$ is $\nu$-ambiguous. The next proposition shows that, if a strategy $\alpha$ is rational given $\mu$ for an ambiguity-neutral DM and prescribes unambiguous actions, then every strategy $\beta$ that is rational given $\mu$ for a strictly ambiguity averse DM must also prescribe unambiguous actions. In other words, if experimentation is not rational for an ambiguity neutral DM, then it cannot be rational for an ambiguity averse DM with the same beliefs. Given prior $\mu$, for any stationary strategy $\alpha$, we call belief-range of $(\alpha, \mu)$ the set of beliefs that the DM may hold with positive probability under $p_\mu$, that is, the set

$$\{\mu(\cdot|h_t) : t\in\mathbb{N}, p_\mu([h_t^\nu = h_t]) > 0\}.$$

**Proposition 5.** Let $(\alpha, \mu)$ be rational under ambiguity neutrality and let $(\beta, \mu)$ be rational under strict ambiguity aversion. Then, for every belief $\nu$ in the belief-range of both $(\alpha, \mu)$ and $(\beta, \mu)$, if $\alpha(\nu)$ is $\nu$-unambiguous, the same holds for $\beta(\nu)$.

In what follows, we restrict our attention to cases where there is a unique $\mu$-ambiguous action. Although this assumption is restrictive, it encompasses interesting stochastic control problems such as two-armed bandit problems with a safe arm. The next proposition establishes that, given a prior $\mu$, the sequence of actions and posteriors induced by a strategy $\beta$ that is rational under strict ambiguity aversion will almost surely converge faster to an SCE (given the true model) than the sequence of actions and posterior beliefs corresponding to a strategy $\alpha$ that is rational under ambiguity neutrality. The intuition is as follows. Under the stated assumptions, the decision problem amounts to deciding how long to experiment, choosing the unique ambiguous action, say $a^*$. Assuming that convergence occurs, there are two possibilities: either the DM never stops experimenting, or he stops at some finite time $t$, choosing ever after the best unambiguous action. In the first case, convergence to an SCE occurs (typically) at infinity. In the second case, it occurs at the stopping time. By Proposition 5 if an ambiguity neutral DM prefers to stop at time $t$, then a strictly ambiguity averse DM prefers to stop as well. Alternatively, if an ambiguity neutral DM chooses $a^*$ forever, an ambiguity averse DM either does the same, or starts choosing the best unambiguous action from some period $t$ onwards and is henceforth trapped in a selfconfirming equilibrium.

**Proposition 6.** Assume that there is a unique $\mu$-ambiguous action $a^*$; let $(\alpha, \mu)$ be rational under ambiguity neutrality and let $(\beta, \mu)$ be rational under strict ambiguity aversion; furthermore, assume that $(\alpha, \mu, \bar{\theta})$ is consistent from some $T$. Then, $\bar{\theta}$-almost surely, the action-belief process $(\beta(\cdot|h^\beta_t), \mu(\cdot|h^\beta_t))$ converges to an SCE (given $\bar{\theta}$) at least as fast as the action-belief process $(\alpha(\cdot|h^\alpha_t), \mu(\cdot|h^\alpha_t))$.

The next proposition shows that an ambiguity averse DM is less likely than an ambiguity neutral DM to eventually play the Nash equilibrium (i.e., objectively optimal) action.

**Proposition 7.** Assume that there is a unique ambiguous action $a^*$ given $\mu$. Let $(\alpha, \mu)$ and $(\beta, \mu)$ be rational under ambiguity neutrality and strict ambiguity aversion, and consistent

---

27 We refer to Battigalli et al. (2018) for an in-depth analysis of the role of this assumption.

28 Additional reasons to focus on this case are provided by Proposition 8.
from some $T_{\alpha}$ and $T_{\beta}$, respectively. Then, $\theta$-almost surely, if \( (a^\beta_t) \) converges to a Nash Equilibrium action (given $\bar{\theta}$), so does \( (a^\alpha_t) \).

We illustrate the previous results in the classical setup introduced by Rothschild (1974) where a monopolist trades-off exploration of the demand curve for his good against exploitation using the price that (subjectively)-maximizes one period profit.

**Example 8.** A monopolist who is uncertain about the demand for its product faces a new customer each period. The cost of producing a unit is $c > 0$. The monopolist can charge either a low price, $p_L$, or a high price, $p_H$, where $p_H > p_L > c \geq 0$. Each new customer has a reservation price in \{ $p_H, p_L, 0$ \}, and he buys the product if and only if his reservation price is weakly larger than the ask price. If the price is set to $p_i$ and a sale is made, the monopolist makes $p_i - c$ in profit; otherwise, he makes 0. Thus, we are considering a build to order production. Here, $A = \{p_L, p_H\}$; $M = \{0, 1\}$ and $S = \{p_H, p_L, 0\}$,

\[
f(a, s) = \begin{cases} 
1 & a \leq s, \\
0 & a > s,
\end{cases}
\]

and

<table>
<thead>
<tr>
<th>Payoff $u(p,m)$</th>
<th>$m = 1$</th>
<th>$m = 0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p = p_L$</td>
<td>$p_L - c$</td>
<td>0</td>
</tr>
<tr>
<td>$p = p_H$</td>
<td>$p_H - c$</td>
<td>0</td>
</tr>
</tbody>
</table>

Imagine that according to his prior $\mu$, the monopolist believes that $\theta = (\theta(p_H), \theta(p_L), \theta(0))$ is either $\theta_1 := (0.8, 0.1, 0.1)$, or $\theta_2 := (0, 0.9, 0.1)$, and that these two models are equally likely. Here, the monopolist is certain that if he posts the low price he sells with probability 0.9, but he is uncertain about the selling probability at a high price, which is — therefore — a $\mu$-ambiguous action.

Moreover, suppose that the correct model is $\theta_1$ and that $0.9(p_L - c) < 0.8(p_H - c)$. Our previous results imply that an ambiguity averse monopolist will stop experimentation with the high price earlier and that he will be more likely to be trapped in the (objectively) suboptimal SCE where he posts the low price (the $\mu$-unambiguous action).

At first sight, the previous results may seem surprising. Indeed, if a DM is ambiguity averse, why does he not experiment more, so as to eliminate (or reduce) the uncertainty about the true model? This reasoning tacitly relies on a different notion of experimentation. It is true that, typically, an ambiguity averse DM is willing to pay more to eliminate the model uncertainty (see e.g., Anderson 2011). However, in an active-learning setting, the DM cannot simply buy information about the true model; learning happens only when actions with ambiguous probabilities of consequences are chosen. Since an ambiguity averse DM dislikes those actions, he will end up resolving less ambiguity than his ambiguity neutral counterpart.

The results above relate to the findings in Anderson (2011). On the theoretical side, his Theorem 1 for two-armed bandits with a safe arm strictly relates to our Proposition 5.
main difference is that Anderson implicitly assumes the possibility to commit to a strategy. Indeed, the Gittins indices used in that paper characterize the *ex-ante optimal* strategy for a decision maker, that is, the strategy that maximizes the value at the initial history. However, like us, he assumes the DM performs Bayesian updating, a feature that paired with ambiguity aversion induces dynamically inconsistent preferences.

On the experimental side, the theoretical predictions of our model are consistent with the findings presented in Anderson (2011): the behavior of an SEU maximizer cannot explain joint data about willingness to pay for information about the stochastic process characterizing the ambiguous arm and the amount of experimentation that is performed. In particular, the resulting experimentation is too low, which is the prediction of our model under ambiguity aversion.

Another reason to focus on the case of a unique ambiguous action is to illustrate how our analysis adds to BCMM. Indeed, the following proposition shows that the set of SCE actions is *invariant* with respect to the (positive) degree of ambiguity aversion captured by the (concave) function $\phi$. Yet, as Propositions [6] and [7] show, ambiguity aversion has “dynamic” effects on the persistence of experimentation and the distribution of long-run outcomes.

**Proposition 8.** Suppose that there is a unique $\mu$-ambiguous action $a^*$. Then, for every concave and strictly increasing $\phi, \phi'$ and every action $a$, if $(a, \mu, \bar{\theta})$ is an SCE under ambiguity attitudes $\phi$, then $a$ is an SCE action (given $\bar{\theta}$) under ambiguity attitudes $\phi'$.

Our running example illustrates. Since there the unique ambiguous action is to bet on yellow, Proposition 8 implies that the monotonicity result of BCMM (the SCE set is weakly increasing in the degree of ambiguity aversion) holds vacuously: the set of equilibria is not affected by ambiguity attitudes. Still, if the true model is $\theta_Y$, the example shows that, under ambiguity neutrality, with positive probability beliefs converge to the true model, while with high ambiguity aversion the process of actions and beliefs is trapped in a non-Nash SCE.

The next example illustrates the relevance of the assumption of a unique ambiguous action.

**Example 9** (Multiple ambiguous actions). There are four possible states, $S = \{g, \bar{g}, b, \bar{b}\}$, and two possible models in $\Theta$, the good model $\theta_g$ and the bad model $\theta_b$ defined as follows:

$$
\theta_g (g) = 0.9 = \theta_b (b) \quad \text{and} \quad \theta_g (\bar{g}) = 0.1 = \theta_b (\bar{b}) .
$$

The DM has three actions: He can bet aggressively (action $a$), bet conservatively (action $c$), or not bet at all (action $n$). The feedback received by the DM is his monetary payoff. We also assume risk-neutrality: for all $\bar{a} \in A$, $m \in M$, $u (a, m) = m$. Feedback and payoffs are summarized in the following table:

<table>
<thead>
<tr>
<th>$f = r$</th>
<th>$g$</th>
<th>$\bar{g}$</th>
<th>$b$</th>
<th>$\bar{b}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a$</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>$c$</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>$n$</td>
<td>4.2</td>
<td>4.2</td>
<td>4.2</td>
<td>4.2</td>
</tr>
</tbody>
</table>
Therefore,

\[ R(a, \theta_g) = 9, \ R(a, \theta_b) = 1, \ R(c, \theta_g) = 5 \] and \[ R(c, \theta_b) = 4 \].

For simplicity, suppose \( \delta = 0 \) and \( \mu(\theta_g) = \mu(\theta_b) = 1/2 \). In this case, in the first period an ambiguity neutral DM bets aggressively \( (a) \) because

\[ \frac{R(a, \theta_g) + R(a, \theta_b)}{2} = 5 > \frac{R(c, \theta_g) + R(c, \theta_b)}{2} = 4.5 \].

Similarly, one can check that a DM with intermediate ambiguity attitudes (i.e. \( \lambda = 1 \)) and with the same belief \( \mu \) bets conservatively. Now, suppose that the true model is \( \theta_g \). First note that, since \( c \) perfectly reveals the model, the ambiguity averse DM discovers the true model at the end of the first period and starts to bet aggressively from the second period. In other words, convergence to Nash equilibrium (since \( a \) is the objectively optimal action under \( \theta_g \)) happens in one period and with probability 1. Now, consider the ambiguity neutral DM. He bets on \( a \) in the first period, and with probability \( 0.1 = \theta_g(\bar{y}) \) he receives message \( m = 0 \), so that the updated belief is

\[ \mu(\theta_g| (a, 0)) = 0.1 \].

With this, from the second period he stops betting (i.e. he chooses \( n \)) and remains in the dark. Thus, there is at least probability 0.1 that the ambiguity neutral DM be trapped in a non-Nash SCE.

In this example there is a misalignment between the most informative action \( (c) \) and the most ambiguous one \( (a) \). Therefore, by avoiding the most ambiguous action, an ambiguity averse DM quickly gathers information about the true model. We believe that this kind of misalignment is unlikely to arise in many applications. Still, we conjecture that the results of this section can be extended to the case of multiple ambiguous actions by first giving an adequate definition of an ambiguity order (see, e.g., Jewitt and Mukerji 2017), and then imposing a condition of comonotonicity between informativeness and ambiguity of actions.\(^{29}\)

### 7.2 Myopic Decision Maker

Sharper versions of our results about comparative dynamics can be given when the DM is myopic, i.e., \( \delta = 0 \). This follows from the fact that, in the present framework where each period consists of a one-stage decision problem, a myopic DM is not vulnerable to dynamic inconsistencies. Despite this simplification, the behavior over time of a myopic DM evolves in interesting ways as he gathers information about the true stochastic process. Indeed, several models of learning in games use the assumption of myopic players (see, e.g., Fudenberg and Kreps 1995, and Fudenberg and Levine 1998). We show that when the DM is myopic, not only can we compare the behavior under ambiguity neutrality and aversion, but the comparative statics results hold for the entire spectrum of ambiguity attitudes.

\(^{29}\)We can prove this for the case of a myopic DM (\( \delta = 0 \)).
For the rest of this section, we assume that $\delta = 0$ and we let $\phi'$ be a strictly increasing and concave transformation of $\phi$, i.e., we assume that the DM with ambiguity attitudes $\phi'$ is strictly more ambiguity averse than the one with ambiguity attitudes $\phi$.

**Proposition 9.** Let $(\alpha, \mu)$ be rational under ambiguity attitudes $\phi$ and let $(\beta, \mu)$ be rational under ambiguity attitudes $\phi'$. Then, for every belief $\nu$ in the belief-range of both $(\alpha, \mu)$ and $(\beta, \mu)$, if $\alpha(\nu)$ is $\nu$-unambiguous, the same holds for $\beta(\nu)$.

Similarly, the speed of convergence to the SCE is monotone for the entire spectrum of ambiguity attitudes.

**Proposition 10.** Assume that there is a unique $\mu$-ambiguous action $a^*$; let $(\alpha, \mu)$ be rational under ambiguity attitudes $\phi$ and let $(\beta, \mu)$ be rational under ambiguity attitudes $\phi'$; furthermore, assume that $(\alpha, \mu, \bar{\theta})$ is consistent from some $T$. Then, $\bar{\theta}$-almost surely, the action-belief process $\left(\beta\left(\mu \left(\cdot | h_t^\phi\right), \mu \left(\cdot | h_t^\phi\right)\right)\right)$ converges to an SCE (given $\bar{\theta}$) at least as fast as the action-belief process $\left(\alpha\left(\mu \left(\cdot | h_t^\phi\right), \mu \left(\cdot | h_t^\phi\right)\right)\right)$.

As a consequence, the probability of converging to a Nash equilibrium is decreasing in ambiguity aversion, too.

**Proposition 11.** Assume that there is a unique ambiguous action $a^*$ given $\mu$. Let $(\alpha, \mu)$ and $(\beta, \mu)$ be rational under ambiguity attitudes $\phi$ and $\phi'$, and consistent from some $T_\alpha$ and $T_\beta$, respectively. Then, $\bar{\theta}$-almost surely, if $\left(a_t^\beta\right)$ converges to a Nash Equilibrium action (given $\bar{\theta}$), so does $\left(a_t^\alpha\right)$.

### 7.3 Alternative updating rules

The decision-theoretic literature does not make a clear stand on whether ambiguity averse players should update beliefs according to the standard rules of conditional probabilities (see, for example, Epstein and Schneider, 2007, Hanany and Klibanoff, 2009). As clarified in Section 3, we take the position that these rules are part of rational cognition, and the adoption of the smooth ambiguity model allows us to describe learning in a standard Bayesian fashion. Here we want to remark that one may conduct a similar analysis considering a DM who uses the Hanany and Klibanoff (2009) updating rule for beliefs.

Yet, the next proposition shows that even considering the alternative, dynamically consistent updating rule proposed by Hanany and Klibanoff, the set of rest points include the set of SCE as by Definition 2. Therefore, the main message of Section 7 that “ambiguity aversion makes convergence to a non-Nash equilibrium more likely” still holds.\(^{30}\)

Formally, assume that $\phi$ is differentiable\(^ {31}\) and let

$$
\mu \left(\theta | H^K (h_{t-1}, a, m)\right) = \frac{\phi'(R(a, \theta))}{\phi'(u(a, m))} \mu \left(\theta | H^K h_{t-1}\right) F (a, \theta) (m) 
\sum_{\theta' \in supp(\phi(\cdot | H^K h_{t-1})) \phi'(u(a, m))} \phi'(R(a, \theta')) \mu \left(\theta' | H^K h_{t-1}\right) F (a, \theta') (m)
$$

\(^{30}\)However, the techniques used here to prove convergence do not easily generalize to that model. The main technical difficulty is that beliefs are not a Martingale under their updating rule, and therefore one cannot use the Martingale Convergence Theorem to show their convergence.

\(^{31}\)This condition is necessary to define the smooth updating rule of Hanany and Klibanoff.
denote the Hanany-Klibano夫’s “smooth posterior” given history $h_t$ with this, we have:

**Proposition 12.** Let $(a, \mu, \bar{\theta})$ be a SCE. Then,

$$\bar{\theta} \left( \{ s : \mu \mid_{HK} f (a, s) = \mu \} \right) = 1.$$ 

In words, if playing the myopic best reply does not induce a change in beliefs under Bayes rule, it does not induce a change in beliefs under the smooth updating rule proposed by Hanany and Klibano夫. Example 10 in the appendix shows that the inclusion is strict: under ambiguity aversion the set of their rest points is strictly higher, and the indeterminacy of the system would be higher considering their rule.

8 Conclusions

The concept of self-confirming equilibrium with standard expected utility maximizing agents has been given a rigorous learning foundation. The literature on stochastic control problems shows that the behavior and beliefs of an ambiguity neutral agent, who faces an unknown i.i.d. process of states affecting the outcome of his actions, almost surely converges to an SCE (see Easley and Kiefer 1988 and Section 5 of the working paper for a detailed analysis of the connection with their work). As for games against other agents, convergence cannot be taken for granted; but if it occurs, the limit point must be an SCE (e.g., Fudenberg and Levine 1993, Fudenberg and Kreps 1995).

This learning foundation cannot be mechanically applied to the case of non-neutral ambiguity attitudes. Ambiguity averse agents typically have dynamically inconsistent preferences over strategies, and dynamic inconsistency prevents us from applying standard dynamic programming techniques. Given such difficulties, to derive results and insights about convergence to SCE under ambiguity aversion, we focus on the case of repeated play against nature. We assume that agents are sophisticated and thus take future incentives into account as they choose actions in earlier periods.

We find that, in several interesting problems, the set of SCE actions is independent of ambiguity attitudes (Proposition 8). Yet, ambiguity aversion affects the dynamics: Higher ambiguity aversion tends to decrease experimentation and therefore makes convergence to Nash equilibrium (best reply to the correct model) less likely. In particular, we show that ambiguity aversion may make it more likely that the agent falls into a “suboptimal certainty trap” whereby he keeps choosing an unambiguous action from which he cannot learn, which prevents him from identifying the objectively optimal action (Proposition 7).

We can give a game-theoretic interpretation of our analysis within a population-game scenario. In this setting, the DM recognizes to be unable to influence the actions of future co-players. Nevertheless, experimentation is valuable for him, since a better understanding of the correct distribution of strategies in co-players’ populations may allow selecting a better
strategy in the following periods (cf. Fudenberg and Levine 1993). The main difference is that Fudenberg and Levine consider an overlapping generations model with finitely lived agents. Since we assume an infinite horizon, we would have to slightly modify our model by introducing a constant probability of death and embed our analysis in an overlapping generations model (cf. Blanchard 1985, Fudenberg and He 2018).

9 Appendix: proofs and related material

Proof of Lemma 1 Note that $R(a, \cdot)$ is a bounded function, since $|R(a, \cdot)| \leq \max_{s \in S} r(a, s)$. Moreover, it is the sum of the continuous functions $r(a, s) \theta(s)$, so it is continuous. Thus, $\int \phi(R(a, \theta)) (\cdot) (d \theta)$ is continuous by the Portmanteau Theorem. Since $\phi$ is strictly increasing on the interval $[\min_{s \in S} r(a, s), \max_{s \in S} r(a, s)]$, by Proposition 6.4.5 in Garling 2013, $\phi^{-1}$ is continuous as well. Since $V(a, \cdot) = \phi^{-1} \circ \int \phi(R(a, \theta)) (\cdot) (d \theta)$, the result follows.

9.1 Models and Learning

Instrumental for the following proofs is the correspondence $\iota_t^S : S^{t-1} \to 2^{S^{\infty}}$ defined by

$$\iota_t^S(s^{t-1}) := I(h_t^S(s^{t-1})) \times S^{\infty} = (h_t^S)^{-1} \circ h_t^S(s^{t-1}) \times S^{\infty}. \quad (5)$$

We can regard $\iota_t^S$ as the identification correspondence determined by $\alpha$ at time $t$. This correspondence models the information about state histories which is available ex-ante at time $t$ to a DM who is acting according to strategy $\alpha$. Clearly, $s^{t-1} \times S^{\infty} \in \iota_t^S(s^{t-1})$, and so the correspondence induces a partition of $S^{t-1}$. We have perfect (state) identification under $\alpha$ when $\iota_t^S(s^{t-1}) = \{s^{t-1}\} \times S^{\infty}$ for each $s^{t-1}$ and each $t > 1$; in this case, the DM knows the actual past history $s^{t-1}$. Otherwise, we have partial identification. This dependence on $\alpha$ of the identification correspondence plays a key role in our results. Of course, there is no such dependence under own-action independence of feedback, in which case we can write $\iota_t(s^{t-1})$; in particular, under perfect feedback, $\iota_t(s^{t-1}) = \{s^{t-1}\} \times S^{\infty}$.

Proof of Lemma 2 Fix $s^{t}$ with $p_\theta(s^{t}) > 0$. Note that $s^{t} \in \iota_{t+1}^\alpha(s^{t}) \subseteq \iota_t^S(s^{t-1})$; thus, $p_\theta(\iota_t^S(s^{t-1})) \geq p_\theta(\iota_{t+1}^\alpha(s^{t})) \geq p_\theta(s^{t}) > 0$. Let $\theta \in \Theta_{t+1}^\alpha(\bar{\theta}) (s^{t-1})$; by definition, $p_\theta(\iota_t^S(s^{t-1})) > 0$. We want to show that $\theta \in \Theta_{t+1}^\alpha(\bar{\theta}) (s^{t})$. To this end, fix $E \in \sigma(h^\alpha)$ with $E \subseteq \iota_{t+1}^\alpha(s^{t})$. Since $\theta \in \Theta_{t+1}^\alpha(\bar{\theta}) (s^{t-1})$, then:

$$\frac{p_\theta(E)}{p_\theta(\iota_t^S(s^{t-1}))} = \frac{p_\theta(E)}{p_\theta(\iota_{t+1}^\alpha(s^{t}))}; \quad \frac{p_\theta(\iota_{t+1}^\alpha(s^{t}))}{p_\theta(\iota_t^S(s^{t-1}))} = \frac{p_\theta(\iota_{t+1}^\alpha(s^{t}))}{p_\theta(\iota_t^S(s^{t-1}))}.$$

The second equality implies $p_\theta(\iota_{t+1}^\alpha(s^{t})) > 0$. Since:

$$\frac{p_\theta(E)}{p_\theta(\iota_{t+1}^\alpha(s^{t}))} \frac{p_\theta(\iota_{t+1}^\alpha(s^{t}))}{p_\theta(\iota_t^S(s^{t-1}))} = \frac{p_\theta(E)}{p_\theta(\iota_{t+1}^\alpha(s^{t}))} \frac{p_\theta(E)}{p_\theta(\iota_t^S(s^{t-1}))},$$

it follows that:

$$\frac{p_\theta(E)}{p_\theta(\iota_{t+1}^\alpha(s^{t}))} = \frac{p_\theta(E)}{p_\theta(\iota_{t+1}^\alpha(s^{t}))}.$$
Hence, \( p_\theta^t (\cdot \mid h_{t+1}^\alpha (s^t)) = p_\theta^t (\cdot \mid h_{t+1}^\alpha (s^t)) \). Since \( \theta \in \text{supp} \mu (\cdot \mid h_t^\alpha (s^{t-1})) \) and \( p_\theta (\epsilon_t^\alpha (s^t)) > 0 \), it follows that \( \theta \in \text{supp} \mu (\cdot \mid h_{t+1}^\alpha (s^t)) \), hence \( \theta \in \Theta_{t+1}^\alpha (\hat{\theta}) (s^t) \). ■

**Lemma 6.** For every \( \hat{\theta} \in \Theta \), the process \( \mu (\hat{\theta} \mid h_t^\alpha) \) is a uniformly bounded martingale in \( (S^\infty, B (S^\infty), p_\mu) \).

**Proof of Lemma 6** Uniform boundedness is immediate from the fact that the process is a sequence of probabilities. For every \( t \), we want to show that:

\[
\mathbb{E}_{p_\mu} \left[ \mu (\hat{\theta} \mid h_t^\alpha) \mid \mu (\hat{\theta} \mid h_{t-1}^\alpha) = k \right] = k.
\]

Let \( h_{t-1} \) be an arbitrary element of \( H_{t-1} \) such that \( \mu (\hat{\theta} \mid h_{t-1}^\alpha) = k \) and \( p_\mu (h_{t-1}) > 0 \). By the Law of Iterated Expectations (see 9.7i in Williams 1991), it is enough to prove that:

\[
\mathbb{E}_{p_\mu} \left[ \mu (\hat{\theta} \mid h_t^\alpha) \mid h_{t-1}^\alpha = h_{t-1} \right] = k.
\]

Recall that the Bayes map yields:

\[
\Delta (\Theta) \times A \times M \rightarrow \Delta (\Theta)
\]

\[
(\mu_{t-1}, a, m) \mapsto B (\mu_{t-1}, a, m) (\theta) = \frac{F (a, \theta)(m) \mu (\theta)}{\sum_{\theta'} F (a, \theta')(m) \mu (\theta')}
\]

for each \( m \) deemed possible according to \( \mu_{t-1} \) given action \( a \), that is, each \( m \) such that the denominator is positive. Define:

\[
M (h_{t-1}) = \left\{ m \in M : \sum_{\theta'} F (a, \theta') (m) \mu (\theta' \mid h_{t-1}) > 0 \right\}.
\]

With this,

\[
\mathbb{E}_{p_\mu} \left[ \mu (\hat{\theta} \mid h_t^\alpha) \mid h_{t-1}^\alpha = h_{t-1} \right] = \sum_{m \in M (h_{t-1})} p_\mu [m_{t-1}^\alpha = m \mid h_{t-1}^\alpha = h_{t-1}] B (\mu (\cdot \mid h_{t-1}^\alpha), \alpha (h_{t-1}), m) (\hat{\theta})
\]

\[
= \sum_{m \in M (h_{t-1})} \sum_{\theta' \in \text{supp} m} \mu (\theta) p_\theta \left( \{ s^\infty : h_{t-1}^\alpha (s^\infty) = h_{t-1} \} \right) B (\mu (\cdot \mid h_{t-1}^\alpha), \alpha (h_{t-1}), m) (\hat{\theta})
\]

\[
= \sum_{m \in M (h_{t-1})} \sum_{\theta' \in \text{supp} m} \mu (\theta) p_\theta \left( \{ s^\infty : h_{t-1}^\alpha (s^\infty) = h_{t-1} \} \right) B (\mu (\cdot \mid h_{t-1}^\alpha), \alpha (h_{t-1}), m) (\hat{\theta})
\]

\[
= \sum_{m \in M (h_{t-1})} F (\alpha (h_{t-1}), \theta) (m) \mu (\theta \mid h_{t-1}) \frac{k F (\alpha (h_{t-1}), \hat{\theta}) (m)}{\sum_{\theta'} F (\alpha (h_{t-1}), \theta') (m)}
\]

\[
= \sum_{m \in M (h_{t-1})} k F (\alpha (h_{t-1}), \hat{\theta}) (m) \sum_{\theta' \in \text{supp} m} \mu_{t-1} (\theta') F (\alpha (h_{t-1}), \theta) (m)
\]

\[
= \sum_{m \in M (h_{t-1})} k F (\alpha (h_{t-1}), \hat{\theta}) (m) = k.
\]

where the first equality comes from the definition of expected value, the second by the definition of \( p_\mu \), the third from the fact that the environment is i.i.d., and the fourth from the definition of conditional probability; the remaining equalities are immediate. ■
Lemma 7. For every $\hat{\theta} \in \text{supp}\mu_0$, the process $(\mu(\cdot|\mathbf{h}^\alpha_t))_{t \in \mathbb{N}_0}$ converges $\hat{\theta}$-a.s. to a random limit $\mu^\alpha_{s^\infty}$.

Proof of Lemma 7 By Lemma 6 the stochastic process $(\mu(\cdot|\mathbf{h}^\alpha_t))_{t \in \mathbb{N}_0}$ is a uniformly-bounded martingale. By the Martingale Convergence Theorem (Billingsley, Theorem 35.5), the limit random variable $\mu^\alpha_{s^\infty}$ exists $p_\mu$-almost surely. This means that there exists a set $E \in \mathcal{B}(S^\infty)$ such that $p_\mu(E) = 1$, so $p_\mu(S^\infty \setminus E) = 0$, and

$$\lim_{t \to +\infty} \mu(\cdot|\mathbf{h}^\alpha_t(s^\infty)) = \mu^\alpha_{s^\infty}$$

for every $s^\infty \in E$. Note that, since $\mu(\cdot|\hat{\theta}) > 0$,

$$p_\mu(S^\infty \setminus E) = 0$$

$$\sum_{\theta \in \text{supp}\mu} p_\theta(S^\infty \setminus E)\mu(\theta) = 0$$

$$\implies p_\hat{\theta}(S^\infty \setminus E) = 0$$

so $p_\hat{\theta}(E) = 1$. 

Proof of Lemma 3 If $\bar{\theta} \in \text{supp}\mu_0$ the result follows immediately from Lemma 7. If $\bar{\theta} \notin \text{supp}\mu_0$, under consistency, the set $O := \{s^\infty : \Theta^\mu_T(\bar{\theta}) (s^{T-1}) \neq \emptyset\}$ has $p_{\bar{\theta}}$-probability 1. Define the set $E^* := \{s^\infty : \text{limit}_{t \to +\infty} \mu(\cdot|\mathbf{h}^\alpha_t(s^\infty)) = \mu^\alpha_{s^\infty}\}$. We have that $E \subseteq E^*$, where $E$ is the set from the proof of Lemma 7. For every $s^\infty \in O$, we can find some $\theta(s^\infty) \in \text{supp}\mu(\cdot|\mathbf{h}^\alpha_t(s^{t-1}))$ such that $p_{\theta}(A) = p_{\theta(s^\infty)}(A)$ for every $A \in \sigma_T(\mathbf{h}^\alpha_t(s^{t-1}))$. In particular, $p_{\bar{\theta}}(E^*) \geq p_{\bar{\theta}}(E) = p_{\theta(s^\infty)}(E) = 1$, so $p_{\bar{\theta}}(E^*) = 1$.

Lemma 8. For every path $s^\infty$, denote by $\mathbf{\alpha}(s^\infty)$ the set of actions played infinitely often (i.o.) under strategy $\mathbf{\alpha}$ along this path. For every $\hat{\theta} \in \text{supp}\mu$, define the set

$$E_{\hat{\theta}} = \left\{s^\infty : \forall (a, \bar{s}) \in \mathbf{\alpha}(s^\infty) \times \text{supp}\hat{\theta}, (\mathbf{a} \left(\mu(\cdot|\mathbf{h}^\alpha_{t-1}(s^\infty))\right), s_t) = (a, \bar{s}) \text{ i.o.}\right\};$$

we have $p_{\hat{\theta}}(E_{\hat{\theta}}) = 1$.

Proof of Lemma 8 For every $(a, \bar{s}) \in \mathbf{\alpha}(s^\infty) \times \text{supp}\hat{\theta}$, denote by $E(a, \bar{s}, n) \subseteq S^\infty$ the set of sequences $s^\infty$ such that $a \in \mathbf{\alpha}(s^\infty)$ but $(\mathbf{a} \left(\mu(\cdot|\mathbf{h}^\alpha_{t-1}(s^\infty))\right), s_t) \neq (a, \bar{s})$ for every $t \geq n$. We have:

$$S^\infty \setminus E_{\hat{\theta}} \subseteq \bigcup_{a \in \mathbf{\alpha}(s^\infty)} \bigcup_{\bar{s} \in \text{supp}\hat{\theta}} \bigcup_{n \in \mathbb{N}} E(a, \bar{s}, n).$$

In turn, for every $k \in \mathbb{N},$

$$E(a, \bar{s}, n) \subseteq \bigcap_{j=1}^k \left\{s^\infty : \exists s_j \in S \setminus \{\bar{s}\}, t_j \geq n, \text{supp}\hat{\theta}, (\mathbf{a} \left(\mu(\cdot|\mathbf{h}^\alpha_{t-1}(s^\infty))\right), s_{t_j}) = (a, s_j)\right\}.$$ 

Thus,
Since \( \bar{s} \in \text{supp} \hat{\theta} \), this inequality implies that \( p_{\hat{\theta}}(E(a, \bar{s}, n)) = 0 \). It follows that \( p_{\hat{\theta}}(S^\infty \setminus E_{\hat{\theta}}) = 0 \), or \( p_{\hat{\theta}}(E_{\hat{\theta}}) = 1 \).

**Lemma 9.** For every \( \hat{\theta} \in \text{supp} \mu \),

\[
\forall a \in \alpha^\infty (s^\infty), \quad \mu_{s^\infty}^\alpha \left( \left\{ \theta \in \Theta : F(a, \theta) = F \left( a, \hat{\theta} \right) \right\} \right) = 1
\]

holds \( \hat{\theta} \)-almost surely.

**Proof of Lemma 9** Define the set \( E_{\hat{\theta}}^\alpha := E \cap E_{\hat{\theta}} \), and fix a sample path \( s^\infty \in \bar{E}_{\hat{\theta}} \). Suppose by way of contradiction that there is some \( a \in \alpha^\infty (s^\infty) \) and some \( m \in M \) such that, for some \( \theta \in \text{supp} \mu \), we have \( F(a, \theta) \neq F \left( a, \hat{\theta} \right) \). This implies that \( B \left( \mu_{s^\infty}^\alpha, a, f(a, \bar{s}) \right) \neq \mu_{s^\infty}^\alpha \), so we can find some \( \varepsilon > 0 \) such that \( \| B \left( \mu_{s^\infty}^\alpha, a, f(a, \bar{s}) \right) - \mu_{s^\infty}^\alpha \| = 2\varepsilon \). By continuity of the Bayes map, there exists some \( \delta > 0 \) such that \( \| \mu \left( \cdot | h_{t-1}^\alpha (s^\infty) \right) - \mu_{s^\infty}^\alpha \| < \delta \) implies:

\[
\| B \left( \mu \left( \cdot | h_{t-1}^\alpha (s^\infty) \right), a, f(a, \bar{s}) \right) - B \left( \mu_{s^\infty}^\alpha, a, f(a, \bar{s}) \right) \| < \varepsilon.
\]

Therefore,

\[
2\varepsilon = \| B \left( \mu_{s^\infty}^\alpha, a, f(a, \bar{s}) \right) - \mu_{s^\infty}^\alpha \|
\leq \| B \left( \mu \left( \cdot | h_{t-1}^\alpha (s^\infty) \right), a, f(a, \bar{s}) \right) - \mu_{s^\infty}^\alpha \|
+ \| B \left( \mu \left( \cdot | h_{t-1}^\alpha (s^\infty) \right), a, f(a, \bar{s}) \right) - B \left( \mu_{s^\infty}^\alpha, a, f(a, \bar{s}) \right) \|

< \| B \left( \mu \left( \cdot | h_{t-1}^\alpha (s^\infty) \right), a, f(a, \bar{s}) \right) - \mu_{s^\infty}^\alpha \| + \varepsilon,
\]

so \( \| B \left( \mu \left( \cdot | h_{t-1}^\alpha (s^\infty) \right), a, f(a, \bar{s}) \right) - \mu_{s^\infty}^\alpha \| > \varepsilon \). Invoking Lemma 7 and since \( s^\infty \in \bar{E}_{\hat{\theta}} \), there exists a sequence of dates \( (n)_{n \in \mathbb{N}} \) such that, for every \( n \),

\[
\left( \alpha \left( \mu \left( \cdot | h_{t-1}^\alpha (s^\infty) \right), s_{tn} \right) \right), \quad \left( \alpha \left( \mu \left( \cdot | h_{t-1}^\alpha (s^\infty) \right), s_{tn} \right) \right) = (a, \bar{s}) \quad \text{and} \quad \| \mu \left( \cdot | h_{t-1}^\alpha (s^\infty) \right) - \mu_{s^\infty}^\alpha \| < \delta,
\]

and so:

\[
\| \mu \left( \cdot | h_{t-1}^\alpha (s^\infty) \right) - \mu_{s^\infty}^\alpha \| = \| B \left( \mu \left( \cdot | h_{t-1}^\alpha (s^\infty) \right), a, f(a, \bar{s}) \right) - \mu_{s^\infty}^\alpha \| > \varepsilon.
\]

This contradicts Lemma 7.

**Proof of Proposition 1** The argument is the same as in the proof of Lemma 3, re-defining the set \( E^* \) as:

\[
E^* := \left\{ s^\infty : \mu_{s^\infty}^\alpha \left( \left\{ \theta \in \Theta : F(a, \theta) = F \left( a, \hat{\theta} \right) \right\} \right) = 1 \right\}.
\]

Further details are omitted.

**Proof of Corollary 1** Since \( \sigma(h^\alpha) = \sigma(s^{t-1}) \) for all \( t > 1 \), it follows that \( \Theta_1^\alpha (\hat{\theta}) = \{ \hat{\theta} \} \) and \( \sigma(h^\alpha) = \sigma(s^\infty) \). By Proposition 1, the statement follows.
9.2 Value

In the rest of the Appendix, we will make use of the fact that the dependence on the state in the value function can be made explicit. Indeed, by definition we have

$$
\sum_{h_r \in H_r} R(\alpha(h_r), \theta) p_\theta([h_r^\alpha = h_r] \mid h_t) = \sum_{s^r \in S^r} r(\alpha_r^\alpha(s^r-1), s_r) p_\theta(s^r \mid h_t),
$$

so that the Value function can be expressed as

$$
V(\alpha, \mu \mid h_t) := \sum_{\tau = t}^{\infty} \delta^{r-1} \phi^{-1} \left( \int_{\Theta} \phi \left( \sum_{s^r \in S^r} r(\alpha_r^\alpha(s^r-1), s_r) p_\theta(s^r \mid h_t) \right) \mu(d\theta \mid h_t) \right).
$$

Proof of Lemma 4 It is immediate to see that the map:

$$
W_{\hat{t}} : \theta \mapsto \phi \left( \sum_{s^\hat{t} \in S^\hat{t}} r(\alpha_{\hat{t}}(s^{\hat{t}-1}), s_{\hat{t}}) p_\theta(s_{\hat{t}} \mid h_t) \right)
$$

is continuous and bounded by $\max_{(a, s) \in A \times S} \phi(r(a, s))$. Moreover, as argued in the proof of Lemma 1, $\phi^{-1}$ is continuous. Since the space of measures endowed with the topology of weak convergence is metrizable, it is enough to show sequential continuity. By continuity of Bayesian updating with respect to positive probability events and $p_\mu(I(h_t)) > 0$,

$$
\mu_n \to \mu \Rightarrow \mu_n(\cdot \mid h_t) \to \mu(\cdot \mid h_t).
$$

Therefore, by definition of weak convergence of measures and continuity of $W_{\hat{t}}$ and $\phi^{-1}$:

$$
\mu_n(\cdot \mid h_t) \to \mu(\cdot \mid h_t) \Rightarrow \phi^{-1}(W_{\hat{t}}(\theta) \mu_n(d\theta \mid h_t)) \to \phi^{-1}(W_{\hat{t}}(\theta) \mu(d\theta \mid h_t)).
$$

Let $\varepsilon > 0$. Since $\delta < 1$ and $W_\tau$ is bounded, there exists a $T$ such that for every $\mu(\cdot \mid h_t)$,

$$
\left| \sum_{\tau = T}^{\infty} \delta^{r-1} \phi^{-1} \left( \int_{\Theta} W_\tau(\theta) \mu(d\theta \mid h_t) \right) \right| < \varepsilon.
$$

But then, let $n$ be such that for every $\tau \leq T$, 

$$
|\phi^{-1}(W_{\hat{t}}(\theta) \mu_n(d\theta \mid h_t)) - \phi^{-1}(W_{\hat{t}}(\theta) \mu(d\theta \mid h_t))| < \varepsilon.
$$

It follows that:

$$
\begin{align*}
|V(\alpha, \mu \mid h_t) - V(\alpha, \mu_n \mid h_t)| & = \left| \sum_{\tau = t}^{\infty} \delta^{r-1} \phi^{-1} \left( \int_{\Theta} W_\tau(\theta) \mu(d\theta \mid h_t) \right) - \sum_{\tau = t}^{\infty} \delta^{r-1} \phi^{-1} \left( \int_{\Theta} W_\tau(\theta) \mu_n(d\theta \mid h_t) \right) \right| \\
& \leq \sum_{\tau = t}^{T} \delta^{r-1} \phi^{-1} \left( \int_{\Theta} W_\tau(\theta) \mu(d\theta \mid h_t) \right) - \sum_{\tau = t}^{T} \delta^{r-1} \phi^{-1} \left( \int_{\Theta} W_\tau(\theta) \mu_n(d\theta \mid h_t) \right) \\
& \quad + \sum_{\tau = T+1}^{\infty} \delta^{r-1} \phi^{-1} \left( \int_{\Theta} W_\tau(\theta) \mu(d\theta \mid h_t) \right) - \sum_{\tau = T+1}^{\infty} \delta^{r-1} \phi^{-1} \left( \int_{\Theta} W_\tau(\theta) \mu_n(d\theta \mid h_t) \right) \\
& \leq (T + 2) \varepsilon.
\end{align*}
$$

34
Since \( \varepsilon \) has been chosen arbitrarily, we have proved the result.

**Proof of Corollary 2** By Proposition 1 the set
\[
E^* = \left\{ s^\infty : \lim_{t \to \infty} \mu(\{ \theta \in \Theta : F(a, \theta) = F(a, \overline{\theta}) \} | h_t^\alpha (s^\infty)) = 1 \right\}
\]
has \( \overline{\theta} \)-probability 1. For each \( s^\infty \in E^* \), consider:
\[
|V(\alpha, \mu | h_t^\alpha (s^\infty)) - V(\alpha, \mu_{s^\infty} | h_t^\alpha (s^\infty))| = \sum_{t=1}^{\infty} \delta^{t-1} \phi^{-1} \left( \int_{\Theta} W(\theta) \mu(d\theta | h_t^\alpha (s^\infty)) \right) - \sum_{t=1}^{\infty} \delta^{t-1} \phi^{-1} \left( \int_{\Theta} W(\theta) \mu_{s^\infty}(d\theta | h_t^\alpha (s^\infty)) \right)
\]
\[
= \sum_{t=1}^{\infty} \delta^{t-1} \phi^{-1} \left( \int_{\Theta} W(\theta) \mu(d\theta | h_t^\alpha (s^\infty)) \right) - \sum_{t=1}^{\infty} \delta^{t-1} \phi^{-1} \left( \int_{\Theta} W(\theta) \mu_{s^\infty}(d\theta | h_t^\alpha (s^\infty)) \right).
\]
The second equality follows from fact that there’s no further updating from \( \mu_{s^\infty} \). By a similar argument as in the proof of Lemma 4 we have that:
\[
|V(\alpha, \mu | h_t^\alpha (s^\infty)) - V(\alpha, \mu_{s^\infty} | h_t^\alpha (s^\infty))| \to 0.
\]
We are done once we note that, by Proposition 1
\[
V(\alpha, \mu_{s^\infty} | h_t^\alpha (s^\infty)) = V(\alpha, \delta_{\overline{\theta}} | h_t^\alpha (s^\infty)).
\]

**Proof of Lemma 5** First note that since \( p_\mu(I(h_t)) \) and \( p_\mu(I(h_t')) \) are strictly positive, then
\[
\frac{\mu(\theta) p_\theta(I(h_t))}{p_\mu(I(h_t'))} = \mu(\theta | h_t') = \mu(\theta | h_t) = \frac{\mu(\theta) p_\theta(I(h_t))}{p_\mu(I(h_t))}.
\]
In particular,
\[
\mu(\theta | h_t') = \mu(\theta | h_t) > 0 \Rightarrow p_\theta(I(h_t)) > 0, \text{ and } p_\theta(I(h_t')) > 0.
\]
That is, the models in the support of the \( \mu(\cdot | h_t') = \mu(\cdot | h_t) \) assign positive probability to the two conditioning events. In turn, this implies that \( p_\theta(\cdot | h_t) \) is well defined. Hence we have:
\[
V(\alpha, \mu | h_t) = \sum_{t=1}^{\infty} \delta^{t-1} \phi^{-1} \left( \int_{\text{supp } \mu(\cdot | h_t)} \phi \left( \sum_{s^t \in S^t} r(\alpha^\alpha_t(s^t-1), s_t) p_\theta(s^t | h_t) \right) \mu(d\theta | h_t) \right)
\]
To show our result, we will prove that for every \( n \in \mathbb{N}_0 \),
\[
\phi^{-1} \left( \int_{\text{supp } \mu(\cdot | h_t)} \phi \left( \sum_{s^{t+n} : h_t^\alpha(s^{t-1})=h_t} r(\alpha^\alpha_{t+n}(s^{t+n-1}), s_{t+n}) p_\theta(s^{t+n} | h_t) \right) \mu(d\theta | h_t) \right)
\]
\[
= \phi^{-1} \left( \int_{\text{supp } \mu(\cdot | h_t')} \phi \left( \sum_{s^{t+n} : h_t^\alpha(s^{t-1})=h_t'} r(\alpha^\alpha_{t+n}(s^{t+n-1}), s_{t+n}) p_\theta(s^{t+n} | h_t') \right) \mu(d\theta | h_t') \right).
\]
Since \( V(\alpha, \mu | h_t) \) and \( V(\alpha, \mu | h_t') \) are defined as the discounted sum from \( n = 0 \) to infinity of, respectively, the first and second line above, the statement will follow.
Let $n \in \mathbb{N}_0$, and $\theta \in \text{supp}\mu(\cdot \mid h_t) = \text{supp}\mu(\cdot \mid h'_t)$, and $(k_0, \ldots, k_n)$ such that $\theta(k_i) \neq 0$ for every $i$ in $\{1, \ldots, n\}$. Define:

$$K(k_0, \ldots, k_n) := \{s^{t+n} | s_t = k_0, \ldots, s_{t+n} = k_n \} \cap \{s^{t+n} : h_t^\alpha(s^{t-1}) = h_t\}$$

and

$$K'(k_0, \ldots, k_n) := \{s'^{t+n} | s'_t = k_0, \ldots, s'_{t+n} = k_n \} \cap \{s'^{t+n} : h'_t^\alpha(s'^{t-1}) = h'_t\}.$$ 

By definition of $p_\theta$:

$$\forall s^{t+n} \in K(k_0, \ldots, k_n), s'^{t+n} \in K'(k_0, \ldots, k_n) \quad p_\theta(s^{t+n} | h_t) = \prod_{i=0}^{n} \theta(k_n) = p_\theta(s'^{t+n} | h'_t).$$

To ease notation, fix $(k_0, \ldots, k_n)$ momentarily and let $K = K(k_0, \ldots, k_n)$ and $K' = K'(k_0, \ldots, k_n)$. Note that

$$t \left( a_{t+n}^\alpha \left( s^{t+n-1} \right), s_{t+n} \right)$$

is constant on $K$. Indeed, we prove by way of induction that for every $j \in \{0, \ldots, n\}$, $a_{t+j}^\alpha \left( s^{t+j-1} \right)$ is constant on $K$. Since for every $s^{t+n} \in K$ we have $h_t^\alpha(s^{t-1}) = h_t$,

$$a_t^\alpha \left( s^{t-1} \right) = \alpha(\mu(\cdot | h_t)).$$

Suppose by way of induction that the statement holds for $j' \leq j$. Thus, for every $s^{t+n} \in K$ we have

$$h_{t+j}^\alpha(s^{t+j-1}) = (h_t, a_t^\alpha \left( s^{t-1} \right), f(a_t^\alpha \left( s^{t-1} \right), s_t), \ldots, f(a_{t+j-1}^\alpha \left( s^{t+j-1} \right), s_{t+j-1})), $$

which, by definition of $K$ and by the inductive hypothesis, is constant on $K$. It follows that:

$$a_{t+j}^\alpha \left( s^{t+j-1} \right) = \alpha(\mu(\cdot | h_{t+j}^\alpha(s^{t+j-1}))).$$

is constant on $K$. Therefore, since $s_{t+n} = k_n$ for every $s^{t+n} \in K$, we have shown that also $t \left( a_{t+n}^\alpha \left( s^{t+n-1} \right), s_{t+n} \right)$ is constant on $K$. A similar argument shows that $t \left( a_{t+n}^\alpha \left( s^{t+n-1} \right), s_{t+n} \right)$ is constant on $K'$. Moreover, we have that, for every $s^{t+n}$ in $K$ and $s'^{t+n}$ in $K'$, for every $j$ in $\{0, \ldots, n\}$,

$$\mu(\cdot | h_{t+j}^\alpha(s^{t+j-1})) = \mu(\cdot | h_{t+j}^\alpha(s'^{t+j-1})).$$

We prove this equality by induction on $j$. By hypothesis, it is true for $j = 0$. Let $j \in \{1, \ldots, n\}$ and suppose that is true for $j - 1$. This implies that:

$$a_{t+j-1}^\alpha \left( s^{t+j-2} \right) = \alpha(\mu(\cdot | h_{t+j-1}^\alpha(s^{t+j-2})))$$

$$= \alpha(\mu(\cdot | h_{t+j-1}^\alpha(s'^{t+j-2})))$$

$$= a_{t+j-1}^\alpha \left( s'^{t+j-2} \right).$$
Therefore:

\[
\begin{align*}
\mu(\theta | h^0_{t+j}(s^{t+j-1})) &= \mu\left(\cdot | h^0_{t+j-1}(s^{t+j-2})\right) F\left(a^0_{t+j-1}(s^{t+j-2}), \theta \right) \left( f\left(a^0_{t+j-1}(s^{t+j-2}), k_{j-1}\right) \right) \\
&= F\left(a^0_{t+j-1}(s^{t+j-2}), \theta \mu(\cdot | h^0_{t+j-1}(s^{t+j-2}))\right) \left( f\left(a^0_{t+j-1}(s^{t+j-2}), k_{j-1}\right) \right) \\
&= \mu\left(\cdot | h^0_{t+j-1}(s^{t+j-2})\right) F\left(a^0_{t+j-1}(s^{t+j-2}), \theta \right) \left( f\left(a^0_{t+j-1}(s^{t+j-2}), k_{j-1}\right) \right) \\
&= \mu(\theta | h^0_{t+j-1}(s^{t+j-2})),
\end{align*}
\]

This in turn implies that, for every \(s^{t+n}\) in \(K\) and \(s'^{t+n}\) in \(K'\),

\[
\begin{align*}
r\left(a^0_{t+n}(s^{t+n-1}), s_{t+n}\right) &= r\left(\alpha\left(\mu\left(\cdot | h^0_{t+n}(s^{t+n-1})\right)\right), k_n\right) \\
&= r\left(\alpha\left(\mu\left(\cdot | h^0_{t+n}(s^{t+n-1})\right)\right), k_n\right) \\
&= r\left(a^0_{t+n}(s^{t+n-1}), s_{t+n}\right).
\end{align*}
\]

Now, we restart to explicitly highlight the dependence on \((k_0, \ldots, k_n)\) of \(K\). Moreover, for every \(n \in \mathbb{N}_0\) and for every \((k_1, \ldots, k_n) \in S^n\), let \(r(k_0, \ldots, k_n) = r\left(a^0_{t+n}(s^{t+n-1}), s_{t+n}\right) = r\left(a^0_{t+n}(s^{t+n-1}), s_{t+n}\right)\), where \(s^{t+n} \in K(k_0, \ldots, k_n)\) and \(s'^{t+n} \in K'(k_0, \ldots, k_n)\). By (6), this quantity is well defined. We have:

\[
\begin{align*}
\sum_{s^{t+n}: h^0_t(s^{t+n})=h_t} r\left(a^0_{t+n}(s^{t+n-1}), s_{t+n}\right) p_\theta(s^{t+n}|h_t) &= \prod_{i=0}^n \theta(k_n) \\
&= \sum_{s'^{t+n}: h^0_{t'}(s'^{t+n})=h'_{t'}} r\left(a^0_{t'+n}(s'^{t+n-1}), s_{t'+n}\right) p_\theta(s'^{t+n}|h'_{t'}).
\end{align*}
\]

Finally, since we have \(\mu(\cdot | h_t) = \mu(\cdot | h'_{t'})\), this implies that:

\[
\begin{align*}
\phi^{-1}\left(\int_{\text{supp} \mu(\cdot | h_t)} \phi \left(\sum_{s^{t+n}: h^0_t(s^{t+n})=h_t} r\left(a^0_{t+n}(s^{t+n-1}), s_{t+n}\right) p_\theta(s^{t+n}|h_t)\right) \mu(\text{d}\theta | h_t)\right) \\
= \phi^{-1}\left(\int_{\text{supp} \mu(\cdot | h'_{t'})} \phi \left(\sum_{s'^{t+n}: h^0_{t'}(s'^{t+n})=h'_{t'}} r\left(a^0_{t'+n}(s'^{t+n-1}), s_{t'+n}\right) p_\theta(s'^{t+n}|h'_{t'})\right) \mu(\text{d}\theta | h'_{t'})\right)
\end{align*}
\]

and the thesis follows.

Lemma 5 and our definition of rationality immediately imply the following Corollary.

**Corollary 4.** Consider \((\alpha, \mu)\) where \(\alpha\) is a stationary strategy. If \((\alpha, \mu)\) is not rational, that is, if there exists \((h_t, a)\) with

\[
V(\alpha/h_t, a, \mu|h_t) > V(\alpha, \mu|h_t),
\]

then for every \(h'_{t'}\) with \(\mu(\cdot | h_t) = \mu(\cdot | h'_{t'})\),

\[
V(\alpha/h'_{t'}, a, \mu|h'_{t'}) > V(\alpha, \mu|h'_{t'}).
\]
In words, if a stationary strategy is not rational at some history, the deviation needed to restore rationality is such that the strategy remains stationary.

**Proof of Corollary 4** Lemma 5 guarantees that the RHS of the two inequalities in the statement are equal. Since the deviation strategies are nonstationary only at histories $h_t$ and $h'_t$ respectively, and the actions prescribed at that histories is $a$ for both of them, we can repeat the argument of Lemma 5 to prove that:

$$
\phi^{-1} \left( \int_{\Omega} \mu(\cdot | h_t) \phi \left( \sum_{s^{t+n}: h_t(s^{t-1}) = h_t} r \left( a_{t+n}^{\alpha/(h_t,a)} (s^{t+n-1}, s_{t+n}) p_\theta (s^{t+n}| h_t) \right) \mu (d\theta | h_t) \right) \right) = \phi^{-1} \left( \int_{\Omega} \mu(\cdot | h'_t) \phi \left( \sum_{s^{t+n}: h'_t(s^{t-1}) = h'_t} r \left( a_{t+n}^{\alpha/(h'_t,a)} (s^{t+n-1}, s_{t+n}) p_\theta (s^{t+n}| h'_t) \right) \mu (d\theta | h'_t) \right) \right).
$$

This implies that also the LHS of the inequality are equal, proving the result.

**Proof of Proposition 2** The result follows by considering an ancillary game. We associate to each possible posterior a player in this new game. This is similar to considering the agent form of the decision problem, with the modification that information sets with the same posterior are associated with the same player. Since the possible information sets are countable, this game has a countable number of players.

We let the set of available actions to every player be $A$. Note that this means that the player associated with posterior $\nu(\cdot | h_t)$ has to choose the same action for every information set where she is asked to play (i.e. the histories where the DM has those beliefs). To complete the specification of the game, suppose that $i$ is a player associated with the belief $\nu$. Then her payoff function when the strategy profile induced by the players is $\beta$ is given by

$$
U_i = \sum_{h_t: \mu(\cdot | h_t) = \nu} \delta^t V (\beta, \mu(\cdot | h_t)).
$$

The ancillary game thus obtained is a perfect information game satisfying all the assumption of Theorem 3 in Hellwig and Leininger (1987), and therefore it admits a Nash equilibrium in pure strategies $\alpha$. By construction, the strategy corresponding to this equilibrium is stationary. Now, no player has incentives to simultaneously deviate from his behavior at every information set. But since the induced profile of strategy is now stationary, Corollary 4 implies that $\alpha$ satisfies the OSDP.

### 9.3 Convergence to SCE

**Lemma 10.** Let $(\alpha, \mu, \bar{\theta})$ be such that:

1. $\mu(\{\theta \in \Theta : p^{\bar{\theta}}_\theta = p^{\theta}_{\bar{\theta}}\}) = 1$;

2. For every action $a$, period $t$, and information history $h_t$, $p_\mu(I(h_t)) > 0 \Rightarrow V(\alpha, \mu | h_t) \geq V(\alpha/(h_t,a), \mu | h_t),$.
Lemma 6

The second equality comes from the fact that the first equality comes from property 1, the strict inequality comes from hypothesis, \( \phi \) prescribes the first period to come and coincide with \( \alpha \) otherwise. However, we have:

\[
V(\alpha, \mu)
\]
\[
= \sum_{s \in S} r(\alpha(\mu), s) \mu(s) + \delta V(\alpha, \mu)
\]
\[
\leq \sum_{s \in S} r(\alpha(\mu), s) \mu(s) + \delta \min_{m : F(\alpha(\mu), (\mu) > 0} V(\alpha, \mu(\cdot | (a, m)))
\]
\[
< \phi^{-1} \left( \int_{\Theta} \phi \left( \sum_{s \in S} r(a, s) \theta(s) \right) \mu(d\theta) \right)
\]
\[
\delta \min_{m : F(\alpha(\mu), (\mu) > 0} \left( \sum_{r = t}^{\infty} \phi^{-1} \left( \int_{\Theta} \phi \left( \sum_{s^r \in S^r} r(a^r_{\mu}(s^{r-1}), s^r) \mu(d\theta | (a, m)) \right) \right) \right)
\]
\[
= \phi^{-1} \left( \int_{\Theta} \phi \left( \sum_{s \in S} r(a, s) \theta(s) \right) \mu(d\theta) \right) +
\]
\[
\delta \min_{m : F(\alpha(\mu), (\mu) > 0} \left( \sum_{r = t}^{\infty} \phi^{-1} \left( \int_{\Theta} \phi \left( \sum_{s^r \in S^r} r(a^r_{\mu}(s^{r-1}), s^r) \mu(d\theta | (a, m)) \right) \right) \right)
\]
\[
\leq \phi^{-1} \left( \int_{\Theta} \phi \left( \sum_{s \in S} r(a, s) \theta(s) \right) \mu(d\theta) \right) +
\]
\[
\delta \left( \sum_{r = t}^{\infty} \phi^{-1} \left( \int_{\Theta} \phi \left( \sum_{s^r \in S^r} r(a^r_{\mu}(s^{r-1}), s^r) \mu(d\theta | (a, m)) \right) \right) \right)
\]
\[
= \phi^{-1} \left( \int_{\Theta} \phi \left( \sum_{s \in S} r(a, s) \theta(s) \right) \mu(d\theta) \right) +
\]
\[
\delta \left( \sum_{r = t}^{\infty} \phi^{-1} \left( \int_{\Theta} \phi \left( \sum_{s^r \in S^r} r(a^r_{\mu}(s^{r-1}), s^r) \mu(d\theta | (a, m)) \right) \right) \right)
\]
\[
= V(\alpha/a, \mu),
\]

where the first equality comes from property 1, the strict inequality comes from hypothesis, the second equality comes from the fact that \( \phi \) is strictly increasing, the third equality by property 2 and the fourth and fifth equalities by the definition of \( \alpha/a \). Note that we will be
done as soon as we prove the first weak inequality, that is:

\[ V(\alpha, \mu) \leq \min_{m:F(\alpha, \theta, \mu)(m) > 0} V(\alpha, \mu(\cdot|(a, m))). \]

Indeed, it would follow that \( V(\alpha, \mu) < V(\alpha/a, \mu) \), a contradiction with the fact that \((\alpha, \mu, \tilde{\theta})\) satisfies 2.

Suppose that there exist \( m \) such that \( F(a, \theta, \mu)(m) > 0 \) with \( V(\alpha, \mu(\cdot|(a, m))) < V(\alpha, \mu) \). The fact that \( F(a, \theta, \mu)(m) > 0 \) implies that \( \theta(\mu(I(a, m))) > 0 \). On the other hand, by property 1, \( \mu\left( \{ \theta \in \Theta : p_\theta^0 = p_\theta^1 \} \right) = 1 \), and in particular:

\[ \mu\left( \{ \theta \in \Theta : F(\alpha, \mu, \theta, m) = F(\alpha(\mu), \tilde{\theta})(m) \} \right) = 1. \]

Then, let \( B = \{ \theta \in \Theta : F(\alpha, \mu, \theta) = F(\alpha(\mu), \tilde{\theta}) \} \). By Bayes rule, we have that:

\[ \mu(B|(a, m)) = \frac{\int_B F(a, \theta)(m) \mu(d\theta)}{F(a, \theta, \mu)(m)} = \mu(B) = 1. \]

But then, it follows that:

\[
V(\alpha, \mu(\cdot|(a, m))) < (1 - \delta) V(\alpha, \mu) + \delta V(\alpha, \mu(\cdot|(a, m))) \\
= \phi^{-1}\left( \int_{\Theta} \phi \left( \sum_{s \in S} r(\alpha(\mu), s) \theta(s) \right) \mu(d\theta) \right) + \delta V(\alpha, \mu(\cdot|(a, m))) \\
= \phi^{-1}\left( \sum_{s \in S} r(\alpha(\mu), s) \theta(s) \right) \mu(d\theta|a, m) + \delta V(\alpha, \mu(\cdot|(a, m))) \\
= \sum_{t=1}^\infty \delta^t \phi^{-1}\left( \int_{\Theta} \phi \left( \sum_{s \in S^t} r^{-1} \left( a_{s^t}^{\alpha(\mu)}(s_{t-1}^t), s_t \right) p_\theta(s^t|h_t) \right) \mu(d\theta|a, m) \right) \\
= V(\alpha/\alpha(\mu), \mu(\cdot|(a, m))).
\]

This contradicts the fact that \((\alpha, \mu, \tilde{\theta})\) satisfies property 2. \(\blacksquare\)

**Proof of Proposition** \(\star\) First, we have that the hypotheses of Lemma \(\star\) are satisfied, so let \( E \) be as in the corresponding proof. By Lemma \(\star\), the value function \(\star\) is continuous in beliefs \(\mu\). Fix \( s^\infty \in E \); for every \( a \) in \( A \),

\[ \lim_{t \to \infty} V(\alpha/a, \mu(\cdot|h_t^\alpha(s_t^t))) = V(\alpha/a, \mu_{s^\infty}). \]

Let \( A_{s^\infty} := \arg\max_{a \in A} V(\alpha/a, \mu_{s^\infty}). \) Note that, in general, our definition of rationality does not require that \( \alpha(\mu_{s^\infty}) \in A_{s^\infty} \). Indeed, if there is no \( h_t \) such that \( p_\mu(I(h_t)) > 0 \) and \( \mu(\cdot|h_t) = \mu_{s^\infty} \), then \( \alpha(\mu_{s^\infty}) \) does not need to satisfy the one-shot-deviation property. Since \( s^\infty \in E \), it follows that \( p_{\tilde{\theta}}(I(h_t^\alpha(s_t^t))) > 0 \) for every finite \( t \). By consistency, if \( t \geq T \), we have that \( p_{\tilde{\theta}}(I(h_t^\alpha(s_t^t))) > 0 \) implies \( p_\mu(I(h_t^\alpha(s_t^t))) > 0 \). Hence,

\[ \alpha(\mu(\cdot|h_t^\alpha(s_t^t))) \in \arg\max_{a \in A} V(\alpha/a, \mu(\cdot|h_t^\alpha(s_t^t))). \]
Now, let $a \notin A_\infty$, and fix $a^* \in A_\infty$. We have that
\[ \lim_{t \to \infty} V(\alpha/a, \mu(\cdot|h_t^\alpha(s^{t-1}))) = V(\alpha/a, \mu_{s^\infty}^\alpha) < \max_{a \in A} V(\alpha/a', \mu_{s^\infty}^\alpha) = V(\alpha/a^*, \mu_{s^\infty}^\alpha) \]
\[ = \lim_{t \to \infty} V(\alpha/a^*, \mu(\cdot|h_t^\alpha(s^{t-1}))). \]
Hence there exists $T_\infty$ such that $a \notin \alpha(\cdot|h_t^\alpha(s^{t-1})))$ for every $t \geq T_\infty$. Let $T_\infty = \max_{a \in A/A_\infty} T_\infty$. Then, from $T_\infty$ onward, the only actions played are in $A_\infty$, that is, they satisfy the one-shot deviation property with respect to the limit beliefs $\mu_{s^\infty}^\alpha$. Let $\widehat{T}_\infty = \max \{ T, T_\infty \}$; we have that from $\widehat{T}_\infty$ onward the action prescribed by strategy $\alpha$, $a_t^\alpha(s^{t-1})$, satisfies the one-shot deviation property with respect to beliefs $\mu_{s^\infty}^\alpha$, and $\mu_{s^\infty}^\alpha$ is confirmed given such action. By Lemma 10 this implies that $(a_t^\alpha(s^{t-1}), \mu_{s^\infty}^\alpha, \bar{\theta})$ is an SCE for every $t \geq \widehat{T}_\infty$.

**Proof of Proposition 4** By hypothesis, we know that $(a_t^\alpha(\cdot), \mu(\cdot|h_t^\alpha))$ converges to an SCE on $s^\infty$ given $\theta$. Therefore, there exists $\widehat{T}$ such that, for every $t \geq \widehat{T}$, the pair $(a_t^\alpha(s^{t-1}), \mu_{s^\infty}^\alpha, \bar{\theta})$ is an SCE, and so
\[ a_t^\alpha(s^{t-1}) \in \arg \max_{a \in A} \phi^{-1} \left( \int_{\Theta} \phi(R(a, \theta)) \mu(\cdot|h_t^\alpha(s^{t-1})) \right) = \{ a^* \}. \]
Therefore, $(a^*, \mu_{s^\infty}^\alpha, \bar{\theta})$ is an SCE. Now, let $a \neq a^*$. By Lemmata 1 and 7 it follows that:
\[ \lim_{t \to \infty} \phi^{-1} \left( \int_{\Theta} \phi \left( \sum_{s \in S} r(a^*, s)\theta(s) \right) \mu(\cdot|h_t^\alpha(s^{t-1})) \right) \]
\[ = \phi^{-1} \left( \int_{\Theta} \phi \left( \sum_{s \in S} r(a^*, s)\theta(s) \right) \mu_{s^\infty}^\alpha(\cdot) \right) \]
\[ > \phi^{-1} \left( \int_{\Theta} \phi \left( \sum_{s \in S} r(a, s)\theta(s) \right) \mu_{s^\infty}^\alpha(\cdot) \right) \]
\[ = \lim_{t \to \infty} \phi^{-1} \left( \int_{\Theta} \phi \left( \sum_{s \in S} r(a, s)\theta(s) \right) \mu(\cdot|h_t^\alpha(s^{t-1})) \right). \]
Thus, there exists $\widehat{T}_a > \widehat{T}$ such that $t > \widehat{T}_a$ implies:
\[ \phi^{-1} \left( \int_{\Theta} \phi \left( \sum_{s \in S} r(a^*, s)\theta(s) \right) \mu(\cdot|h_t^\alpha(s^{t-1})) \right) \]
\[ > \phi^{-1} \left( \int_{\Theta} \phi \left( \sum_{s \in S} r(a, s)\theta(s) \right) \mu(\cdot|h_t^\alpha(s^{t-1})) \right). \]
Let $\widehat{T}_\infty = \max_{a \in A/a^*} \widehat{T}_a$. We have that $t > \widehat{T}_\infty$ implies
\[ \phi^{-1} \left( \int_{\Theta} \phi \left( \sum_{s \in S} r(a^*, s)\theta(s) \right) \mu(\cdot|h_t^\alpha(s^{t-1})) \right) \]
\[ = \max_{a \in A} \phi^{-1} \left( \int_{\Theta} \phi \left( \sum_{s \in S} r(a, s)\theta(s) \right) \mu(\cdot|h_t^\alpha(s^{t-1})) \right). \]
The thesis follows.

**Proof of Corollary 3** By Proposition 3, there exists a $E \subseteq S^\infty$ with $p_\theta (E) = 1$ such that convergence to an SCE happens on that set. By Proposition 1, there exists a $E^* \subseteq S^\infty$ with $p_\theta (E^*) = 1$ such that

$$s^\infty \in E^* \Rightarrow (\forall a \in \alpha^\infty (s^\infty) \quad \mu(\{ \theta \in \Theta : F(a, \theta) = F(a, \theta') \} \ | \ h^\alpha_t (s^\infty) = 1).$$

Let $G = E \cap E^*$. Perfect feedback implies that:

$$s^\infty \in G \Rightarrow (\forall a \in A \quad \mu(\{ \theta \in \Theta : F(a, \theta) = F(a, \theta') \} \ | \ h^\alpha_t (s^\infty) = 1).$$

But then,

$$s^\infty \in G \subseteq E^* \Rightarrow \phi^{-1} \left( \int_{\Theta} \phi(R(a, \theta)) \mu^\alpha_{s^\infty}(d\theta) \right) = R(a, \theta); \quad (7)$$

that is, the value of each action under the limit belief is equal to the objective value. Since $s^\infty \in G \subseteq E$, there exists a finite time $t$ such that $(a^\alpha_t (s^\tau-1), \mu^\alpha_{s^\infty}, \theta), \tau \geq t$, forms an SCE. By definition of SCE and Equation (7), this means that for $\tau \geq t$ only the objective myopic best reply is played. 

**9.4 Comparative dynamics**

Since in this section we deal with different levels of ambiguity attitudes, we are going to make explicit the dependence of the value function on the ambiguity attitudes by writing $V_\phi (\alpha, \mu)$ in place of $V (\alpha, \mu)$.

**Proof of Proposition 5** Since $(\alpha, \mu)$ is rational under ambiguity neutrality, it satisfies the one-shot deviation property for every $h_t$ such that $p_\mu (I(h_t)) > 0$ and $\nu = \mu (\cdot | h_t)$:

$$\forall a \in A, V_{id} (\alpha, \nu) \geq V_{id} (\alpha / (a, h_t), \nu).$$

Since the spaces of action and state are finite and $\delta < 1$, our problem is continuous at infinity. By ambiguity neutrality and Theorem 4.2 in Fudenberg and Tirole (1991), this implies that for every alternative strategy $\gamma$:

$$V_{id} (\alpha, \mu) \geq V_{id} (\gamma, \mu).$$

Let $(\beta, \mu)$ be rational under ambiguity aversion. Suppose, by way of contradiction, that $\beta (\nu)$ prescribes an ambiguous action:

$$\exists \theta, \theta' \in \text{supp } \nu : \quad F(\alpha (\mu), \theta) \neq F(\alpha (\mu), \theta').$$

Then,

$$V_\phi (\beta, \nu) = \sum_{\tau=1}^{\infty} \delta^{\tau-1} \phi^{-1} \left( \int_{\Theta} \phi \left( \sum_{s^\tau \in S^\tau} r \left( a^\beta_{s^\tau-1} (s^\tau), s^\tau \right) p_\theta^\beta (s^\tau) \right) \nu (d\theta) \right) \geq V_\phi (\beta / (\alpha (\nu), h_1), \nu) \geq R (\alpha (\nu), p_\nu) + \delta V_\phi (\beta, \mu)$$

42
where the inequality follows from the one-shot deviation property, while the second equality follows from the assumption that \( \alpha (\nu) \) is unambiguous. Therefore,

\[
V_\phi (\beta, \nu) \geq \frac{R(\alpha (\nu), p_\nu)}{1 - \delta}.
\]

By Jensen’s inequality, under strict concavity of \( \phi \),

\[
\frac{R(\alpha (\nu), p_\nu)}{1 - \delta} = V_{Id} (\alpha, \nu) \geq V_{Id} (\beta, \nu) > V_\phi (\beta, \nu) \geq \frac{R(\alpha (\nu), p_\nu)}{1 - \delta},
\]

a contradiction.

Proof of Proposition 7

Let \( p_\beta (\mathbf{h}_t^\alpha (s^{t-1})) > 0 \), and let \( (\alpha (\mu (\cdot | \mathbf{h}_t^\alpha (s^{t-1}))), \mu (\cdot | \mathbf{h}_t^\alpha (s^{t-1})), \bar{\theta}) \) be an SCE under ambiguity neutrality. By the absolute continuity assumption, an action that is unambiguous given \( \mu \) will be unambiguous given \( \mu (\cdot | \mathbf{h}_t^\alpha (s^{t-1})) \). Therefore, the sequence \( (\alpha (\mu (\cdot | \mathbf{h}_t^\alpha (s^{t-1}))))_{t=1}^T \) has the form \( (a^*, ..., a^*, a, ..., a) \) for some \( a \in A \)\(^{33}\).

By Proposition 5, \( \alpha (\mu (\cdot | \mathbf{h}_t^\alpha (s^{t-1}))) \neq a^* \) implies \( \beta (\mu (\cdot | \mathbf{h}_t^\beta (s^{t-1}))) \neq a^* \). Therefore, \( (\beta (\mu (\cdot | \mathbf{h}_t^\beta (s^{t-1}))))_{t=1}^T \) has the form \( (a^*, ..., a^*, a, ..., a) \) with a (weakly) shorter sequence of \( a^* \). If \( \beta (\mu (\cdot | \mathbf{h}_t^\beta (s^{t-1}))) \neq a^* \), \( (\beta (\mu (\cdot | \mathbf{h}_t^\beta (s^t))), \mu (\cdot | \mathbf{h}_t^\beta (s^t)), \bar{\theta}) \) is an SCE by Lemma 10 and rationality of \( (\beta, \mu) \). If \( (\beta (\mu (\cdot | \mathbf{h}_t^\beta (s^{t-1}))))_{t=1}^T = a^* \), then \( \alpha \) and \( \beta \) have prescribed the same action \( (a^*) \) at every node, and therefore:

\[
(\alpha (\mu (\cdot | \mathbf{h}_t^\alpha (s^{t-1}))), \mu (\cdot | \mathbf{h}_t^\alpha (s^{t-1})), \bar{\theta}) = (\beta (\mu (\cdot | \mathbf{h}_t^\beta (s^{t-1}))), \mu (\cdot | \mathbf{h}_t^\beta (s^{t-1})), \bar{\theta}).
\]

Again, by Lemma 10, the fact that the LHS is an SCE with ambiguity neutrality, and the definition of SCE, the RHS is an SCE under ambiguity aversion.

Proof of Proposition 7

First, suppose that the objectively-optimal action played in the SCE is \( a^* \). Then, given the original belief \( \mu \) and infinite history \( s^{\infty} \), \( (\mathbf{a}_t^\beta, \mu (\cdot | \mathbf{h}_t^\beta)) \) converges to an SCE \( (a^*, \mu^*, \bar{\theta}) \) if and only if \( \beta (\mu (\cdot | \mathbf{h}_t^\beta (s^{\infty}))) = a^* \) for every \( t \). But Proposition 5 guarantees that this can happen only if \( \alpha (\mu (\cdot | \mathbf{h}_t^\alpha (s^{\infty}))) = a^* \) for every \( t \), and therefore \( (\mathbf{a}_t^\alpha, \mu (\cdot | \mathbf{h}_t^\alpha)) \) also converges to \( (a^*, \mu^*, \bar{\theta}) \).

Second, suppose that the objectively-optimal action is \( a \neq a^* \). Consider \( s^{\infty} \) where convergence of \( (\mathbf{a}_t^\beta, \mu (\cdot | \mathbf{h}_t^\beta)) \) to an SCE where \( a \) is played happens. Then, by definition of SCE, \( R (a, \bar{\theta}) = R (a, \bar{\theta}) \) is constant on \( \text{supp} \mu \). At the same time, Proposition 3 guarantees convergence of \( (\mathbf{a}_t^\alpha, \mu (\cdot | \mathbf{h}_t^\alpha)) \) to an SCE \( (\bar{a}, \bar{\mu}, \bar{\theta}) \). The absolute continuity hypothesis implies that \( \text{supp} \mu (\cdot | \mathbf{h}_t^\alpha) \subset \text{supp} \mu (\cdot) \) for every \( h_t \) with \( p_\beta (h_t) > 0 \), and therefore \( R (a, \bar{\theta}) = R (a, \bar{\theta}) \) is constant on \( \text{supp} \bar{\mu} \). Since by definition of SCE \( R (\bar{a}, \theta) = R (\bar{a}, \theta) \) is also constant on \( \text{supp} \bar{\mu} \) and

\[
\hat{a} \in \arg \max_{a' \in A} \phi^{-1} \left( \int \phi \left( \sum_{s \in S} r(a', s) \theta(s) \right) \mu^\ast (d\theta) \right),
\]

\(33\)Possibly including the cases \( (a^*, ..., a^*) \) and \( (a, ..., a) \)
we must have $R(\tilde{a}, \theta) \geq R(a, \theta)$. By the objective optimality of $a$, we must have in fact that $R(\tilde{a}, \theta) = R(a, \theta)$, so $\tilde{a}$ is objectively optimal.

**Proof of Proposition**

Let $(a, \mu, \tilde{\theta})$ be an SCE under ambiguity attitude $\phi$. The optimality condition for an SCE gives:

$$a \in \arg\max_{a' \in A} \phi^{-1}\left(\int_{\Theta} \phi\left(\sum_{s \in S} r(a', s) \theta(s)\right) \mu(d\theta)\right). \tag{8}$$

Therefore, there exists $\theta$ in $\text{supp} \mu$ with:

$$\sum_{s \in S} r(a, s) \theta(s) \geq \sum_{s \in S} r(a^*, s) \theta(s).$$

Next, consider every $\tilde{a} \neq a^*$; such actions are unambiguous given $\mu$. By (8):

$$\sum_{s \in S} r(\tilde{a}, s) \theta(s) = \phi^{-1}\left(\int_{\Theta} \phi\left(\sum_{s \in S} r(\tilde{a}, s) \theta(s)\right) \mu(d\theta)\right) \geq \phi^{-1}\left(\int_{\Theta} \phi\left(\sum_{s \in S} r(a^*, s) \theta(s)\right) \mu(d\theta)\right) = \sum_{s \in S} r(a^*, s) \theta(s).$$

We have established that, for every $a_0 \in A$, we can find some $\theta' \in \text{supp} \mu$ such that $\sum_{s \in S} r(a_0, s) \theta'(s) \geq \sum_{s \in S} r(a^*, s) \theta'(s)$. Now, define the function $\tilde{v} : A \times \Delta(\text{supp} \mu) \rightarrow \mathbb{R}$ as $\tilde{v}(a', \nu) = V_{\text{id}}(a, \nu) - V_{\text{id}}(a', \nu)$, where $V_{\text{id}}(a', \mu)$ is the value function for an ambiguity-neutral DM. Then, $\tilde{v}(a', \delta_{\theta'}) \geq 0$, and so:

$$\min_{a' \in A} \max_{\nu \in \Delta(\text{supp} \mu)} \tilde{v}(a', \nu) \geq 0.$$ 

By Sion’s minimax theorem in Sion (1958),

$$\max_{\nu \in \Delta(\text{supp} \mu)} \min_{a' \in A} \tilde{v}(a', \nu) = \min_{\nu \in \Delta(\text{supp} \mu)} \max_{a' \in A} \tilde{v}(a', \nu) \geq 0.$$ 

The function $\tilde{v}$ is continuous and its domain is compact, so the set of maximizers is non-empty.\[34\] Pick any:

$$\mu' \in \arg\max_{\nu \in \Delta(\text{supp} \mu)} \left(\min_{a' \in A} h(a', \nu)\right).$$

We have, for every $a' \in A$,

$$V_{\text{id}}(a, \mu') - V_{\text{id}}(a', \mu') \geq \min_{a' \in A} h(a', \mu') \geq 0$$

It follows that $(a, \mu', \tilde{\theta})$ is an SCE under ambiguity neutrality. By Theorem 1 in Battigalli et al. (2015), $(a, \mu', \theta)$ is an SCE under ambiguity attitude $\phi'$.

\[34\] Compactness of the domain of $\tilde{v}$ follows from the fact that $\text{supp} \mu (\cdot | h_r)$ is a closed subset of $\Delta$. 

44
Proof of Proposition 9 Since \((\alpha, \mu)\) is rational under \(\phi\), and \(\delta = 0\), the one-shot deviation property reads:

\[
\forall a \in A, \phi^{-1} \left( \int_{\Theta} \phi \left( R \left( \alpha(\nu), \hat{\theta} \right) \right) \nu \left( d\hat{\theta} \right) \right) \geq \phi^{-1} \left( \int_{\Theta} \phi \left( R \left( a, \hat{\theta} \right) \right) \nu \left( d\hat{\theta} \right) \right).
\]

Let \((\beta, \mu)\) be rational under \(\phi'\). Suppose, by way of contradiction, that \(\beta(\nu)\) prescribes an ambiguous action:

\[
\exists \theta, \theta' \in \text{supp } \nu : 
F \left( \alpha(\nu), \theta \right) \neq F \left( \alpha(\nu), \theta' \right).
\]

Then,

\[
V_{\phi'}(\beta, \nu) = (\phi')^{-1} \left( \int_{\Theta} \phi' \left( R \left( \beta(\nu), \hat{\theta} \right) \right) \nu \left( d\hat{\theta} \right) \right)
\geq (\phi')^{-1} \left( \int_{\Theta} \phi' \left( R \left( \alpha(\nu), \hat{\theta} \right) \right) \nu \left( d\hat{\theta} \right) \right) = R(\alpha(\nu), p_{\nu}),
\]

where the inequality follows from the one-shot deviation property, while the second equality follows from the assumption that \(\alpha(\nu)\) is unambiguous. Therefore,

\[V_{\phi'}(\beta, \nu) \geq R(\alpha(\nu), p_{\nu}).\]

By Jensen’s inequality, and since \(\phi'\) is a strictly concave transformation of \(\phi\),

\[R(\alpha(\nu), p_{\nu}) = V_{\phi}(\alpha, \nu) \geq V_{\phi}(\beta, \nu) > V_{\phi'}(\beta, \nu) \geq R(\alpha(\nu), p_{\nu}),\]

a contradiction. ■

Proof of Proposition 10 The proof is very similar to that of Proposition 6, invoking Proposition 9 instead of Proposition 5. Further details are omitted. ■

Proof of Proposition 11 As with the previous proposition, invoke Proposition 9 instead of Proposition 5. ■

Proof of Proposition 12 Let \(s \in \text{supp } \hat{\theta}\), so that \(F \left( a, \hat{\theta} \left( f(a, s) \right) \right) > 0\). For all \(\theta' \in \text{supp } \mu\), we have \(F \left( a, \theta' \left( f(a, s) \right) \right) > 0\). Then, the expected utility under model \(\theta'\) given outcome \(f(a, s)\) and action \(a\) is well defined and equal to:

\[
\forall \theta' \in \text{supp } \mu \quad \frac{\sum_{s' \in f^{-1}(f(a, s))} \theta'(s') u(a, f(a, s'))}{F \left( a, \theta' \left( f(a, s) \right) \right)} = u(a, f(a, s))
\]

Therefore, Definition 3.1 and Theorem 3.1 in Hanany and Klibanoff (2009), imply that

\[
\mu \left( \theta \right| H_{K} a, f(a, s) \right) = \frac{\frac{\phi'(R(a, \theta))}{\phi(u(a, f(a, s)))} \mu(\theta) F \left( a, \theta \left( f(a, s) \right) \right)}{\sum_{\theta' \in \text{supp } \mu} \frac{\phi'(R(a, \theta'))}{\phi(u(a, f(a, s)))} \mu(\theta') F \left( a, \theta' \left( f(a, s) \right) \right)}.
\]
But then, recall that by Remark 1 \( \phi' (R (a, \theta')) \) is constant on \( \text{supp} \mu \), so that the previous expression simplifies into

\[
\mu (\theta |^{HK} a, f (a, s)) = \frac{\mu (\theta) F (a, \theta) (f (a, s))}{\sum_{\theta' \in \text{supp} \mu} \mu (\theta') F (a, \theta') (f (a, s))}.
\]

Finally, condition 1 in Definition 2 implies that \( F (a, \theta') (f (a, s)) \) so that

\[
\mu (\theta |^{HK} a, f (a, s)) = \frac{\mu (\theta)}{\sum_{\theta' \in \text{supp} \mu} \mu (\theta')} = \mu (\theta),
\]

proving the result.

The following example shows that the set of rest points under the HK smooth updating rule is strictly larger than that of Bayesian updating.

**Example 10.** Consider an urn that contains black (\( B \)) and green (\( G \)) balls. At each time \( t \), the DM is asked to bet 1 euro on the color of the ball that will be drawn from the urn; therefore the possible bets are blue (\( b \)) and green (\( g \)). Ex post, after the draw, he learns the result of his bet, namely, whether or not he wins 1 euro. The feedback function is described in the following table:

<table>
<thead>
<tr>
<th>( f )</th>
<th>( B )</th>
<th>( Y )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( b )</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>( y )</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Suppose that the DM is characterized by the following beliefs:

<table>
<thead>
<tr>
<th>Marginals</th>
<th>( B )</th>
<th>( Y )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \theta^{uni} )</td>
<td>( \frac{1}{2} )</td>
<td>( \frac{1}{2} )</td>
</tr>
<tr>
<td>( \theta^{Y} )</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Prior</th>
<th>( \theta^{Y} )</th>
<th>( \theta^{uni} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \mu )</td>
<td>( \frac{1}{2} )</td>
<td>( \frac{1}{2} )</td>
</tr>
</tbody>
</table>

Moreover, suppose that the true model is \( \theta^{Y} \). Suppose that \( u = \text{Id}, \delta = 0 \) and \( \phi (x) = \exp (-0.57536 x) \) so that \( \phi' (x) = -0.57536 \exp (-0.57536 x) \). With this belief, and since the DM is myopic, she is going bet on \( y \). With probability 1, she will observe a success, updating the beliefs in favor of model \( \theta^{Y} \). However, the smooth updating rule prescribes no updating

\[
\mu ((\theta^{Y}) |^{HK} (y, 1)) = \frac{\phi' (1) \mu (\theta^{Y}) \theta^{Y} (y)}{\phi' (1) \mu (\theta^{Y}) \theta^{Y} (y) + \phi' (\frac{1}{2}) \mu (\theta^{uni}) \theta^{uni} (y)}
\]

\[
= \frac{\phi' (1) \theta^{Y} (y)}{\phi' (1) \theta^{Y} (y) + \phi' (\frac{1}{2}) \theta^{uni} (y)}
\]

\[
= \frac{\phi' (1) \frac{3}{4}}{\phi' (1) \frac{1}{2} + \phi' (\frac{1}{2}) \frac{3}{4}} = \frac{1}{2}.
\]

Therefore, even if \( (y, \mu, \theta^{Y}) \) is not a SCE, it is a rest point of the process of actions and beliefs under the smooth updating rule.
References


