INNOCENZO
GASPARINI
INSTITUTE
FOR
ECONOMIC
RESEARCH

Institutional Members: CEPR, NBER and Università Bocconi

WORKING PAPER SERIES

Learning and Self-confirming Long-Run Biases

P. Battigalli, A. Francetich, G. Lanzani, M. Marinacci

Working Paper n. 588

This Version: April 2, 2017

IGIER – Università Bocconi, Via Guglielmo Röntgen 1, 20136 Milano –Italy
http://www.igier.unibocconi.it

The opinions expressed in the working papers are those of the authors alone, and not those of the Institute, which takes non institutional policy position, nor those of CEPR, NBER or Università Bocconi.

Learning and Self-confirming Long-Run Biases*

P. Battigalli,[a] A. Francetich,[b] G. Lanzani,[a] M. Marinacci[a]

[a]Department of Decision Sciences and IGIER - Università Bocconi

[b]Department of Economics - University of Washington Bothell

2nd April 2017

Abstract

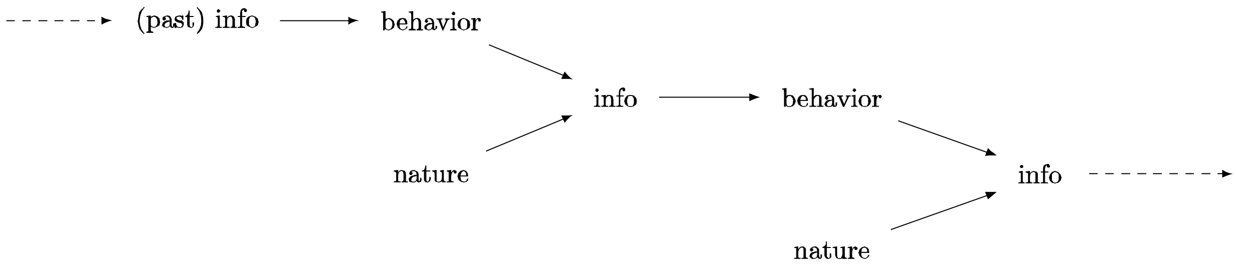
We consider an uncertainty averse, sophisticated decision maker facing a recurrent decision problem where information is generated endogenously. In this context, we study self-confirming strategies as the outcomes of a process of active experimentation. We provide inter alia a learning foundation for self-confirming equilibrium with model uncertainty (Battigalli et al., 2015). We also argue that ambiguity aversion tends to stifle experimentation, increasing the likelihood that decision maker get 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). The flow of actions and information can be diagrammed as follows:

We consider an ambiguity averse, not infinitely patient DM who is uncertain about the data generating process followed by nature. This uncertainty is represented by means of 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 beliefs about models are updated according to Bayes rule. However, note that an uncertainty

*We thank seminar audiences at Chicago, Tel Aviv and at RUD2013, Nicodemo De Vito, Ignacio Esponda, Elia Sartori, 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) and the AXA Research Fund is gratefully acknowledged.



averse DM may feature instances of dynamically inconsistent preferences (see, e.g., Siniscalchi 2011). We allow for reversals of preferences, but we assume a sophisticated DM. Indeed, he figures out these possible inconsistencies and formulates a strategy that is *dynamically consistent* because it satisfies the one-shot-deviation property: there is no situation where 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.

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 Section 5 for a formal connection between the two setups. However, note that 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 a 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 into the support of the DM's subjective belief. Finally, the first part of our paper considers the more general case of not i.i.d. data generating process. Therefore, the different underlying hypotheses lead us to provide a novel convergence result under a consistency assumption.

We study how self-confirming strategies arise from an active experimentation process. We show that a long-run bias emerges that favors "tested" actions, namely, actions on which information has been collected over time. The intuition behind our result is that tested actions become "certainty traps": 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 then implies a bias toward tested actions—"exploitation" rather than "exploration." More precisely, we show that the stochastic process of beliefs and actions converges with probability one to a random limit 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.

In the particular case of ambiguity neutrality and i.i.d. data generating process, our result is a version and restatement of the convergence result of EK: We connect our framework to EK and clarify that under subjective expected utility maximization (even if the DM cares about the future) the limit action-belief pair must be a self-confirming equilibrium of the game repeatedly played by the DM against nature, that is, the action is a short-run best reply to the DM's belief about nature's randomized strategy, and this belief is consistent with the long-run frequencies of observable outcomes.[1] Since the latter 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. While this may seem obvious ex-post, it is nonetheless worth noting given the almost exclusive reliance of economic theory on the Nash equilibrium concept.[2]

Although our results are derived in a single-person framework, their relevance extends to games more generally. We can interpret our DM as an agent in a large population of individuals playing a game recurrently in a given role, against agents drawn at random from large populations to play in different roles. The state of nature is then interpreted as the action profile of the other agents with whom DM has been matched. In particular, the i.i.d. case obtains in an environment similar to the steady-state learning model of Fudenberg and Levine (1993), where individual agents learn through their life, but the populations' 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 limit belief is confirmed by the evidence generated by the action and the steady state distribution of opponents strategies. Since 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

[1] 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).

[2] Self-confirming equilibrium is discussed in the monograph by Fudenberg and Levine (1998), but—to our knowledge—it is notably absent in published textbooks of microeconomic theory and game theory.

can only become less attractive as ambiguity aversion increases. We also show by example that higher ambiguity aversion tends to hinder experimentation and makes convergence to objectively optimal behavior (the best reply to the correct model) less likely.

The paper is structured as follows. Section 2 presents the static and dynamic decision framework and some preliminary notation. Section 3 describes the endogenous information process. Section 4 presents self-confirming equilibria and our learning results. Section 5 discusses the relationships between our setup and the one used in the literature on active learning, as exemplified by Easley and Kiefer (1988). Section 6 briefly relates our analysis to the literature on belief learning in games and concludes. All proofs are collected in the Appendix.

2 Framework

2.1 Static environment

Let S be a finite space of *states of nature* and let C be a *consequence 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 consequences by means of a *consequence function* $\rho : A \times S \to C$. The triple (A, S, ρ) is the basic structure of the decision problem.

Given any probability measure p on S, each action a is evaluated through its expected utility:

$$R(a,p) := \mathbb{E}_p \left[u \circ \rho_a \right] = \sum_{s \in S} \left(u \circ \rho \right) (a,s) \, p\left(s\right),$$

where $\rho_a = \rho(a, \cdot)$ is the section of ρ at a, and $u : C \to \mathbb{R}$ is a von Neumann-Morgenstern utility function. It is often convenient to write the criterion in the expected payoff form $R(a,p) = \sum_{s \in S} r(a,s)\, p(s)$, where $r : A \times S \to \mathbb{R}$ is the *payoff* (or *reward*) *function* $r := u \circ \rho$.

Let Δ be the collection of all probability measures on S. We identify Δ with the simplex of dimension $|S| - 1$, and we endow it with the Borel σ-algebra $\mathcal{B}(\Delta)$. If the probability model p is unknown, namely, if there is model uncertainty (cf. Marinacci 2015), the DM posits a closed set $P \subseteq \Delta$ of possible probability models and ranks actions according to the *smooth ambiguity* criterion of Klibanoff et al.:

$$V\left(a,\mu\right) := \phi^{-1}\left(\int_P \phi\left(R\left(a,p\right)\right) \mu\left(\mathrm{d}p\right)\right), \tag{1}$$

where μ is a *prior probability* measure on $(P, \mathcal{B}(P))$, and ϕ is a strictly increasing and continuous real-valued function that describes attitudes towards ambiguity. In particular, a concave ϕ captures ambiguity aversion, while a linear ϕ (e.g., the identity) corresponds to ambiguity neutrality:

$$V(a,\mu) = \int_P R(a,p)\,\mu(\mathrm{d}p) = \int_S r(a,s)\,p_\mu(\mathrm{d}s) = R(a,p_\mu),$$

where $p_\mu \in \Delta$ is the *predictive probability* given by $p_\mu(E) := \int_P p(E)\,\mu(\mathrm{d}p)$ for all $E \subseteq S$. This is the *classical subjective expected utility criterion* (Cerreia-Vioglio et al. 2013). Finally, note that:

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

(ii) The limit case of criterion (1) as ambiguity aversion increases is a version of the maxmin criterion $\inf_{p \in \operatorname{supp}\mu} R(a,p)$ of Gilboa and Schmeidler (1989).

The static decision problem can be summarized by the seven-tuple:

$$\Gamma = (A, S, C, \rho, u, P, \phi). \tag{2}$$

2.2 Dynamic environment

Notation Finite spaces are endowed with their power sets as σ-algebras. For each time t, let Z_t be a corresponding finite space of possible "realizations," where Z_t may be independent of t.[3] We let $Z^t = \prod_{i=1}^t Z_i$ and $Z^\infty = \prod_{t=1}^\infty Z_t$; Z^t and Z^∞ are, respectively, the spaces of finite and infinite histories. The space Z^∞ is endowed with the Borel σ-algebra, $\mathcal{B}(Z^\infty)$, corresponding to the product topology on Z^∞; this is the same as the σ-algebra generated by the elementary cylinders $\{z_1\} \times \cdots \times \{z_t\} \times Z \times \cdots$. We denote by $z^t = (z_1, ..., z_t) \in Z^t$ both the histories and the elementary cylinders that they identify by means of 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^∞. Finally, we denote by $\Delta(Z^\infty)$ the collection of all probability measures defined on $(Z^\infty, \mathcal{B}(Z^\infty))$, and by $\mathcal{B}(\Delta(Z^\infty))$, the Borel σ-algebra on $\Delta(Z^\infty)$ generated by the weak*-star topology, which coincides with the sigma-algebra generated by the evaluation maps $p \mapsto p(B)$ for all $B \in \mathcal{B}(Z^\infty)$.

[3] Unless 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.

Environment Given S, let $\left(S^{\infty}, \mathcal{B}\left(S^{\infty}\right), \bar{p}\right)$ be the probability space on which a coordinate state process $\left(\mathbf{s}_1, \mathbf{s}_2, ...\right)$ is defined, with $\mathbf{s}_t : S^{\infty} \to S$ for each t.[4] We will use the less demanding notation $\mathbf{s}^{\infty}$ for the state process.

The *state process* $\mathbf{s}^{\infty}$ describes the exogenous uncertainty in the decision problem. It can be seen as the environment of the problem. Its realizations are denoted by $s^{\infty} \in S^{\infty}$. To ease notation, we set $\mathbf{s}^t = \left(\mathbf{s}_1, ..., \mathbf{s}_t\right)$. We denote by $\sigma\left(\mathbf{s}^t\right)$ the σ-algebra generated by the random variables $\mathbf{s}_1, ..., \mathbf{s}_t$, namely, by the process up to time t. Thus, $\left(\sigma\left(\mathbf{s}^t\right)\right)$ is the basic filtration for the problem. We denote by $\sigma\left(\mathbf{s}^0\right)$ the trivial σ-algebra $\left\{\emptyset, S^{\infty}\right\}$.

For each time t, we can also regard $\sigma\left(\mathbf{s}^t\right)$ as a subset of 2^{S^t} by identifying the elements of $\sigma\left(\mathbf{s}^t\right)$ with their projection on S^t. In what follows, we often maintain this small abuse of notation for the natural filtration as well as for any sub-filtration.

Actions and outcomes The DM's choices are described by a sequence $\left(a_t\right) \in A^{\infty}$ that consists of an action for each time t. At each such t, there is a time-independent one-period *consequence function* $\rho : A \times S \to C$. Here, $\rho\left(a_t, s_t\right)$ 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. For convenience, we assume that the problem is stationary, i.e., its elements are time invariant.

Information feedback In a dynamic setting, the DM may receive some feedback, that is, a "message" generated by the outcomes of past actions. For example, he may observe the (random) consequences of such actions. This feedback is a source of "endogenous" (choice dependent) information about states that is peculiar to the dynamic setting and that will play a key role in the paper.

This endogenous information is modelled through a time invariant *feedback function* $f : A \times S \to M$, where M is a (finite) message space. By selecting action $a_t \in A$ at time t, the DM receives ex post a *message* $m_t = f\left(a_t, s_t\right)$ if state s_t realizes. A DM who selects action a_t and ex post receives the message m_t knows that the true state s_t belongs to the set $\left\{s_t \in S : f\left(a_t, s_t\right) = m_t\right\} = f_{a_t}^{-1}\left(m_t\right)$.[5]

Actions and messages 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 messages $m^{t-1} = \left(m_1, ..., m_{t-1}\right)$ received in the previous periods as a result of the history of actions $a^{t-1} = \left(a_1, ..., a_{t-1}\right)$ and states $s^{t-1} = \left(s_1, ..., s_{t-1}\right)$.[6]

[4] We use boldface letters for random variables and normal letters for realizations.

[5] As in the case of ρ_a, f_a denotes the section of f at a, $f_a(\cdot) := f(a, \cdot)$.

[6] We distinguish three point in time within each period: the ex ante time (prior to the decision), the decision time, and the ex post time (after the decision). Any information available ex post at

We assume that consequences are observable: There exists a function $\gamma : A \times M \to C$ such that

$$\forall a \in A, \ \rho_a = \gamma_a \circ f_a. \tag{3}$$

This condition ensures that the feedback function reflects at least the information on states determined by outcome observability. In particular, ex post information about the state is typically endogenous, that is, the partition

$$\left\{ f_a^{-1}(m) : m \in M \right\} \subseteq 2^S$$

of the state space S induced by the messages depends, in general, on the choice of action a.[7]

If the DM receives the same information about states regardless of his action, namely, if

$$\forall a, a' \in A, \left\{ f_a^{-1}(m) : m \in M \right\} = \left\{ f_{a'}^{-1}(m) : m \in M \right\},$$

we say that feedback satisfies *own-action independence.* This is the case, for instance, if $f_a = f_{a'}$ for all $a, a' \in A$. In particular, there is *perfect feedback* when the DM ex post observes the realized state s_t; that is, f_a is injective for each $a \in A$.

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 from the urn; therefore the possible bets are b, g, and 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 $P = \{p \in \Delta : p(B) = 1/3\}$. Ex post, after the draw, he only learns the result of his bet, namely, whether or not he wins 1 euro. In other words, the messages are the obtained prizes. Here, $S = \{B, G, Y\}$, $A = \{b, g, y\}$ and $C = M = \{0, 1\}$. The consequence function is

$$\begin{aligned} \rho(b, B) &= \rho(y, Y) = \rho(g, G) = 1; \\ \rho(b, Y) &= \rho(b, G) = \rho(g, B) = \rho(g, Y) = \rho(y, B) = \rho(y, G) = 0. \end{aligned}$$

The feedback function coincides with the consequence function, $f = \rho$, and is described by the following table:

ρ, f	B	Y	G
b	1	0	0
y	0	1	0
g	0	0	1

period t is also available ex ante at $t + 1$.

[7] If f_a is not onto, the collection of subsets $\left\{ f_a^{-1}(m) : m \in M \right\}$ also contains the empty set, hence it is not a partition according to the standard meaning. We neglet for simplicity this minor detail, which does not affect our analysis.

Therefore, we have:

$$
\begin{aligned}
f_b^{-1}(1) &= \{B\}, \quad f_b^{-1}(0) = \{Y, G\}, \\
f_y^{-1}(1) &= \{Y\}, \quad f_y^{-1}(0) = \{B, G\}, \\
f_g^{-1}(1) &= \{G\}, \quad f_g^{-1}(0) = \{B, Y\}.
\end{aligned}
$$

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 $\{\{G\}, \{B, Y\}\}$. If, instead, we assume that $M = \{0, 1\} \times S$ and

$$\forall (a, s) \in A \times S,\ f(a, s) = (\rho(a, s), s),$$

then the DM observes the color of the ball drawn—on top of the payoff—and we have perfect feedback. In this case, betting on any color always yields the partition $\{\{B\}, \{G\}, \{Y\}\}$. ▲

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, $S = \{B_I B_{II}, B_I G_{II}, G_I B_{II}, G_I G_{II}\}$, $A = \{I, II\}$, and $C = M = \{0, 1\}$; the messages are the prizes. The consequence function and feedback function coincide and are described by the following table:

ρ, f	$B_I B_{II}$	$B_I G_{II}$	$G_I B_{II}$	$G_I G_{II}$
I	0	0	1	1
II	0	1	0	1

Therefore:

$$
\begin{aligned}
f_I^{-1}(1) &= \{G_I B_{II}, G_I G_{II}\}, \quad f_I^{-1}(0) = \{B_I B_{II}, B_I G_{II}\}, \\
f_{II}^{-1}(1) &= \{B_I G_{II}, G_I G_{II}\}, \quad f_{II}^{-1}(0) = \{B_I B_{II}, G_I B_{II}\}.
\end{aligned}
$$

Thus, betting on the first urn yields the partition $\{\{G_I B_{II}, G_I G_{II}\}, \{B_I B_{II}, B_I G_{II}\}\}$, while betting on the second urn yields the partition $\{\{B_I G_{II}, G_I G_{II}\}, \{B_I B_{II}, G_I B_{II}\}\}$. 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 messages, a^{t-1} and m^{t-1}. The *ex ante information history* h_t at time t is given by:

$$h_1 = \left(a^0, m^0\right); \quad \forall t > 1,\ h_t = \left(a^{t-1}, m^{t-1}\right) = (h_{t-1}, a_{t-1}, m_{t-1}),$$

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

$$H_{t+1} = \left\{(a^t, m^t) \in A^t \times M^t : \exists s^t \in S^t, \forall k \in \{1, ..., t\}, m_k = f(a_k, s_k)\right\}.$$

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 $\alpha_t : H_t \to A$ for each t. Since $H_1 = \{(a^0, m^0)\}$ is a singleton, the first term α_1 prescribes a non-contingent action.

Information and strategies A state process $\mathbf{s}^\infty$ and a strategy $\alpha = (\alpha_t)$ recursively induce an action process $(\mathbf{a}_t^\alpha)$, a message process $(\mathbf{m}_t^\alpha)$ and an information process $\mathbf{h}^\alpha = (\mathbf{h}_t^\alpha)$, as follows:

(i) $\mathbf{a}_1^\alpha = \alpha_1 (a^0, m^0)$ and $\mathbf{a}_t^\alpha = \alpha_t (\mathbf{h}_t^\alpha)$ for each $t > 1$;

(ii) $\mathbf{m}_t^\alpha = f (\mathbf{a}_t^\alpha, \mathbf{s}_t)$ for each t;

(iii) $\mathbf{h}_1^\alpha = (a^0, m^0)$ and $\mathbf{h}_{t+1}^\alpha = (\mathbf{h}_t^\alpha, \mathbf{a}_t^\alpha, \mathbf{m}_t^\alpha)$ for each t.

In words, at each period t, an action a_t is selected according to the time-t strategy α_t based on the information history $h_t = (h_{t-1}, a_{t-1}, m_{t-1})$. In turn, its execution generates a message m_t that may be considered in subsequent periods. Note that α_1 prescribes only one action, $\alpha_1 (s^0, m^0)$, which, together with realization s_1 of $\mathbf{s}_1$, initializes the recursion by sending message m_1.

The sequence of σ-algebras $(\sigma (\mathbf{h}_t^\alpha))$ on S^∞ generated by the information process $(\mathbf{h}_t^\alpha)$ is a filtration that describes the information structure generated and used by strategy α. Since feedback will typically not be perfect, this filtration is coarser than the one generated by the state process $\mathbf{s}$; that is, $\sigma (\mathbf{h}_t^\alpha) \subseteq \sigma (\mathbf{s}^{t-1})$ for each $t > 1$. For this reason, without loss of generality, we can regard $\mathbf{h}_t^\alpha$, and also $\mathbf{a}_t^\alpha$ and $\mathbf{m}_{t-1}^\alpha$, as functions defined on S^{t-1}.[8]

Identification correspondence Information history $h_t = (a^{t-1}, m^{t-1})$ yields the *information set:*

$$I (h_t) := \prod_{\tau=1}^{t-1} f_{a_\tau}^{-1}(m_\tau) = \left\{s^{t-1} \in S^{t-1} : \forall \tau \in \{1, \ldots, t-1\}, f (a_\tau, s_\tau) = m_\tau\right\}.$$

This is what information history h_t says about the past state history s^{t-1}. A priori, a strategy α can reach all information histories that belong to the image of $\mathbf{h}_t^\alpha$; for

[8] Recall that $\sigma (\mathbf{s}^0)$ is the trivial σ-algebra.

any reachable h_t, it holds that $I(h_t) = (\mathbf{h}_t^\alpha)^{-1}(h_t)$.[9] This is what h_t says about s^{t-1} to a DM who reached it by using strategy α. In particular, fix a state history $\bar{s}^{t-1}$; the set $I(\mathbf{h}_t^\alpha(\bar{s}^{t-1}))$ is the collection of all state histories s^{t-1} that are observationally equivalent to $\bar{s}^{t-1}$ given information $\mathbf{h}_t^\alpha(\bar{s}^{t-1})$. In view of this, define $\iota_t^\alpha : S^{t-1} \to 2^{S^{t-1}}$ by

$$\iota_t^\alpha\left(s^{t-1}\right) := I\left(\mathbf{h}_t^\alpha\left(s^{t-1}\right)\right) = \left((\mathbf{h}_t^\alpha)^{-1} \circ \mathbf{h}_t^\alpha\right)\left(s^{t-1}\right). \tag{4}$$

We can regard ι_t^α as the *identification correspondence* determined by α 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 α.

Clearly, $s^{t-1} \in \iota_t^\alpha(s^{t-1})$, and so the correspondence induces a partition. We have *perfect* (*state*) *identification* under α when $\iota_t^\alpha(s^{t-1}) = \{s^{t-1}\}$ 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 α of the identification correspondence will play a key role in our results. Of course, there is no such dependence under own-action independence of feedback; in this case, we can write $\iota_t(s^{t-1})$. In particular, under perfect feedback we have $\iota_t(s^{t-1}) = \{s^{t-1}\}$.

Identification algebra As noted above, the collection:

$$\left\{I \subseteq S^{t-1} : \exists s^{t-1} \in S^{t-1}, I = \iota_t^\alpha\left(s^{t-1}\right)\right\} = \left\{I \subseteq S^{t-1} : \exists h_t \in \operatorname{Im} \mathbf{h}_t^\alpha, I = I(h_t)\right\} \tag{5}$$

is a partition of S^{t-1}. The sigma algebra $\sigma(\mathbf{h}_t^\alpha) \subseteq 2^{S^{t-1}}$ is in fact generated by this partition; for this reason, we call it the *identification* (*sigma*) *algebra* determined by strategy α at time t.[10] Under perfect feedback, we have $\sigma(\mathbf{h}_t^\alpha) = 2^{S^{t-1}}$; otherwise, $\sigma(\mathbf{h}_t^\alpha)$ can be coarser than $2^{S^{t-1}}$.

Thus, the filtration $(\sigma(\mathbf{h}_t^\alpha))$ models the evolution of the DM's information about the state histories s^t. Events that belong to this filtration are called *α-observable.* The filtration is increasing; that is, $\sigma(\mathbf{h}_t^\alpha) \subseteq \sigma\left(\mathbf{h}_{t+1}^\alpha\right)$ for each t. We denote by $\sigma(\mathbf{h}^\alpha) = \sigma(\cup_t \sigma(\mathbf{h}_t^\alpha))$ the σ-algebra generated by this filtration. In other words, $\sigma(\mathbf{h}^\alpha)$ is the sigma algebra generated by the *α-observable events.* Under own-action independence, information does not depend on strategy α. For example, under perfect feedback, we have $\sigma(\mathbf{h}^\alpha) = \mathcal{B}(S^\infty)$ for each strategy α.

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:

ρ, f	B	Y	G
b	1	0	0
y	0	1	0

[9] More precisely, $I(h_t)$ is the projection on S^{t-1} of the pre-image of h_t under $\mathbf{h}_t^\alpha$.

[10] More precisely, the identification algebra is the projection on S^{t-1} of $\sigma(\mathbf{h}_t^\alpha)$.

Throughout our leading example, we will consider two strategies, denoted by α^{NE} (No Experimentation) and α^{E} (Experimentation). Strategy α^{NE} bets on black forever; while strategy α^{E} experiments with yellow in period 1, and, from period 2 onwards, the action selected depends on the result of this experimentation: If a success is observed in period 1, y is chosen forever, otherwise b is chosen forever.[11] Formally:

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

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

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},

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

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

$$\mathbf{h}_{t+1}^{\alpha^{NE}}\left(s^{t}\right) = \begin{cases} \left(\mathbf{h}_t^{\alpha^{NE}}\left(s^{t-1}\right), b, 1\right) & \text{if } s_t = B, \\ \left(\mathbf{h}_t^{\alpha^{NE}}\left(s^{t-1}\right), b, 0\right) & \text{if } s_t \in \{Y, G\}, \end{cases} .$$

Moreover,

$$I(\mathbf{h}_2^{\alpha^{NE}}(s_1)) = \begin{cases} \{B\} & \text{if } s_1 = B, \\ \{G, Y\} & \text{if } s_2 \in \{G, Y\}, \end{cases}$$

and, for each t and s^{t-1},

$$I(\mathbf{h}_{t+1}^{\alpha^{NE}}\left(s^{t-1}, B\right)) = I(\mathbf{h}_t^{\alpha^{NE}}\left(s^{t-1}\right)) \times \{B\},$$

$$I(\mathbf{h}_{t+1}^{\alpha^{NE}}\left(s^{t-1}, G\right)) = I(\mathbf{h}_{t+1}^{\alpha^{NE}}\left(s^{t-1}, Y\right)) = I(\mathbf{h}_t^{\alpha^{NE}}\left(s^{t-1}\right)) \times \{G, Y\}.$$

The identification algebra is then $\sigma(\mathbf{h}_t^{\alpha^{NE}}) = \{\emptyset, \{B\}, \{G, Y\}, S\}^{t-1}$. In particular, $I(\mathbf{h}_{t+1}^{\alpha^{NE}}(s^t)) = \{s^t\}$ if and only if $s^t = B^t$ (B t-times). Thus, strategy α^{NE} only allows partial identification.

[11] In this paragraph, we do not explicitly comment on behavior at information histories ruled out by the strategy itself.

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

$$\alpha_t^E\left(h_t\right) = \begin{cases} y & \text{if } t = 1, \\ b & \text{if } t > 1 \text{ and } a_{t-1} = b, \\ y & \text{if } t > 1,\ a_{t-1} = y \text{ and } (y,1) \in \{(a_1, m_1), \dots, (a_{t-1}, m_{t-1})\}, \\ b & \text{if } t > 1,\ a_{t-1} = y \text{ and } (y,1) \notin \{(a_1, m_1), \dots, (a_{t-1}, m_{t-1})\}. \end{cases}$$

The only difference between this strategy and α^{NE} is the action chosen in the first period.

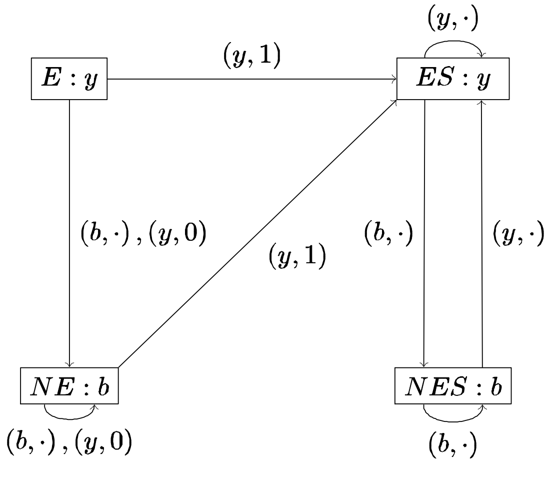
Such strategies are more easily understood as finite automata. In particular, we can consider a common set of states and the same decision and transition functions, the only difference being given by the initial state:

- The set of states is $Q = \{E, NE, ES, NES\}$; intuitively, states of the form XS (with $X = E, NE$) are reached after having experienced at least one Success betting on yellow.
- The initial states for α^{NE} and α^E are, respectively, $q^{\alpha^{NE}} = NE$ and $q^{\alpha^E} = E$;
- The decision function is

$$d\left(q\right) = \begin{cases} y & \text{if } q \in \{E, ES\}, \\ b & \text{if } q \in \{NE, NES\}. \end{cases}$$

- The transition function is

$$\tau\left(q, a, m\right) = \begin{cases} NE & \text{if } (q, a, m) \in \left\{\left(E, b, \cdot\right), \left(E, y, 0\right), \left(NE, b, \cdot\right), \left(NE, y, 0\right)\right\}, \\ ES & \text{if } (q, a, m) \in \left\{\left(\cdot, y, 1\right), \left(NES, y, 0\right), \left(ES, y, 0\right)\right\}, \\ NES & \text{if } (q, a, m) \in \{\left(ES, b, \cdot\right), \left(NES, b, \cdot\right)\}. \end{cases}$$



Next we describe the induced processes of actions and messages:

$$\mathbf{a}_1^{\alpha^E} = y,$$

$$\mathbf{m}_1^{\alpha^E}(s_1) = \begin{cases} 1 & \text{if } s_1 = Y, \\ 0 & \text{if } s_1 \in \{B, G\}, \end{cases}$$

$$\mathbf{h}_2^{\alpha^E}(s_1) = \begin{cases} (y, 1) & \text{if } s_1 = Y, \\ (y, 0) & \text{if } s_1 \in \{B, G\}, \end{cases}$$

and, for each $t > 1$ and s^{t-1},

$$\mathbf{a}_t^{\alpha^E}\left(s^{t-1}\right) = \begin{cases} y & \text{if } s_1 = Y, \\ b & \text{else}, \end{cases}$$

$$\mathbf{m}_t^{\alpha^E}\left(s^t\right) = \begin{cases} 1 & \text{if } s_1 = Y \text{ and } s_t = Y, \\ 1 & \text{if } s_1 \in \{B, G\} \text{ and } s_t = B, \\ 0 & \text{else}, \end{cases}$$

$$\mathbf{h}_{t+1}^{\alpha^E}\left(s^t\right) = \begin{cases} \left(\mathbf{h}_t^{\alpha^E}\left(s^{t-1}\right), y, 1\right) & \text{if } s_1 = Y \text{ and } s_t = Y, \\ \left(\mathbf{h}_t^{\alpha^E}\left(s^{t-1}\right), b, 1\right) & \text{if } s_1 \in \{B, G\} \text{ and } s_t = B, \\ \left(\mathbf{h}_t^{\alpha^E}\left(s^{t-1}\right), y, 0\right) & \text{if } s_1 = Y \text{ and } s_t \in \{B, G\}, \\ \left(\mathbf{h}_t^{\alpha^E}\left(s^{t-1}\right), b, 0\right) & \text{if } s_1 \in \{B, G\} \text{ and } s_t \in \{Y, G\}. \end{cases}$$

Since

$$I\left(\mathbf{h}_{t+1}^{\alpha^E}\left(s^t\right)\right) = \begin{cases} \{Y^t\} & \text{if } s^t = Y^t, \\ \{Y^{t-1}\} \times \{B, G\} & \text{if } s^t \in \{Y^{t-1}\} \times \{B, G\}, \\ \cdots & \cdots \\ \{B, G\} \times \{B^{t-1}\} & \text{if } s^t = B^t, \\ \{B, G\} \times \{Y, G\}^{t-1} & \text{if } s^t = G^t, \end{cases}$$

the identification algebra is

$$\sigma\left(\mathbf{h}_{t+1}^{\alpha^E}\right) = \left\{\emptyset, \{Y^t\}, \{Y^{t-1}\} \times \{B, G\}, \ldots, \{B, G\} \times \{B^{t-1}\}, \{B, G\} \times \{(Y, G)^{t-1}\}, S\right\}.$$

For instance, if $t = 3$, we have:

$$\begin{aligned} \sigma\left(\mathbf{h}_3^{\alpha^E}\right) &= \big\{\emptyset, \{Y^3\}, \{Y^2\} \times \{B, G\}, \{Y\} \times \{B, G\} \times \{Y\}, \\ &\quad \{Y\} \times \{(B, G)^2\}, \{B, G\} \times \{B^2\}, \{B, G\} \times \{B\} \times \{Y, G\} \\ &\quad \{B, G\} \times \{Y, G\} \times \{B\}, \{B, G\} \times \{(Y, G)^2\}, S\big\}. \end{aligned}$$

Note that $I(\mathbf{h}_t^{\alpha^E}(s^t)) = \{s^{t-1}\}$ if and only if $s^t = Y^t$: The DM learns the past state history if and only if yellow happens to be drawn at each time. ▲

3 Models, learning, and payoffs

3.1 Distributions and updating

Distributions Fix a probability model of the stochastic process of states $p : \mathcal{B}(S^\infty) \to [0,1]$. If information history h_t is reached, the conditional probability $p(\cdot \mid h_t) : \mathcal{B}(S^\infty) \to [0,1]$ is defined as $p(B \mid h_t) = p(B \cap I(h_t)) / p(I(h_t))$ for each $B \in \mathcal{B}(S^\infty)$, if $p(I(h_t)) > 0$. The regular conditional probability $p(\cdot \mid \mathbf{h}_t^\alpha(\cdot))$ keeps track of the information histories that strategy α can actually reach at each period t. Recall that, for each event $B \in \mathcal{B}(S^\infty)$, $p(B \mid \mathbf{h}_t^\alpha(\cdot))$ is a $\sigma(\mathbf{h}_t^\alpha)$-measurable function.[12] The conditional probability at time t, $p(\cdot \mid \mathbf{h}_t^\alpha(\cdot))$, is an α-dependent random measure, because it is a function of the random information history $\mathbf{h}_t^\alpha$ generated by strategy α.

Predictive and posterior probabilities A measure $\mu : \mathcal{B}(\Delta(S^\infty)) \to [0,1]$ is called a *prior probability.* A prior induces a *predictive distribution* $p_\mu \in \Delta(S^\infty)$ defined by $p_\mu(B) = \int_{\Delta(S^\infty)} p(B)\,\mu(\mathrm{d}p)$ for all $B \in \mathcal{B}(S^\infty)$. Moreover, for each t, we denote by $\mu(\cdot \mid h_t)$ the *posterior* of μ given information history h_t.[13] Recall that, for each $D \in \mathcal{B}(\Delta(S^\infty))$,

$$\mu(D|h_t) = \frac{1}{p_\mu(I(h_t))} \int_D p(I(h_t))\,\mu(\mathrm{d}p),$$

provided that $p_\mu(I(h_t)) > 0$. The *conditional predictive distribution* $p_\mu(\cdot \mid h_t)$ is such that, for each $B \in \mathcal{B}(S^\infty)$,

$$p_\mu(B \mid h_t) = \int_{\Delta(S^\infty)} p(B \mid h_t)\,\mu(\mathrm{d}p \mid h_t).$$

Observationally equivalent models Given any $p \in \Delta(S^\infty)$, denote by $p^\alpha : \sigma(\mathbf{h}^\alpha) \to [0,1]$ its restriction to the sigma algebra $\sigma(\mathbf{h}^\alpha)$ generated by the α-observable events. Fix a true model $\bar{p}$; the $\sigma(\mathbf{h}_t^\alpha)$-measurable correspondence $\hat{P}_t^{\alpha,\mu}(\bar{p}) : S^{t-1} \to 2^{\Delta(S^\infty)}$ given by

$$\hat{P}_t^{\alpha,\mu}(\bar{p})\left(s^{t-1}\right) = \left\{p \in \mathrm{supp}\mu\left(\cdot|\mathbf{h}_t^\alpha\left(s^{t-1}\right)\right) : p^\alpha\left(\cdot|\mathbf{h}_t^\alpha\left(s^{t-1}\right)\right) = \bar{p}^\alpha\left(\cdot|\mathbf{h}_t^\alpha\left(s^{t-1}\right)\right)\right\}$$

represents the collection of models that are deemed possible and that, conditional on $\mathbf{h}_t^\alpha(s^{t-1})$, are *observationally equivalent* to the true model $\bar{p}$ under strategy α and prior μ.[14] Note that, for some s^{t-1}, the set $\hat{P}_t^{\alpha,\mu}(\bar{p})(s^{t-1})$ may be empty if $\bar{p} \notin \mathrm{supp}\,\mu$.

The next lemma establishes a monotonicity property of this correspondence with respect to time.

[12] See Appendix 7.1 for a formal definition of regular conditional probability.

[13] See Appendix 7.1 for a formal definition.

[14] It is actually enough to require $p^\alpha\left(E \mid \mathbf{h}_t^\alpha\left(s^{t-1}\right)\right) = \bar{p}^\alpha\left(E \mid \mathbf{h}_t^\alpha\left(s^{t-1}\right)\right)$ for all $E \in \cup_{t\geq 1}\sigma(\mathbf{h}_t^\alpha)$. That is, observational equivalence is determined by the α-observable events.

Lemma 1. *For every period t, $\hat{P}_t^{\alpha,\mu}(\bar{p})(\cdot) \subseteq \hat{P}_{t+1}^{\alpha,\mu}(\bar{p})(\cdot)$ $\bar{p}$-almost surely.*

The intuition behind the lemma is as follows. The set $\hat{P}_t^{\alpha,\mu}(\bar{p})(s^t)$ may contain models that disagree with $\bar{p}$ on the relative probabilities of past events (up to t), but that agree with $\bar{p}$ on the relative probabilities of future events (from $t+1$). Almost every model that agrees with $\bar{p}$ on future events conditional on information up to t also agrees on future events conditional on information up to $t+1$.

It follows from the lemma that, $\bar{p}$-a.s.,

$$\hat{P}_1^{\alpha,\mu}(\bar{p}) := \{p \in \operatorname{supp}\mu : p^\alpha = \bar{p}^\alpha\} \subseteq \hat{P}_t^{\alpha,\mu}(\bar{p})(\cdot)$$

for every t. The set $\hat{P}_1^{\alpha,\mu}(\bar{p})$ represents the irreducible model uncertainty that, when $\bar{p}$ is the true model, the DM faces if he plays α and holds belief μ.[15] When $\hat{P}_1^{\alpha,\mu}(\bar{p}) = \operatorname{supp}\mu$, such uncertainty does not allow any learning, as all the models that the DM initially deems possible are α-observationally equivalent to the true model. The opposite is true when $\hat{P}_1^{\alpha,\mu}(\bar{p}) = \{\bar{p}\}$, since in this case the DM will assign probability arbitrarily close to 1 to the true model as he accumulates observations.

Identification of models We say that p and $\bar{p}$ are *α-orthogonal*, which we denote by $p^\alpha \perp \bar{p}^\alpha$, if there is some event $E \in \sigma(\mathbf{h}^\alpha)$ such that $p(E) = 0$ and $\bar{p}(E) = 1$; in words, they are distinguishable—at least in the long run, or asymptotically—by some observable event, given strategy α.

The collection of models that are distinguishable from $\bar{p}$ conditional on $\mathbf{h}_t^\alpha$ under strategy α and prior μ is a $\sigma(\mathbf{h}_t^\alpha)$-measurable correspondence $\perp_t^{\alpha,\mu}(\bar{p}) : S^{t-1} \to 2^{\Delta(S^\infty)}$ given by:

$$\perp_t^{\alpha,\mu}(\bar{p})\left(s^{t-1}\right) = \left\{p \in \operatorname{supp}\mu\left(\cdot|\mathbf{h}_t^\alpha\left(s^{t-1}\right)\right) : p^\alpha\left(\cdot|\mathbf{h}_t^\alpha\left(s^{t-1}\right)\right) \perp \bar{p}^\alpha\left(\cdot|\mathbf{h}_t^\alpha\left(s^{t-1}\right)\right)\right\}.$$

The notions of observationally equivalent models and orthogonal models motivate the following definition. In what follows, we will often study the property of a triple $(\alpha, \mu, \bar{p})$. The interpretation is the following: α is the strategy that is carried on by the DM, μ is a prior probability that corresponds to his belief over models at period 0, and the probability measure $\bar{p}$ on $(S^\infty, \mathcal{B}(S^\infty))$ is the "correct" model, that is, the one characterizing the data generating process. In view of this, we study the triple $(\alpha, \mu, \bar{p})$ in order to understand what happens to a DM who follows strategy α when his prior probability is μ and the data generating process is $\bar{p}$. Therefore, we have a particular interest in the statements that hold $\bar{p}$-a.s., that is, that are almost surely true with respect to the correct model.

[15] 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.

Definition 1. *The triple $(\alpha, \mu, \bar{p})$, where α is a strategy, μ is a prior probability and $\bar{p}$ is a probability measure on $(S^\infty, \mathcal{B}(S^\infty))$, is* consistent *at time t if (i) $\hat{P}_t^{\alpha,\mu}(\bar{p})(\cdot) \neq \emptyset$ and (ii)* $\operatorname{supp}\mu(\cdot|\mathbf{h}_t^\alpha(\cdot)) = \hat{P}_t^{\alpha,\mu}(\bar{p})(\cdot) \cup \perp_t^\alpha(\bar{p})(\cdot)$ *$\bar{p}$-a.s.*

In words, the triple $(\alpha, \mu, \bar{p})$ is consistent[16] at time t if, conditional on the available information $\mathbf{h}_t^\alpha$,

(i) at least one model deemed possible is α-observationally equivalent to the true model;

(ii) the set of models deemed possible is partitioned into models that are α-observationally equivalent to the true model and those that are (asymptotically) α-distinguishable from it.

Suppose that $\operatorname{supp}\mu$ consists of i.i.d. models p_π parameterized by their marginals $\pi \in \Delta(S)$, so that it makes sense to write $\mu \in \Delta(\Delta(S))$ and $\hat{P}_t^{\alpha,\mu}(\bar{\pi})$. Let

$$\sigma_{\geq t}(\mathbf{h}^\alpha(s^{t-1})) = \left\{E \subseteq S^\infty : I\left(\mathbf{h}_t^\alpha\left(s^{t-1}\right)\right) \times E \in \sigma(\mathbf{h}^\alpha)\right\}$$

be the sigma-algebra of α-observable events from date t onwards given s^{t-1}. Then, $\hat{P}_t^{\alpha,\mu}(\bar{\pi})(s^{t-1})$ and $\perp_t^{\alpha,\mu}(\bar{\pi})(s^{t-1})$ are given by, respectively,

$$\left\{\pi \in \operatorname{supp}\mu\left(\cdot|\mathbf{h}_t^\alpha\left(s^{t-1}\right)\right) : \forall E \in \sigma_{\geq t}\left(\mathbf{h}^\alpha\left(s^{t-1}\right)\right), p_\pi^\alpha(E) = p_{\bar{\pi}}^\alpha(E)\right\}$$

and

$$\left\{\pi \in \operatorname{supp}\mu\left(\cdot|\mathbf{h}_t^\alpha\left(s^{t-1}\right)\right) : \exists E \in \sigma_{\geq t}\left(\mathbf{h}^\alpha\left(s^{t-1}\right)\right), p_\pi^\alpha(E) = 1 - p_{\bar{\pi}}^\alpha(E) \in \{0,1\}\right\}.$$

Hence, $(\alpha, \mu, \bar{p})$ is consistent at t if, for $\bar{p}$-almost every s^{t-1}, $\hat{P}_t^{\alpha,\mu}(\bar{\pi})(s^{t-1}) \neq \emptyset$ and, for each $p_\pi \in \operatorname{supp}\mu(\cdot \mid \mathbf{h}_t^\alpha(s^{t-1}))$, either $p_\pi^\alpha(E) = p_{\bar{\pi}}^\alpha(E)$ for all $E \in \sigma_{\geq t}(\mathbf{h}^\alpha(s^{t-1}))$ or $p_\pi^\alpha(E) = 1 - p_{\bar{\pi}}^\alpha(E) \in \{0,1\}$ for some $E \in \sigma_{\geq t}(\mathbf{h}^\alpha(s^{t-1}))$. Tail events are particularly useful in understanding if two models are orthogonal. Indeed, by Kolmogorov's 0-1 law**,** we know that these events have either probability 1 or probability 0 under a model. Therefore, if two models disagree over the probability of a tail event, they are automatically orthogonal.

The previous framework can be applied both to the case of a prior with a finite or an infinite support. However, to ease the analysis, from now on, we maintain the following assumption:

[16] The word consistent may recall the consistency criterion imposed in Arrow and Green (1973). However, note that 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.

Assumption The prior belief μ has finite support.

In view of Lemma 1, for a triple $(\alpha, \mu, \bar{p})$ it is easier to meet the conditions for consistency as t gets larger. We denote by $T = T(\alpha, \mu, \bar{p})$ 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.

Proposition 1. *Let $(\alpha, \mu, \bar{p})$ be consistent from some period $T \geq 1$. Then,*

$$\lim_{t \to \infty} \mu(\hat{P}_T^{\alpha,\mu}(\bar{p})(\cdot) \mid \mathbf{h}_t^{\alpha}(\cdot)) = 1 \;\; \bar{p}^{\alpha}\text{-}a.s. \tag{6}$$

In words, a triple $(\alpha, \mu, \bar{p})$ that is consistent from some period T allows the DM to learn in the long run the α-observable implications of the true model $\bar{p}$.[17]

The mapping $p \mapsto \hat{P}_{T(\alpha,\mu,p)}^{\alpha,\mu}(p)(\cdot)$ can be viewed as the (long-run) *model-identification map* determined by α and μ. We have *perfect model identification* when $\hat{P}_{T(\alpha,\mu,p)}^{\alpha,\mu}(p)(\cdot) = \{p\}$ for each p; in this case, the DM who holds belief μ and plays strategy α 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, $\hat{P}_1^{\alpha,\mu}(\bar{p}) = \{\bar{p}\}$.

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

$$\mu(\bar{p} \,|\cdot) \to 1 \;\; \bar{p}\text{-}a.s.$$

In terms of predictive distributions, Proposition 1 implies that predictive and true conditional distributions merge on events in $\sigma(\mathbf{h}^{\alpha})$, which can be thus regarded as the learnable events.

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

$$\left| p_{\mu\left(\cdot|\mathbf{h}_T^{\alpha}(\cdot)\right)}(B_t \mid \mathbf{h}_t^{\alpha}(\cdot)) - \bar{p}(B_t \mid \mathbf{h}_t^{\alpha}(\cdot)) \right| \to 0 \;\; \bar{p}^{\alpha}\text{-}a.s. \tag{7}$$

for each sequence of events (B_t) with range in $\sigma(\mathbf{h}^{\alpha})$.

For example, if $B_t = S^{t-1} \times \{(s_t, ..., s_{t+k})\} \times S^{\infty} \in \sigma(\mathbf{h}^{\alpha})$ for each t, then (7) becomes:

$$\left| p_{\mu\left(\cdot|\mathbf{h}_T^{\alpha}(\cdot)\right)}(s_t, ..., s_{t+k} \mid \mathbf{h}_t^{\alpha}(\cdot)) - \bar{p}(s_t, ..., s_{t+k} \mid \mathbf{h}_t^{\alpha}(\cdot)) \right| \to 0 \;\; \bar{p}^{\alpha}\text{-a.s.}$$

[17] Since $\hat{P}_T^{\alpha,\mu}(\bar{p})(\cdot) \subseteq \hat{P}_t^{\alpha,\mu}(\bar{p})(\cdot)$ ($\bar{p}$-a.s.) for all $t > T$, (6) implies that $\mu\left(\hat{P}_t^{\alpha,\mu}(\bar{p})(\cdot)\,|\mathbf{h}_t^{\alpha}\right) \to 1$ $\bar{p}^{\alpha}$-a.s. if the triple is consistent at any $t > T$.

That is, the predictive and the true conditional distributions of the future finite history $(s_t, ..., s_{t+k})$ merge, provided that such history is learnable (that is, that it belongs to $\sigma(\mathbf{h}^\alpha)$).

Finally, from (7) it follows that, for every $E \in \sigma(\mathbf{h}^\alpha)$, if we define $p(\cdot|\mathbf{h}^\alpha(\cdot)) := \lim_{t\to\infty} p(\cdot|\mathbf{h}^\alpha_t(\cdot))$,

$$p_{\mu\left(\cdot\middle|\mathbf{h}^\alpha_T(\cdot)\right)}(E \mid \mathbf{h}^\alpha(\cdot)) = \bar{p}(E \mid \mathbf{h}^\alpha(\cdot)) \in \{0,1\} \quad \bar{p}^\alpha\text{-a.s.}$$

That is, the DM asymptotically learns whether or not the events in $\sigma(\mathbf{h}^\alpha)$ obtain. Of course, the result is non-trivial only for events in the tail (e.g., E_i in the example below). Otherwise, if $E \in \sigma(\mathbf{h}^\alpha_k)$, i.e., $E = \bigcup_{s^{k-1}\in\bar{S}^{k-1}} \iota^\alpha_k(s^{k-1})$ for some $\bar{S}^{k-1} \subseteq S^{k-1}$, then, for each $t \geq k$, either $\iota^\alpha_t(s^{t-1}) \subseteq E$ or $\iota^\alpha_t(s^{t-1}) \cap E = \emptyset$, thus $p_{\mu\left(\cdot\middle|\mathbf{h}^\alpha_T(s^{T-1})\right)}(E \mid \mathbf{h}^\alpha_t(s^{t-1})) = \bar{p}(E \mid \mathbf{h}^\alpha_t(s^{t-1})) = 1$ for $t \geq k$ if $\iota^\alpha_t(s^{t-1}) \subseteq E$, and 0 if $\iota^\alpha_t(s^{t-1}) \cap E = \emptyset$.

Example 4 (Act II)**.** We consider i.i.d. models $p \in \Delta(S^\infty)$ parameterized by their marginals $\text{marg}_S\, p = \pi \in \Delta(S)$. Each of these models describes a possible composition of the urn. The prior $\mu \in \Delta(\Delta(S))$ is thus directly defined on the marginals.

Suppose that the DM:

1. knows that 1/3 of the balls are black (and so all his models π are such that $\pi(B) = 1/3$);

2. has a 3-point prior μ with $\text{supp}\,\mu = \{\pi^Y, \pi^{uni}, \pi^G\}$ and believes it is equally likely that the true model is either π^Y (with $\pi^Y(Y) = 2/3$), the uniform model π^{uni}, or π^G (with $\pi^G(G) = 2/3$):

Marginals	B	Y	G
π^Y	$\frac{1}{3}$	$\frac{2}{3}$	0
π^{uni}	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$
π^G	$\frac{1}{3}$	0	$\frac{2}{3}$

Prior	π^Y	π^{uni}	π^G
μ	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

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

$$\forall \pi \in \text{supp}\,\mu,\ \mu\left(\pi|\mathbf{h}^{\alpha^{NE}}_t(\cdot)\right) = \mu(\pi).$$

Here, $T(\alpha^{NE}, \mu, \bar{\pi}) = 1$ and $\hat{P}^{\alpha^{NE},\mu}_1(\bar{p}) = \text{supp}\,\mu$; strategy α^{NE} only allows partial identification.

For strategy α^E, if $\bar{\pi} \in \operatorname{supp} \mu$, $T(\alpha^E, \mu, \bar{\pi}) = 2$. To see why this is the case, note that $\operatorname{supp} \mu\left(\cdot|(y,1)\right) = \{\pi^Y, \pi^{\text{uni}}\}$, $\operatorname{supp} \mu\left(\cdot|(y,0)\right) = \{\pi^Y, \pi^{\text{uni}}, \pi^G\}$, and

$$\begin{aligned}
\hat{P}_2^{\alpha^E,\mu}\left(\bar{\pi}\right)\left(s_1\right) &= \begin{cases} \{\bar{\pi}\} & \text{if } \mathbf{h}_2^{\alpha^E}\left(s_1\right) = (y,1)\,, \\ \left\{\pi^Y, \pi^{uni}, \pi^G\right\} & \text{if } \mathbf{h}_2^{\alpha^E}\left(s_1\right) = (y,0)\,, \end{cases} \\
&= \begin{cases} \{\bar{\pi}\} & \text{if } s_1 = Y, \\ \left\{\pi^Y, \pi^{uni}, \pi^G\right\} & \text{else.} \end{cases}
\end{aligned}$$

We have that $p_{\pi^Y}^{\alpha^E} \perp p_{\pi^{\text{uni}}}^{\alpha^E}$ on $\sigma_{\geq 2}(\mathbf{h}^{\alpha^E})(Y)$, because betting on yellow forever reveals its objective probability as the long-run frequency of winning (2/3 under π^Y and 1/3 under π^{uni}). Formally, for $i \in \{1,2\}$, let:

$$E_i := \left\{ s^\infty \in S^\infty : \lim_{t\to\infty} \frac{1}{t} \sum_{k=2}^{t+1} \mathbf{1}_Y(s_k) = \frac{i}{3} \right\}.$$

We have $E_1, E_2 \in \sigma_{\geq 2}(\mathbf{h}^{\alpha^E})(Y)$, $p_{\pi^{\text{uni}}}^{\alpha^E}(E_1) = p_{\pi^Y}^{\alpha^E}(E_2) = 1$ and $p_{\pi^{\text{uni}}}^{\alpha^E}(E_2) = p_{\pi^Y}^{\alpha^E}(E_1) = 0$.

This establishes that the triple $\left(\alpha^E, \mu, \bar{\pi}\right)$ is minimally consistent at $T = 2$. By Proposition 1,

$$\mu\left(\cdot|\mathbf{h}_t^{\alpha^E}\right) \to \begin{cases} \delta_{\bar{\pi}} & \text{if } \mathbf{h}_2^{\alpha^E} = (y,1)\,, \\ \mu\left(\cdot \mid (y,0)\right) & \text{if } \mathbf{h}_2^{\alpha^E} = (y,0)\,. \end{cases}$$

If experimentation yields a success, the true model is asymptotically learned. Otherwise, if $h_2 = (y,0)$, posterior beliefs attain their limit value already in the second period and the DM remains in the dark. ▲

3.2 Value

At each time t there is a (time invariant) *instantaneous utility function* $u : C \to \mathbb{R}$, and an *instantaneous payoff function* $r = u \circ \rho$. If h_t is observed, the DM ranks strategy α given prior μ according to the present value, discounted by a factor $\delta \in [0,1)$, of the continuation stream of utility certainty equivalents:[18]

$$V\left(\alpha, \mu \mid h_t\right) := \sum_{\tau=t}^{\infty} \delta^{\tau-t} \phi^{-1}\left(\int_{\Delta(S^\infty)} \phi\left(\sum_{s^\tau \in S^\tau} r\left(\mathbf{a}_\tau^\alpha\left(s^{\tau-1}\right), s_\tau\right) p\left(s^\tau \mid h_t\right)\right) \mu\left(\mathrm{d}p \mid h_t\right)\right).$$

This criterion ranks at each point of time the current 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:

[18] Recall that $\mathbf{a}_t^\alpha\left(s^{t-1}\right) = \alpha_t\left(\mathbf{h}_t^\alpha\left(s^{t-1}\right)\right)$.

(i) $V(\alpha, \mu \mid h_t) = \sum_{\tau=t}^{\infty} \delta^{\tau-t} \sum_{s^\tau \in S^\tau} r\left(\mathbf{a}_\tau^\alpha\left(s^{\tau-1}\right), s_\tau\right) p_\mu\left(s^\tau \mid h_t\right)$ when ϕ is linear;

(ii) $V(\alpha, \mu \mid h_t) = \sum_{\tau=t}^{\infty} \delta^{\tau-t} \sum_{s^\tau \in S^\tau} r\left(\mathbf{a}_\tau^\alpha\left(s^{\tau-1}\right), s_\tau\right) p\left(s^\tau \mid h_t\right)$ when $\operatorname{supp} \mu = \{p\}$.

We remark that, except for the benchmark case of ambiguity neutrality, this 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 this inconsistencies in our setting.

Example 5 (Dynamic inconsistency)**.** We consider a modified version of our leading example. To ease calculations, we consider the two-periods truncated problem.[19] Assume also, as in Act I, that only bets on either black or yellow are possible, not on green. However, we assume that it is also possible to bet on black *and* to observe the color of the selected ball, action *bo*. Finally, we normalize payoffs as $u(0) = 0$ and $u(1) = 1$. With two outcomes, risk aversion is irrelevant and we can set $u(c) = c$, so that $r = u \circ \rho = \rho$, whereas f is described by the table ($*$ means "no direct observation of the color"):

f	B	Y	G
b	$1, *$	$0, *$	$0, *$
y	$0, *$	$1, *$	$0, *$
bo	$1, B$	$0, Y$	$0, G$

We consider i.i.d. models $p \in \Delta(S^\infty)$ parameterized by their marginals $\operatorname{marg}_S p = \pi \in \Delta(S)$. Each of these models describes a possible composition of the urn. The prior $\mu \in \Delta(\Delta(S))$ is thus directly defined on the marginals.

Suppose that the decision maker:

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

Summing up:

Marginals	B	Y	G
$\overline{\pi}^Y$	$\frac{1}{3}$	$\frac{5}{12}$	$\frac{1}{4}$
$\overline{\pi}^G$	$\frac{1}{3}$	$\frac{1}{4}$	$\frac{5}{12}$

Prior	$\overline{\pi}^Y$	$\overline{\pi}^G$
μ	$\frac{1}{2}$	$\frac{1}{2}$

[19] The two-periods problem can be easily framed into our infinite horizon setting. More precisely, it is obtained if all the models assign probability one to the same deterministic outcome after period 2, that is, all the uncertainty is resolved after the first two periods.

Let $\phi(x) = -e^{-10x}$; the *ex-ante* optimal strategy is:

α': "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."[20]

However, α' does not satisfy the one-shot deviation property. Indeed, after having observed yellow, the DM prefers to bet on black.

The ex-ante value of strategy α' is:

$$\begin{aligned} V\left(\alpha', \mu, \mid \left(a^0, m^0\right)\right) &= \sum_{\tau=t}^{2} \delta^{\tau-t} \phi^{-1} \left(\int_{\Delta(S)} \phi \left(\sum_{s^\tau \in S^\tau} r\left(\mathbf{a}_\tau^{\alpha'}\left(s^{\tau-1}\right), s_\tau\right) \pi\left(s^\tau \mid h_t\right) \right) \mu\left(\mathrm{d}\pi \mid \left(a^0, m^0\right)\right) \right) \\ &= \frac{1}{3} + \delta\phi^{-1} \begin{array}{c} \left(\mu\left(\overline{\pi}^Y \mid \left(a^0, m^0\right)\right) \phi\left(1 \cdot \overline{\pi}^Y (Y)^2 + 1 \cdot \left(1 - \overline{\pi}^Y (Y)\right) \left(\overline{\pi}^Y (B)\right)\right)\right. \\ \left. +\mu\left(\overline{\pi}^G \mid \left(a^0, m^0\right)\right) \phi 1 \cdot \overline{\pi}^G (Y)^2 + 1 \left(1 - \overline{\pi}^G (Y)\right) \left(\overline{\pi}^G (B)\right)\right) \end{array} \\ &= 0.\overline{3} + \delta 0.3364. \end{aligned}$$

On the other hand, the posterior belief after having chosen *bo* and having observed yellow, is:

$$\mu(\overline{\pi}^Y | (bo, Y)) = \mu(\overline{\pi}^Y | \left(a^0, m^0\right)) \frac{\overline{\pi}^Y (Y)}{\pi_\mu(Y)} = \frac{1}{2} \cdot \frac{5}{12} / \frac{1}{3} = \frac{5}{8},$$

that is:

Posterior	$\overline{\pi}^Y$	$\overline{\pi}^G$
$\mu(\cdot\|(bo, Y))$	$\frac{5}{8}$	$\frac{3}{8}$.

Hence, we obtain

$$\begin{aligned} & V\left(\alpha', \mu\left(\cdot|(bo, Y)\right) \mid (bo, Y)\right) \\ = \; & \phi^{-1}\left(\mu\left(\overline{\pi}^Y \mid (bo, Y)\right) \phi\left(1 \cdot \overline{\pi}^Y (Y)\right) + \mu\left(\overline{\pi}^G \mid (bo, Y)\right) \phi\left(1 \cdot \overline{\pi}^G (Y)\right)\right) \\ = \; & \phi^{-1}\left(\frac{5}{8}\phi\left(1 \cdot \frac{5}{12}\right) + \frac{3}{8}\phi\left(1 \cdot \frac{1}{4}\right)\right) \\ \cong \; & 0.3207 \\ < \; & \frac{1}{3} = V\left(\alpha', \mu\left(\cdot|(y, 1)\right) \mid (y, 1)\right). \end{aligned}$$

This is a typical example of dynamically inconsistent preferences. At period 0, the DM would like to commit to condition his behavior to 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 the true model is $\overline{\pi}^G$, Y will not be the first period draw. Therefore, even under model $\overline{\pi}^G$ this strategy presents a moderate expected value. However, after having observed (bo, Y), even if the posterior probability of $\overline{\pi}^G$ is lower,

[20] 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 are irrelevant in determining ex-ante optimality.

the DM considers the consequences of choosing action y too ambiguous. Indeed, the expected value under model $\overline{\pi}^G$, 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. ▲

If we make explicit the information histories that the process $(\mathbf{h}_t^\alpha)$ can actually reach, we have the $\sigma(\mathbf{h}_t^\alpha)$-measurable functions $V(\alpha, \mu \mid \mathbf{h}_t^\alpha(\cdot))$. In particular, we have the following:

(i) Under own-action independence of feedback about the state, $V(\alpha, \mu \mid \mathbf{h}_t^\alpha(\cdot)) = V(\alpha, \mu \mid \mathbf{h}_t(\cdot))$. In this case, neither the conditional nor the posterior probabilities depend on strategies. There is a separation between information and decision (see Witsenhausen 1971). This is a key feature that, in general, fails.

(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 $V(\alpha, \mu \mid \mathbf{h}_t^\alpha(\cdot)) = V(\alpha, \mu \,| \mathbf{s}^{t-1}(\cdot))$.

The dynamic structure $D = (A, S, M, \rho, f, u, \delta, \phi)$ enriches the static structure (2). The feedback function f and the discount factor δ are the genuine dynamic notions in the structure of the problem. In particular, we write $D = (\Gamma, \delta, f)$ to emphasize both ambiguity attitudes and feedback, our main objects of interest.

4 Self-confirming equilibrium

4.1 Steady-state analysis

We consider a stage decision problem Γ faced recurrently by a DM. He acts according to an overall strategy α, which at each time t prescribes some action a_t as a function of the information history $h_t = (a_1, m_1, ..., a_{t-1}, m_{t-1})$.

To introduce our main equilibrium concept, we need some notation. For any strategy α, 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 α otherwise.

Definition 2. *Triple* $(\alpha, \mu, \bar{p})$ *is a* self-confirming equilibrium *(SCE) if:*

(i) $\mu(p \in \Delta(S^\infty) : p^\alpha = \bar{p}^\alpha) = 1;$

(ii) For every action a, period t, and information history h_t,

$$p_\mu(I(h_t)) > 0 \Rightarrow V(\alpha, \mu \mid h_t) \geq V(\alpha/(h_t, a), \mu \mid h_t). \tag{8}$$

Condition (i) says that the DM only deems possible those models that coincide with the true model on events that are observable, at least in the long run, under the strategy played. In other words, his beliefs are concentrated on models that yield the same distribution of data, given the strategy, as the true model.

Condition (ii) is the one-shot deviation property which says that—for every information history h_t that the DM deems reachable with positive probability— action $\alpha_t(h_t)$ maximizes the continuation value conditional on h_t given that α will be followed in the future. The motivation is the following: Strategy α 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 the decision variable he controls maximizes his value given the 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 ϕ is positively affine—, the one-shot deviation principle implies that strategy α in an SCE $(\alpha, \mu, \bar{p})$ is subjectively optimal given μ.

Our definition of SCE is closely related to the notion of subjective equilibrium (Kalai and Lehrer 1995, henceforth KL). Besides minor details, there are two key differences. First, we consider arbitrary beliefs over probability models, while KL only consider Dirac beliefs over probability models. This is without loss of generality under the assumption of subjective expected utility maximization, because only predictive probability matter, hence a belief μ can be replaced by its predictive p_μ. Since we allow for non-neutral ambiguity attitudes, such simplification is precluded (see Sections 2.1 and 3.2). Second, the analysis of KL encompasses both strategic interaction and single-agent decision making, whereas we focus on the latter. Relatedly, unlike KL, we assume that the state process is exogenous, that is, the DM's actions cannot influence the probabilities of states in future periods, which is implausible in long-run interactions where the states are the co-players' stage-game choices.

However, there is a specific game theoretic framework that justifies our exogeneity assumption, the large population game. One way to interpret our setup is that it presents the point of view of a DM who plays recurrently a game with other agents independently drawn from large populations. The DM recognizes to be unable to influence the evolution of the environment with his actions. With this interpretation, the probability models describe the evolution of the distributions of actions in the co-players populations. In KL, instead, the set of interacting players is fixed once and for all. Finally, we emphasize that some results rely on the assumption of an i.i.d. environment that can be hardly reconciled with a situation of long-run interaction, but is instead consistent with a steady-state learning environment *à la* Fudenberg and Levine (1993).

4.2 I.i.d. environment

For some results, we will focus on the case where the real and the posited models are i.i.d. We denote by π in $\Delta(S)$ the (marginal) distribution of states of nature. Since the models are assumed to be i.i.d., this marginal uniquely pins down the model. More precisely, we define p_π in $\Delta(S^\infty)$ in the following way. For every t in $\mathbb{N}$, and for every s^t in S^t,

$$p_\pi(s^t \times S \times ...) = \prod_{\tau=1}^{t} \pi\left(s_\tau\right). \tag{9}$$

In this way, we have defined p_π over all the elementary cylinders. We call p_π the unique extension of this measure on $\mathcal{B}(S^\infty)$. Formally, beliefs are probability measures over the measurable space $(\Delta(S^\infty), \mathcal{B}(\Delta(S^\infty))$. However, since each model p_π is parametrized by its marginal π, we can directly consider beliefs as probability measures over marginal distributions, that is, as elements of $\Delta(\Delta(S))$. We use the letter ν to denote generic elements of $\Delta(\Delta(S))$. If the true model and the posited set of models are i.i.d., we say that the *environment is i.i.d.*: $\bar{p} = p_{\bar{\pi}}$ for some $\bar{\pi} \in \Delta\left(S\right)$ and $\operatorname{supp}\mu \subseteq \{p \in \Delta(S^\infty) : \exists \pi \in \Delta\left(S\right), p = p_\pi\}$. In the i.i.d. environment, we consider triples $(\alpha, \nu, \bar{\pi})$, where $\bar{\pi}$ is the one-period marginal of the correct model and ν is the probability measure over marginals induced by a prior μ in the natural way. In this context, with the term belief, we refer to the probability measure ν over marginals.

Standard results guarantee that it is without loss of generality to consider only stationary strategies in an i.i.d. environment. A stationary strategy α is a function $\alpha : \Delta(\Delta(S)) \to A$ that specifies actions as a function of the DM's (updated) beliefs. Formally, the strategy $\alpha : H_t \to A$ and the beliefs $\nu(\cdot|\cdot)$ are such that

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

In other words, the strategy must be measurable with respect to the beliefs. Note that this restriction is imposed on the pair of strategy and beliefs, not just on the strategy. Therefore, throughout this work, when the environment is assumed to be i.i.d., the strategy is implicitly assumed to be stationary.

Finally, a belief ν induces a *predictive marginal* and a *predictive distribution.* The former is defined as

$$\pi_\nu(s) = \int_{\Delta(S)} \pi\left(s\right)\nu\left(\mathrm{d}\pi\right),$$

whereas the latter is the corresponding p_{π_ν} defined as in (9).

In view of all this, we can adapt our value function to this i.i.d. environment, obtaining:

$$V\left(\alpha,\nu|h_t\right)=\sum_{\tau=t}^{\infty}\delta^{\tau-t}\phi^{-1}\left(\int_{\Delta(S)}\phi\left(\sum_{s^\tau\in S^\tau}r\left(\mathbf{a}_\tau^\alpha\left(s^{\tau-1}\right),s_\tau\right)p_\pi(s^\tau|h_t)\right)\nu\left(\mathrm{d}\pi\mid h_t\right)\right). \tag{10}$$

The following result shows that in an i.i.d setting, the value function depends on the history only through beliefs. Hence, without loss of generality, we can write $V\left(\alpha,\nu\right)$ to indicate the evaluation of a stationary strategy α under beliefs ν.

Lemma 2. *If $p_{\pi_\nu}(I\left(h_t\right))$ and $p_{\pi_\nu}(I\left(h'_{t'}\right))$ are strictly positive, then $\nu(\cdot\mid h_t)=\nu(\cdot\mid h'_{t'})$ implies $V\left(\alpha,\nu|h_t\right)=V\left(\alpha,\nu|h'_{t'}\right)$.*

We can also specialize the notion of SCE to the i.i.d. case.

Remark 1. The triple $(\alpha,\nu,\bar{\pi})$ is an SCE in an i.i.d. environment if:

- $\operatorname{supp}\nu\subseteq\{\pi\in\Delta(S):p_\pi^\alpha=p_{\bar{\pi}}^\alpha\}$;
- for every action a, period t, information history h_t, and ,

$$p_{\pi_\nu}(I(h_t))>0\Rightarrow V\left(\alpha,\nu(\cdot\mid h_t)\right)\geq V(\alpha/(h_t,a),\nu(\cdot\mid h_t)).$$

4.3 Learning dynamics

While 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 definition of pre self-confirming equilibrium.

Definition 3. *Triple $(\alpha,\mu,\bar{p})$ is a* pre self-confirming equilibrium *if, for every period t and information history h_t,*

(i) if $h_t\in\operatorname{Im}\mathbf{h}_t^\alpha$,

$$\bar{p}(I(h_t))>0\Rightarrow p_\mu(I(h_t))>0;$$

(ii) for every action a,

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

The difference between Definitions 2 and 3 lies in condition (i). Here, this first condition is weaker; we do not presume that beliefs have already "converged." However, beliefs must still be disciplined: The DM cannot be "surprised," in the sense that information sets $I(h_t)$ that have positive objective probability under strategy α (i.e., such that $\bar{p}(I(h_t)) > 0$ for $h_t \in \operatorname{Im} \mathbf{h}_t^{\alpha}$) have also positive subjective probability (i.e., they are such that $p_\mu(I(h_t)) > 0$). This is related to the absolute continuity condition of Kalai and Lehrer (1995).

Example 6 (Act III)**.** If we normalize payoffs as $u(0) = 0$ and $u(1) = 1$, we have $r = \rho = f$. Messages are thus the bets' payoffs. Moreover, we assume that $\phi(u) = -e^{-\lambda u}$, so that higher (absolute) ambiguity aversion corresponds to higher λ (see Klibanoff et al.). We maintain the i.i.d. hypothesis.

Suppose that the DM features the prior μ presented in Act II. We consider the strategies α^{NE} and α^{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:

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

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. While 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 3-point prior μ with $\operatorname{supp} \mu = \{\pi^Y, \pi^{uni}, \pi^G\}$ and believes it is equally likely that the true model is either π^Y, the uniform model π^{uni}, or π^G. If we let $\mu(\cdot|h_t) := (\mu(\pi^{uni}|h_t), \mu(\pi^Y|h_t), \mu(\pi^G|h_t))$ 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 α^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 μ, the continuation-value function after a success in period 1 is:

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

For the initial history, we have:

$$
\begin{aligned}
V(\alpha^E,\mu) :&= V(\alpha^E,\mu|\left(a^0,m^0\right)) \\
&= \phi^{-1}\left(\frac{1}{3}\phi(1/3)+\frac{1}{3}\phi(2/3)+\frac{1}{3}\phi(0)\right) \\
&+\frac{\delta}{1-\delta}\phi^{-1} \begin{array}{c} \left(\frac{1}{3}\phi\left(\pi^{uni}(Y)\cdot\frac{1}{3}+\pi^{uni}(G\cup B)\cdot\frac{1}{3}\right)+\right. \\ \left.\frac{1}{3}\phi\left(\pi^{Y}(Y)\cdot\frac{2}{3}+\pi^{Y}(G\cup B)\cdot\frac{1}{3}\right)+\frac{1}{3}\phi\left(\pi^{G}(Y)\cdot 0+\pi^{G}(G\cup B)\cdot\frac{1}{3}\right)\right) \end{array} \\
&= \phi^{-1}\left(\frac{1}{3}\phi(1/3)+\frac{1}{3}\phi(2/3)+\frac{1}{3}\phi(0)\right) \\
&+\frac{\delta}{1-\delta}\phi^{-1}\left(\left[\frac{1}{3}\phi\left(\frac{1}{9}+\frac{2}{9}\right)+\frac{1}{3}\phi\left(\frac{4}{9}+\frac{1}{9}\right)+\frac{1}{3}\phi\left(\frac{1}{3}\right)\right]\right) \\
&= \phi^{-1}\left(\frac{1}{3}\phi(1/3)+\frac{1}{3}\phi(2/3)+\frac{1}{3}\phi(0)\right)+\frac{\delta}{1-\delta}\phi^{-1}\left(\frac{1}{3}\phi\left(\frac{1}{3}\right)+\frac{1}{3}\phi\left(\frac{5}{9}\right)+\frac{1}{3}\phi\left(\frac{1}{3}\right)\right).
\end{aligned}
$$

Two forces affect the option value of experimentation: ambiguity aversion (the higher the value of λ, the lower the value of experimentation) and patience (the higher the value of δ, the higher the value of experimentation). In view of this, strategy α^{NE} is preferred if either $\delta = 0$ or λ is high enough given $\delta > 0$; if so, the triple $\left(\alpha^{NE},\mu,\overline{\pi}\right)$ is a pre self-confirming equilibrium for each $\overline{\pi} \in \left\{\pi^Y,\pi^{uni},\pi^G\right\}$. As for strategy α^E, if δ is sufficiently high and λ is sufficiently low, e.g., $\lambda = 1$ and $\delta = 0.39$, strategy α^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 π^G is false and updates his belief from $(1/3, 1/3, 1/3)$ to $(1/3, 2/3, 0)$. 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 $\lambda > 0$,

   $$
   \begin{aligned}
   V(\alpha^E,(1/3,2/3,0)) &= \frac{\phi^{-1}\left(\frac{1}{3}\phi(\pi^{uni}(Y))+\frac{2}{3}\phi(\pi^{Y}(Y))\right)}{1-\delta} \\
   &= \frac{\phi^{-1}\left(\frac{1}{3}\phi(1/3)+\frac{2}{3}\phi(2/3)\right)}{1-\delta} \\
   &> 1/3+\delta\frac{\phi^{-1}\left(\frac{1}{3}\phi(1/3)+\frac{2}{3}\phi(2/3)\right)}{1-\delta} = V(\alpha^E/b,(1/3,2/3,0)).
   \end{aligned}
   $$

   Moreover, at every subsequent period, the updating rule implies that the prior will be of the form $(1-k,k,0)$, with $k \in (0,1)$. It is easy to see that

   $$
   \begin{aligned}
   V(\alpha^E,(1-k,k,0)) &= \frac{\phi^{-1}\left(\left(1-k\right)\phi(1/3)+k\phi(2/3)\right)}{1-\delta} \\
   &> 1/3+\delta\frac{\phi^{-1}\left((1-k)\phi(1/3)+k\phi(2/3)\right)}{1-\delta} = V(\alpha^E/b,(1-k,k,0)).
   \end{aligned}
   $$

2. If experimentation is unsuccessful (i.e., $s_1 \in \{B, G\}$), the posterior of μ lowers the weight of model π^Y relative to models π^{uni} and π^G, so that $p_\mu(Y \mid (y,0)) < p_\mu(B \mid (y,0)) = 1/3$. Thereupon, strategy α^E recommends switching (and sticking) to black, so that the continuation value is the same as that under α^{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 δ, or for sufficiently high λ,

$$\begin{aligned}
& V\left(\alpha^E, \left(\frac{1}{3}, \frac{1}{6}, \frac{1}{2}\right)\right) \qquad (11) \\
= \;& \frac{1}{3}\frac{1}{1-\delta} \\
> \;& \phi^{-1}\left(\frac{1}{3}\phi(1/3) + \frac{1}{6}\phi(2/3) + \frac{1}{2}\phi(0)\right) + \frac{\delta}{1-\delta}\phi^{-1}\left(\frac{1}{3}\phi\left(\frac{1}{3}\right) + \frac{1}{6}\phi\left(\frac{5}{9}\right) + \frac{1}{2}\phi\left(\frac{1}{3}\right)\right) \\
= \;& \phi^{-1}\left(\frac{1}{3}\phi(1/3) + \frac{1}{6}\phi(2/3) + \frac{1}{2}\phi(0)\right) + \frac{\delta}{1-\delta}\phi^{-1}\left(\frac{1}{3}\phi\left(\frac{1}{3}\right) + \frac{1}{6}\phi\left(\frac{5}{9}\right) + \frac{1}{2}\phi\left(\frac{1}{3}\right)\right) \\
= \;& V\left(\alpha^E/y, \left(\frac{1}{3}, \frac{1}{6}, \frac{1}{2}\right)\right).
\end{aligned}$$

In particular, this inequality holds with $\lambda = 1$ and $\delta = 0.39$, and we have already proved that (α^E, μ) satisfies the one-shot deviation property at the root; therefore, $(\alpha^E, \mu, \overline{\pi})$ is a pre self-confirming equilibrium for each $\overline{\pi} \in \operatorname{supp} \mu$. ▲

This example suggests the following idea: *As ambiguity aversion increases, experimentation becomes less attractive.* Suppose for simplicity that the consequence and feedback functions coincide ($C = M$ and $\rho = f$) and that the utility function $v : C \to \mathbb{R}$ is injective. Then, to obtain evidence on the correct model, that is, to experiment, the DM has to choose an action that does not induce the same probability measure over payoffs under all the models he deems possible. This is exactly the kind of *ambiguous* choice that, other things being equal, an ambiguity averse DM avoids. On the other hand, if there is an action inducing the same probabilities of payoffs under all the models that the DM deems possible, high ambiguity aversion makes it attractive, but such action is not expected to be informative about the underlying probability model; indeed, the DM is certain that his next-period belief will be the same as the current belief if he chooses such "unambiguous" action.

The argument applied to the current-period expected payoffs can be extended to the value of experimentation: Experimentation at time t leads the DM to condition his behavior on the collected experimental evidence. In particular, he will choose an action with a large payoff under the models whose likelihood has been reinforced by the collected evidence. However, for this reason, the next periods' expected payoff will be model-dependent, that is, "ambiguous." Therefore, an increase in ambiguity aversion reduces also the value of experimentation. To sum up, there is a trade-off between choosing unambiguous actions and choosing informative actions. As we will see, this fact has a key implication for the long-run limit of the process.

4.4 Convergence to SCE

We are interested in studying the limit behavior of pre self-confirming equilibria. In particular, we investigate the conditions that imply convergence to self-confirming equilibria. To do this, we introduce the concept of ε-self-confirming equilibrium, which adapts the definition of ε-subjective equilibrium proposed in Kalai and Lehrer (1993) to a context of model uncertainty. The idea is that a triple $(\alpha, \mu, \bar{p})$ is an ε-self-confirming equilibrium if it satisfies two requirements. First, strategy α satisfies the one-shot deviation property given belief μ; second, belief μ assigns at least probability $1-\varepsilon$ to the set of models p that are observationally equivalent to the true data generating model $\bar{p}$. This second requirement is a weakening of condition (i) of SCE in Definition 2.

Definition 4. *The triple* $(\alpha, \mu, \bar{p})$ *is an* ε-self-confirming equilibrium *if:*

(i) $\mu\left(p \in \Delta(S^\infty) : p^\alpha = \bar{p}^\alpha\right) \geq 1 - \varepsilon$*;*

(ii) For every action a*, period* t*, and information history* h_t*,*

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

We study the evolution of actions and beliefs starting from a pre self-confirming equilibrium. Recall that a pre self-confirming equilibrium characterizes a sequentially rational DM who holds beliefs that do not assign probability 0 to observable events that can occur under the true model, that is, he cannot be completely surprised. Now, consider a pre self-confirming equilibrium $(\alpha, \mu, \bar{p})$ consistent from period T. Every history h_t, with positive probability under $\bar{p}$, induces a reinitialized triple $\left(\alpha_{h_t}, \mu_{h_t}, \bar{p}_{h_t}\right)$. More precisely, strategy α_{h_t} corresponds to the continuation strategy induced by α for the information histories that follow h_t, which corresponds to the new empty history. The reinitialized true model $\bar{p}_{h_t}$ is the restriction of $\bar{p}$ on the events that are consistent with h_t, and μ_{h_t} is the probability measure over similarly restricted models obtained from the posterior $\mu(\cdot|h_t)$.[21]

Next we show that, after a sufficiently long history, the reinitialized triple will $\bar{p}$-almost surely converge to an ε-self-confirming equilibrium. Given a path s^∞, we say that a triple $(\alpha, \mu, \bar{p})$ *converges to an* ε*-self-confirming equilibrium* on s^∞ if from a finite time t onward, the reinitialized triple $\left(\alpha_{\mathbf{h}^\alpha_{t+1}(s^t)}, \mu_{\mathbf{h}^\alpha_{t+1}(s^t)}, \bar{p}_{\mathbf{h}^\alpha_{t+1}(s^t)}\right)$ will form an ε-self-confirming.

[21] See the proof of Proposition 1 for a formal definition of these objects.

Proposition 2. *Let $(\alpha, \mu, \overline{p})$ be a pre self-confirming equilibrium. If $(\alpha, \mu, \overline{p})$ is consistent from some period $T \geq 1$, then, for every $\varepsilon > 0$, $(\alpha, \mu, \overline{p})$ converges $\bar{p}$-almost surely to an ε-self-confirming equilibrium.*

It is interesting to investigate the strength of our convergence result. Indeed, even if beliefs converge to a limit measure that assigns probability 1 to models that, given the adopted strategy, are observationally equivalent to the true one, the implications of this convergence in terms of predictive probabilities are not obvious. Therefore, in order to relate our ε-self-confirming equilibrium to the definition proposed by Kalai and Lehrer, we show that the predictive measure on observable events induced by the beliefs becomes ε-close to the objective one. Thus, we show that our first requirement for an ε-self-confirming equilibrium implies an analog of condition (c) of Kalai and Lehrer's definition.

Definition 5. *Let $\varepsilon > 0$ and let p, q be two probability measures defined on a measurable space (Ω, Σ). We say that p is ε-*close *to q if there exists $E \in \Sigma$ such that:*

(i) $p(E)$ and $q(E)$ are greater than $1 - \varepsilon$,

(ii) for every $E' \in \Sigma$ with $E' \subseteq E$,

$$|p(E') - q(E')| \leq \varepsilon q(E'). \tag{12}$$

The strength of this definition with respect to other definitions of ε-closeness (such as the one proposed by Blackwell and Dubins 1962), derives from the approximation restriction for small probability events. As underlined by Kalai and Lehrer (1993) in a repeated-game framework "being correct in small probability events is important since even significant events may have small probability if they occur late in the game." Given a path s^∞, we say that *beliefs becomes ε-close to the true model* on s^∞ if, from a finite period t onward, the predictive measures becomes ε-close to the true model.

Proposition 3. *Let $(\alpha, \mu, \overline{p})$ be a pre self-confirming equilibrium. If $(\alpha, \mu, \overline{p})$ is consistent from some period $T \geq 1$, then, for every $\varepsilon > 0$, predictive beliefs become $\bar{p}$-almost surely ε-close to the true model.*

In words, the predictive probabilities induced by beliefs and the true model almost surely merge on the observable events. Note that this result extend the "local" result of Corollary 2 to a "global" one. On the one hand, Corollary 2 states that, given a specific sequence of events, the predictive and the true conditional distributions of these events merge, provided they are learnable. On the other hand, this proposition shows that the predictive and the true conditional distributions merge "globally" on observable events.

The first condition of the Definition 4 ensures that the probability assigned to the set of models observationally equivalent to $\overline{p}$ converges to 1. Now we show that this implies that the DM understands the payoff relevant implications of the adopted strategy. More precisely, we show that its subjective value converges to the objective one.

Proposition 4. *Let $(\alpha, \mu, \overline{p})$ be a pre self-confirming equilibrium. If $(\alpha, \mu, \overline{p})$ is consistent from some period $T \geq 1$, then, $\bar{p}$-almost surely:*

$$\lim_{t \to \infty} \left| V\left(\alpha, \mu | \mathbf{h}_t^\alpha\right) - V\left(\alpha, \delta_{\bar{p}} | \mathbf{h}_t^\alpha\right) \right| = 0.$$

4.5 Convergence to static SCE

We can introduce in our framework the counterpart of the SCE notion of BCMM, which we call "static SCE." The key feature of the equilibrium concept of BCMM is that the chosen action is a myopic best reply to confirmed beliefs; therefore, we consider the following definition:[22]

Definition 6. *A triple $(a^*, \nu^*, \bar{\pi}) \in A \times \Delta(\Delta(S)) \times \Delta(S)$ of actions, beliefs, and models is a* static SCE *if*

(i) $\nu^* \left(\pi \in \Delta(S) : \pi \circ f_{a^*}^{-1} = \bar{\pi} \circ f_{a^*}^{-1}\right) = 1$*;*

(ii) $a^* \in \arg\max_{a \in A} \phi^{-1} \left(\int_{\Delta(S)} \phi \left(\sum_{s \in S} r(a, s) \pi(s) \right) \nu^*(\mathrm{d}\pi) \right)$.

The second condition says that a^* is a (myopic, or one-period) best response to ν^* given the ambiguity attitudes determined by ϕ. The first condition is the self-confirming property adapted to the static framework: the distribution of messages that the DM "observes" in the long run if he always plays a^* is exactly what he expects. Since payoffs are observable, the self-confirming property implies that a^* is *unambiguous* for ν^* and the expected distribution of payoffs coincides with the one implied by the true model $\bar{\pi}$. Indeed, by (3),

$$\forall \pi \in \operatorname{supp} \nu,\ \pi \circ \rho_{a^*}^{-1} = \pi \circ f_{a^*}^{-1} \circ \gamma_{a^*}^{-1} = \bar{\pi} \circ f_{a^*}^{-1} \circ \gamma_{a^*}^{-1} = \bar{\pi} \circ \rho_{a^*}^{-1}.$$

Does our result of convergence of pre-SCE to SCE imply convergence to a static SCE? It is clear that condition (i) of SCE (Definition 3) implies condition (i) of Definition 6. Moreover, at an SCE in an i.i.d. environment, since beliefs are confirmed and the strategy is stationary, a unique action $\alpha\left(\nu^*\right)$ is played. However, while the definition of static SCE (a^*, ν^*, π^*) requires a^* to be the myopic best reply to ν, in an SCE, strategy α is required to satisfy the one-shot deviation property. The following result sheds light on the relation between SCE and static SCE.

[22] We adopt the standard "pushforward" notation: given $\pi \in \Delta(S)$ and $f_a : S \to M$, the induced measure on M is $\pi \circ f_a^{-1}$, where $\left(\pi \circ f_a^{-1}\right)(m) = \pi\left(f_a^{-1}(m)\right)$ for each m.

Proposition 5. *Fix an i.i.d. environment and let the triple $(\alpha, \nu, \bar{\pi})$ be an SCE. Then the triple $(\alpha(\nu), \nu, \bar{\pi})$ is a static SCE.*

Then, we show that, under our assumption, beliefs converge almost surely to a random limit $\nu^{\alpha}_{\mathbf{s}^{\infty}}$

Lemma 3. *Let $(\alpha, \nu, \bar{\pi})$ be a pre self-confirming equilibrium in an i.i.d. environment. If $(\alpha, \nu, \bar{\pi})$ is consistent from some period $T \geq 1$, then beliefs converge almost surely to a random limit $\nu^{\alpha}_{\mathbf{s}^{\infty}}$.*

We remark that, in an i.i.d. environment, consistency simply requires that the DM assigns positive probability to the set of model observationally equivalent—under the adopted strategy—to the true model $\bar{\pi}$.[23]

Finally, we can combine Lemma 3 and Proposition 5 to provide a learning foundation to the concept proposed by BCMM. Given a path s^{∞}, we say that a triple $(\alpha, \nu, \bar{\pi})$ *converges to a static SCE* on s^{∞} if from a finite time t onward, $(\mathbf{a}^{\alpha}_{t}(s^{t-1}), \nu^{\alpha}_{s^{\infty}}, \bar{\pi})$ forms a static SCE. Note that the tail sequence of actions $(\mathbf{a}^{\alpha}_{\tau}(s^{\tau-1}))_{\tau \geq t}$ is not required to be constant.

Proposition 6. *Let $(\alpha, \nu, \bar{\pi})$ be a pre self-confirming equilibrium in an i.i.d. environment. If $(\alpha, \nu, \bar{\pi})$ is consistent from some period $T \geq 1$, $(\alpha, \nu, \bar{\pi})$ converges $\bar{p}$-almost surely to a static SCE.*

The intuition is as follows. Under the stated assumptions, beliefs converge almost surely to a random limit $\nu^{\alpha}_{\mathbf{s}^{\infty}}$. Since the action set A is finite, after a random time $\widehat{T}_{\mathbf{s}^{\infty}}$, each action chosen by α is played infinitely often and must be a best reply to the limit belief $\nu^{\alpha}_{\mathbf{s}^{\infty}}$, because it is (asymptotically) a best reply to beliefs arbitrarily close to $\nu^{\alpha}_{\mathbf{s}^{\infty}}$. Since $\nu^{\alpha}_{\mathbf{s}^{\infty}}$ assigns probability 1 to the set of models that are α-observationally equivalent to $\bar{\pi}$, all such actions must yield, with $(\nu^{\alpha}_{\mathbf{s}^{\infty}}, \bar{\pi})$, a static SCE. Note that the realized sequence of actions $(\mathbf{a}^{\alpha}_{t}(s^{t-1}))$ converges if there is a unique myopic best reply to the limit belief $\nu^{\alpha}_{s^{\infty}}$, but such uniqueness is not guaranteed. Yet, if the myopic best reply is indeed unique, say action a^{*}, the action sequence is eventually constant at $a^{*}$and $(a^{*}, \nu^{\alpha}_{s^{\infty}}, \bar{\pi})$ is a static SCE. Moreover, after a finite time, the agent chooses an action that maximizes one-period value with respect to current beliefs (and not only limit ones), that is, exploration (experimentation) becomes irrelevant, all that matter is one-period exploitation.

Corollary 3. *Let $(\alpha, \nu, \bar{\pi})$ be a pre self-confirming equilibrium in an i.i.d. environment that converges to a static SCE on path s^{∞}. If*

$$\arg\max_{a \in A} \phi^{-1}\left(\int_{\Delta(S)} \phi\left(\sum_{s \in S} r(a,s)\pi(s)\right) \nu^{\alpha}_{s^{\infty}}(\mathrm{d}\pi)\right) = \{a^{*}_{s^{\infty}}\}$$

[23] This is intuitive, but perhaps not obvious. The proof is available upon request.

for some $a^*_{s^\infty} \in A$, then there exist $\widehat{T}_{s^\infty}$ such that, for every $t > \widehat{T}_{s^\infty}$,

$$\mathbf{a}^\alpha_t(s^{t-1}) = a^*_{s^\infty}.$$

Moreover, there exists $\bar{T}_{s^\infty}$ such that

$$t > \bar{T}_{s^\infty} \Rightarrow a^*_{s^\infty} \in \arg\max_{a\in A} \phi^{-1}\left(\int_{\Delta(S)} \phi\left(\sum_{s\in S} r(a,s)\pi(s)\right)\nu(\mathrm{d}\pi|\mathbf{h}^\alpha_t\left(s^{t-1}\right))\right).$$

It is important to stress that this convergence does *not* imply that the limit belief is the Dirac measure supported by the correct model. However, the limit pairs of beliefs and actions almost surely satisfies the usual properties of stochastic limits in the (expected utility) stochastic control limit literature. Indeed, the realization $(a^*_{s^\infty}, \nu^\alpha_{s^\infty})$ is such that we have:

- *(confirmed beliefs)* $\nu^\alpha_{s^\infty}$ assigns probability 1 to the models that are observationally equivalent given $a^*_{s^\infty}$, (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 on exploration.

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

Example 7 (Act IV)**.** Consider the strategy α^E of the previous acts. Again, recall by Act II that the DM has a 3-point prior μ with $\operatorname{supp}\mu = \left\{\pi^Y, \pi^{uni}, \pi^G\right\}$ and believes it is equally likely that the true model is either π^Y, the uniform model π^{uni}, or π^G. In Act II, we have shown that $(\alpha^E, \mu, \overline{\pi})$ is consistent from period 2, whereas in Act III, we have proved that with parameters $\lambda = 1$ and $\delta = 0.39$ it is a pre self-confirming equilibrium. We can show how our convergence result obtains in this simple specific case. Suppose that $\overline{\pi} = \pi^Y$. From Proposition 3, we have that the limit beliefs attained are the following:

$$\mu^{\alpha^E}_{s^\infty} = \left(\mu^{\alpha^E}_{s^\infty}(\pi^{uni}), \mu^{\alpha^E}_{s^\infty}(\pi^Y), \mu^{\alpha^E}_{s^\infty}(\pi^G)\right) = \begin{cases} (1/3, 1/6, 1/2) & \text{if } s_1 \in \{B, G\}, \\ (0, 1, 0) & \text{if } s_1 = Y. \end{cases}$$

Indeed, if the experimentation is unsuccessful, the posterior of μ lowers the weight of model π^Y relative to models π^{uni} and π^G, thereupon, strategy α^E recommends switching (and sticking) to black, and so there is no additional updating. On the other hand, if the experimentation is successful, strategy α^E prescribes to stick on yellow thereupon, and then the correct model π^Y is asymptotically identified.

Since we are in an i.i.d. context, Proposition 6 holds. In particular, if $s_1 \in \{B, G\}$, by (11), for every $t > 1$,

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

for every $t > 1$, and $(b, (1/3, 1/6, 1/2), \pi^Y)$ is the static SCE that obtains in the limit. Note that in this case the DM will end up choosing an *objectively sub-optimal* action.

If $s_1 = Y$,

$$\mathbf{a}_t^{\alpha^E}(s^{t-1}) = y,$$

for every $t > 1$, and $(y, (0, 1, 0), \pi^Y)$ is the static SCE that obtains in the limit. Indeed, it is immediate to see that these actions maximize one-period value with respect to limit beliefs and that the distribution of probability over messages confirms them.

Finally, consider strategy α^{NE}. In Act III, we have argued that it is a pre self-confirming equilibrium if the DM is sufficiently ambiguity averse. In this case, regardless of the correct marginal $\bar{\pi} \in P = \left\{\pi^{\mathrm{Y}}, \pi^{\mathrm{uni}}, \pi^{\mathrm{G}}\right\}$,we have almost sure convergence to a static SCE from period 1. Indeed, the DM sticks on black from the first period, and black is the myopic best reply to the confirmed prior $\mu = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$. However, note that if the correct model is π^{Y}*,* betting on black is *objectively sub-optimal.*

As argued earlier, our leading example suggests the idea that ambiguity aversion tends to stifle experimentation. Therefore, the DM is more likely to end up not discovering the correct model. If instead the true model $\bar{\pi}$ is identified in the limit, by Proposition 6, eventually the DM chooses the one-period best reply to $\bar{\pi}$, which is the counterpart of Nash Equilibrium in our framework*.*

This suggests the following conjecture: As ambiguity aversion increases, the DM reduces experimentation and he is more likely to converge to a non-Nash SCE. However, we can show by example that this conjecture is incorrect.[24]

Despite this *caveat*, our leading example and the previous results cast a new light on the relation between ambiguity attitudes and self-confirming equilibrium. In the example, the set the set of static SCE of the game is invariant with respect to ambiguity attitudes, a property that extends to a large class of situations (see, e.g., Battigalli et al. 2016a, and Battigalli et al. 2016b). Yet, even though ambiguity attitudes do not affect the set of long-run outcomes, they may have a "dynamic" effect: Suppose that the true model in the example is π^Y; then, for moderate ambiguity aversion the process converges to the Nash equilibrium with positive probability, with high ambiguity aversion the process is stuck in a non-Nash SCE.

[24] Similarly, a lower discount factor does not necessarily make convergence to a non-Nash SCE more likely. Counterexamples are available upon request. We remark that, in the example about comparative ambiguity aversion, only payoffs are observable; this implies the coincidence between actions that allow for learning and ambiguous actions.

4.6 Characterization of SCE beliefs

In what follows, we borrow from Easley and Kiefer (1988) and derive a necessary property of SCE. In view of our Proposition 6, we can also interpret it as a long-run result.

Let

$$V_1(\nu) = \max_{a \in A} \phi^{-1} \left(\int_{\Delta(S)} \phi \left(\sum_{s \in S} r(a,s)\pi(s) \right) \nu(\mathrm{d}\pi) \right).$$

That is, $V_1(\nu)$ is the value of the one-period truncated problem.

Proposition 7. *Let $(\alpha, \nu^*, \overline{\pi})$ be an SCE in an i.i.d. environment. Then*

$$\nu^* \in \arg \min_{\nu : \operatorname{supp} \nu \subseteq \operatorname{supp} \nu^*} V_1(\nu).$$

The idea behind this proposition is the following: For every model $\pi \in \operatorname{supp} \nu^*$ the self-confirming action $\alpha(\nu^*)$ yields the same objective expected payoff as under the true model $\bar{\pi}$. Therefore, the same holds for beliefs ν with a smaller support. Then the self-confirming action has the same one-period value under ν and ν^*. Since this action is feasible, but may be (subjectively) sub-optimal under belief ν, the maximal one-period value under ν must be at least as high as the true objective expected payoff, which is the maximal one-period value under ν^*. This result is particularly useful because the one-period value function is easy to calculate, and it sheds light on the beliefs that can support an SCE.

Proposition 7 formalizes the intuition that information is valuable: if sharper information is available, that is, a smaller set of models are deemed possible, than the value of the problem is higher.

5 Discussion: stochastic control problems

5.1 Two frameworks

In this section we connect our setting with the one used by the literature on active learning in stochastic control problems, as exemplified by the well known work of Easley and Kiefer (1988). To this end, we relate the SCE concept for decision problems with feedback with the limit behavior of solutions to stochastic control problems. To ease matters, we consider (classical subjective) expected utility—i.e., ambiguity neutrality—and we assume that all the relevant sets are *finite*.

Feedback framework Consider a static decision problem with feedback and observable payoffs

$$(A, S, C, M, P, u, f, \gamma) \tag{13}$$

where $\gamma : A \times M \to C$ is the function introduced in (3). As a result, the objective expected utility of a given probability model p is $R(a,p) = \sum_{s \in S} u\left(\gamma\left(a, f(a,s)\right)\right) p\left(s\right)$. Let

$$R(a,\mu) := \sum_{p \in P} R(a,p)\mu\left(p\right) \tag{14}$$

In view of Definition 6, an SCE is a triple of actions, beliefs, and models $(a^*, \mu^*, \bar{p})$ such that:

(i) $\mu^*\left(p \in P : p \circ f_{a^*}^{-1} = \bar{p} \circ f_{a^*}^{-1}\right) = 1$,

(ii) $a^* \in \arg\max_{a \in A} R(a, \mu^*)$.

An SCE given $\bar{p} \in P$ is a pair (a^*, μ^*) such that $(a^*, \mu^*, \bar{p})$ is an SCE.

Consider the infinite repetition of the static decision problem, with i.i.d. models with unknown marginal distribution $\bar{p}$. Stationary strategies $\alpha : \Delta\left(P\right) \to A$ (Section 4.2) induce, via Bayes rule, a sequence $(\mathbf{a}_t^\alpha, \boldsymbol{\mu}_t^\alpha)$ of random actions and beliefs, with $\mathbf{a}_t^\alpha = \alpha\left(\boldsymbol{\mu}_t^\alpha\right)$ and $\boldsymbol{\mu}_{t+1}\left(p\right) = \left(p \circ f_{\mathbf{a}_t^\alpha}^{-1}\right)\left(\mathbf{m}_t^\alpha\right)\boldsymbol{\mu}_t\left(p\right) / \sum_{p \in P}\left(p \circ f_{\mathbf{a}_t^\alpha}^{-1}\right)\left(\mathbf{m}_t^\alpha\right)\boldsymbol{\mu}_t\left(p\right)$.[25] Given a discount factor $\delta \in [0,1)$ and prior belief μ_0, the DM then solves $\max_\alpha \mathbb{E}_{\mu_0} \sum_{t=1}^\infty \delta^t R\left(\mathbf{a}_t^\alpha, \boldsymbol{\mu}_t^\alpha\right)$ over the set of stationary strategies.

Stochastic control framework Easley and Kiefer [9, 1988] (henceforth EK) analyze the following problem of discrete-time stochastic control. We use our own notation and terminology, but we report theirs in brackets to ease the comparison:

- $a \in A$, *actions* [EK: $x \in X$];
- $m \in M$, *messages* [EK: *observations, or outcomes* $y \in Y$];
- $r : A \times M \to \mathbb{R}$, *payoff function* [EK: $r : X \times Y \to \mathbb{R}$];
- $\theta \in \Theta$, *parameters*;
- $\varphi(\cdot|\cdot,\cdot) : M \times A \times \Theta \to \mathbb{R}$, *conditional density*, a function $(a,\theta) \mapsto \varphi\left(\cdot|a,\theta\right) \in \Delta\left(M\right)$ in the finite case [EK: $f(\cdot|\cdot,\cdot) : Y \times X \times \Theta \to \mathbb{R}$];[26]
- *prior/posterior beliefs* $\nu \in \Delta\left(\Theta\right)$, with ν_0 being the *prior*, that is, the initial belief [EK: $\mu \in \Delta\left(\Theta\right)$];

[25] Throughout the section we adopt the convention $0/0 = 0$.

[26] Densities require some reference measure over Y. For instance, when $Y \subseteq \mathbb{R}^n$ is full dimensional, the reference measure is the Lebesgue measure; when Y is finite, the reference measure is the uniform measure.

In particular, the current (one-period) expected reward of action a given belief ν is

$$\tilde{R}(a,\nu) = \sum_{\theta\in\Theta}\left(\sum_{m\in M} r(a,m)\varphi(m|a,\theta)\right)\nu(\theta).$$

where the tilde distinguishes this reward from (14).

This setting can be summarized by the sextuple

$$(A, M, r, \Theta, \nu_0, \varphi). \tag{15}$$

In an infinite repetition of this static decision problem, with i.i.d. models, the state space is $\Delta(\Theta)$ and the law of motion is Bayes rule. Let α and $(\mathbf{a}_t^\alpha, \boldsymbol{\nu}_t^\alpha)$ denote, respectively, the stationary strategy of the DM and the implied sequence of random actions and Bayesian posteriors, with $\mathbf{a}_t^\alpha = \alpha(\boldsymbol{\nu}_t^\alpha)$ and

$$\nu_{t+1}(\theta|a,m) = \frac{\varphi(m|a,\theta)\,\nu_t(\theta)}{\sum_{\theta'\in\Theta}\varphi(m|a,\theta')\,\nu_t(\theta')},$$

where ν_t is a realization of $\boldsymbol{\nu}_t^\alpha$. Given a discount factor $\delta\in[0,1)$ and prior belief ν_0, the DM then solves $\max_\alpha \mathbb{E}_{\nu_0}\sum_{t=0}^{\infty}\delta^t\tilde{R}(\mathbf{a}_t^\alpha, \boldsymbol{\nu}_t^\alpha)$ with respect to his stationary strategies.

5.2 Unification of the two frameworks

Let us unify the two previous frameworks, noting that A, M, and the discount factor δ are common to both.

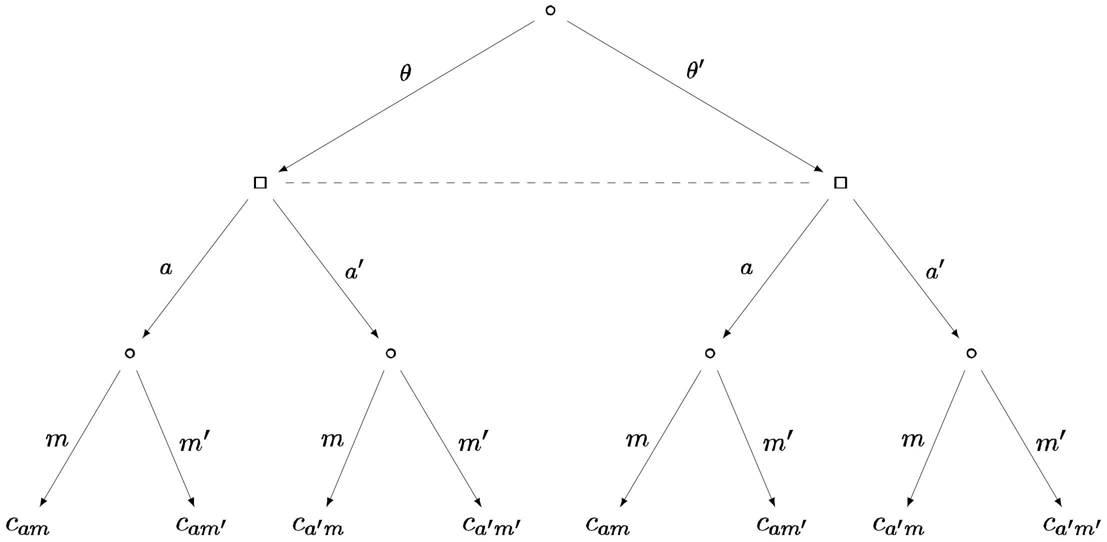
From feedback to the stochastic control framework Given a decision problem with feedback and observable payoffs $(A, S, C, M, P, u, f, \gamma)$ and a prior μ_0, the basic elements of a stochastic control problem $(A, M, r, \Theta, \nu_0, \varphi)$ are derived as follows:

- $r = u\circ\gamma$;
- $\Theta = P \subseteq \Delta(S)$ and $\nu_0 = \mu_0$;
- $\varphi(\cdot|a,p) = p\circ f_a^{-1}\in\Delta(M)$ for all $(a,p)\in A\times P = A\times\Theta$.

The stochastic control problem can be thus seen as a reduced form of a decision problem with feedback and observable payoffs.

Vice versa Now we fix the elements $(A, M, r, \Theta, \nu_0, \varphi)$ of a stochastic control problem, and derive $(A, S, C, M, P, u, f, \gamma)$ and the prior μ_0 as follows:

- *States of nature:* $S = \Theta \times M^{A\times\Theta}$. States are thus the pairs $(\theta, \eta) \in \Theta \times M^{A\times\Theta}$, that is, the "strategies of nature," where nature chooses θ before a (and the DM does not observes θ) or simultaneously, and in a second stage of the same period nature chooses the outcome m as a function of (a, θ). The following extensive game form of a (binary) decision problem illustrates this construction:



This is the most important connection between the two frameworks.

- *Stochastic models:* The issue is how to define the set P of possible stochastic models given the function $(a, \theta) \mapsto \varphi(\cdot|a, \theta) \in \Delta(M)$. This map defines a "second-stage behavioral strategy" of nature, and P can be defined as the set of "mixed strategies of nature" that pick a particular $\bar{\theta}$ with probability 1 and are consistent with such second-stage behavioral strategy:

$$\left\{ p \in \Delta\left(\Theta \times M^{A\times\Theta}\right) : \begin{array}{l} \exists\theta^* \in \Theta, \forall\,(a, m) \in A \times M, \mathrm{marg}_\Theta\, p = \delta_{\theta^*}, \\ p\left((\theta, \eta) \in S : \theta = \theta^*, \eta(a, \theta) = m\right) = \varphi\left(m|a, \theta\right)\right) \end{array} \right\}$$

However, one would like to parametrize P so that it is isomorphic to Θ, because all that matters for the DM are the objective probabilities

$$p\left((\theta, \eta) \in S : \theta = \theta^*, \eta(a, \theta) = m\right) = \varphi\left(m|a, \theta\right)$$

and the true model θ^*. Because of the finiteness of all the sets, P can be fully parametrized by θ by restricting the set of probability models in the same way as Kuhn (1953) goes from behavioral strategies to mixed strategies.[27] Let $\Theta \subseteq$

[27] See also Selten [29, 1975].

$\Delta(\Theta)$ be the canonical embedding of Θ into $\Delta(\Theta)$, i.e., $\Theta = \{\delta_\theta\}_{\theta\in\Theta}$. Given $\varphi \in \Delta(M)^{A\times\Theta}$, define the following measure on $M^{A\times\Theta}$: for all $\eta \in M^{A\times\Theta}$,

$$p^{\varphi}_{M^{A\times\Theta}}(\eta) = \prod_{(a,\theta)\in A\times\Theta} \varphi\left(\eta\left(a,\theta\right)|a,\theta\right).$$

That is, $p^{q}_{M^{A\times\Theta}}(\eta)$ is the probability of the profile $(\eta(a,\theta))_{(a,\theta)\in A\times\Theta}$ of contingent choices by nature under the assumption that such choices are independent across nodes. Then, write the set of probability models as follows:

$$P = \left\{p \in \Delta\left(\Theta \times M^{A\times\Theta}\right) : \exists\theta\in\Theta, p = \delta_\theta \times p^{\varphi}_{M^{A\times\Theta}}\right\},$$

where $\delta_\theta \times p^{q}_{M^{A\times\Theta}}$ is the product measure obtained from δ_θ and $p^{q}_{M^{A\times\Theta}}$. This yields the bijection $\varsigma : \Theta \to P$ given by:

$$\varsigma(\theta) = \delta_\theta \times p^{\varphi}_{M^{A\times\Theta}}.$$

This is the second most important connection between the two frameworks. The usual realization-equivalence argument *à la* Kuhn (1953) shows that, for each $(a,\theta,m) \in A\times\Theta\times M$,

$$\mathbb{P}_{\delta_a,\varsigma(\theta)}\left(a,\theta,m\right) = \varphi\left(m|a,\theta\right),$$

where, in general, for every "mixed strategy pair" $(\alpha,p) \in \Delta(A)\times\Delta\left(\Theta\times M^{A\times\Theta}\right)$,

$$\mathbb{P}_{\alpha,p}\left(a,\theta,m\right) = \alpha(a)p\left((\theta',\eta')\in\Theta\times M^{A\times\Theta} : \theta'=\theta, \eta'(a,\theta)=m\right)$$

denotes the induced probability of (a,θ,m).

- *Prior:* $\mu_0 = \nu_0 \circ \varsigma^{-1}$.

- *Consequences, consequence function, and utility:* $C = \operatorname{Im} r \subseteq \mathbb{R}$, $u = \mathrm{Id}_C$, and $\gamma = r$. Note that the stochastic control problem does not specify a consequence space and a consequence function because of its reduced-form nature. Therefore, the specification of C and u has to be somewhat arbitrary. The specification above is natural when there are monetary consequences and risk neutrality. An alternative and equally salient specification is $C = A\times M$ and $u = r$.

- *Feedback:* $f\left(a,(\theta,\eta)\right) = \eta(a,\theta)$ for all $(a,(\theta,\eta)) \in A\times S = A\times\Theta\times M^{A\times\Theta}$.

5.3 Convergence to self-confirming equilibria

Fix a decision problem with feedback and observable payoffs (A,S,C,M,P,u,f,γ), a prior $\mu_0 \in \Delta(P)$, and some discount factor $\delta \in [0,1)$. In this section, we have assumed so far that all sets are finite. Now assume that:

(i) the sets S and P are finite,

(ii) the sets A and M are, possibly infinite, subsets of an Euclidean space.

(iii) the optimum problem $\max_{a \in A} R(a, \mu)$ has, for each $\mu \in \Delta(P)$, a unique solution, denoted $a^*(\mu)$.

The analysis of the previous part of this section still holds, *mutatis mutandis*, under the weaker cardinality assumptions (i) and (ii). Let $p^* \in \Delta(P)$ denote the (unknown) true distribution of states. Then the associated stochastic control problem defined in Section 5.2 satisfies all the assumptions of Easley and Kiefer (1988). In particular, they give conditions on the stochastic control problem so that, given the prior ν_0 and the true parameter θ_0, for every optimal stationary strategy α^* the induced stochastic process of actions and posterior beliefs $(\mathbf{a}_t^*, \boldsymbol{\nu}_t^*)$ converges, θ_0-almost surely,[28] to a random limit $(\mathbf{a}_\infty^*, \boldsymbol{\nu}_\infty^*)$ such that, θ_0-almost surely,

$$\mathbf{a}_\infty^* = a^*(\boldsymbol{\nu}_\infty^*)$$

and

$$\operatorname{supp} \boldsymbol{\nu}_\infty^* \subseteq \{\nu^* \in \Delta(\Theta) : \nu^*(\{\theta : \varphi(\cdot|a^*(\nu^*), \theta) = \varphi(\cdot|a^*(\nu^*), \theta_0)\}) = 1\}.$$

In terms of our feedback setting, this means that for every stationary expected utility maximizing strategy α^* of the repeated decision problem (given the prior μ_0), the induced stochastic process $(\mathbf{a}_t^*, \boldsymbol{\mu}_t^*)$ of actions and posterior beliefs converges p^*-almost surely to a random pair $(\mathbf{a}_\infty^*, \boldsymbol{\mu}_\infty^*)$ such that, p^*-almost surely,

$$\mathbf{a}_\infty^* = a^*(\boldsymbol{\mu}_\infty^*)$$

and

$$\operatorname{supp} \boldsymbol{\mu}_\infty^* \subseteq \left\{\mu^* \in \Delta(\Theta) : \mu^*\left(\left\{p : p \circ f_{a^*(\mu^*)}^{-1} = p^* \circ f_{a^*(\mu^*)}^{-1}\right\}\right) = 1\right\}.$$

This can be verified by bookkeeping. In particular, given that $\Theta = P$, $\theta_0 = p^*$, $\varphi(\cdot|a, \varsigma(p)) = p \circ f_a^{-1}$ for all $(a, p) \in A \times P$, and $\nu^* = \mu^*$, the condition

$$\nu^*(\theta : \varphi(\cdot|a^*(\nu^*), \theta) = \varphi(\cdot|a^*(\nu^*), \theta_0)) = 1$$

becomes $\mu^*(p : p \circ f_{a^*(\mu^*)}^{-1} = p^* \circ f_{a^*(\mu^*)}^{-1}) = 1$. But then, for each pair (a^*, μ^*) in the support of the random limit, we have:

(i) (confirmed belief) $\mu^*(p : p \circ f_{a^*(\mu^*)}^{-1} = p^* \circ f_{a^*(\mu^*)}^{-1}) = 1$,

(ii) (subjective best reply) $a^* = \arg\max_{a \in A} R(a, \mu^*)$.

We conclude that the stochastic process of actions and beliefs implied by expected utility maximization converges, with probability one, to an SCE given p^*.

[28] That is, almost surely with respect to the i.i.d. process determined by θ_0.

6 Conclusions

The concept of self-confirming equilibrium (SCE) characterizes stable pairs of behaviors and beliefs. This stability is ensured through two conditions: First, the behavior must be subjectively optimal given beliefs and (smooth ambiguity) preferences. Second, these beliefs must be confirmed, that is, they must be consistent with the evidence obtained playing the equilibrium strategy. SCE with standard expected utility maximizing agents can be given a rigorous learning foundation. Indeed, the literature on stochastic control problems shows that the behavior and beliefs of an ambiguity neutral agent, who face an unknown i.i.d. process of states affecting the outcome of his actions, almost surely converge to an SCE, although such equilibrium concept was not explicitly emphasized (see Easley and Kiefer 1988 and Section 5). 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. First, it is not even apparent from the decision theoretic literature that ambiguity averse players are supposed to update beliefs according to the standard rules of conditional probabilities (see, for example, Epstein and Schneider 2007, Hanany and Klibanoff 2009). On this issue, we take instead 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. Second, ambiguity averse agents typically have dynamically inconsistent preferences over strategies. We assume that agents are sophisticated and thus take future incentives into account as they choose actions in earlier periods. This is modeled by the requirement that the adopted strategy satisfies the one-shot-deviation property, which in a finite horizon problem is equivalent to "folding-back" planning. However, dynamic inconsistency prevents us from applying standard dynamic programming techniques.

Given such difficulties, in this paper, we focused on the case of repeated play against nature to derive results and insights about convergence to SCE. Although we are mostly interested in the case of i.i.d. states, we consider the more general case of an exogenous stochastic process of states. Under smooth ambiguity, beliefs about the correct stochastic model are key. Therefore we are interested in the evolution of such beliefs, rather than the updated predictive probabilities of the states. Another essential feature of the SCE literature and our analysis is that feedback about the realized state is typically imperfect and endogenous, i.e., choice-dependent.

First, we prove convergence of beliefs under rather mild assumptions (Proposition 1). In particular, we do not require that the DM deems the actual model possible. In the i.i.d. case, we just require that the DM assigns positive probability to the set of models that are observationally equivalent to the true one under the adopted strategy. Next, this result is used to obtain almost sure convergence to a version of SCE adapted to our dynamic environment (Propositions 2, 4), and—for the i.i.d. case—to static SCE as defined by BCMM (Proposition 6). Thus, the comparative statics result of BCMM implies that higher ambiguity aversion allows for a larger set of possible long-run outcomes and therefore makes the limit behavior less predictable.

Since we do not assume that the DM assigns positive probability to the correct data generating process, our analysis belongs to the literature that studies agents with misspecified models. In particular, we relate to Esponda and Pouzo (2016). 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 the models that minimize the Kullback-Leibler divergence from the correct one. In our case, the divergence is zero due to our consistency condition (see Definition **??**). 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.[29]

We remark that our analysis provides additional insight. In several interesting decision problems or games the set of SCE, hence of possible long-run behaviors, is independent of ambiguity attitudes. Yet, ambiguity aversion affects the dynamics. We point out that higher ambiguity aversion tends to decrease experimentation and therefore makes convergence to Nash equilibrium (best reply to the correct model) more unlikely. Although this is not a general result, the intuition for this tendency is quite clear: The DM can learn only from the actions that imply an unknown likelihood of observable outcomes (otherwise, under Bayesian updating, the next-period belief would be the same as the current belief); if uncertainty about observable outcomes translates into uncertainty about payoff-relevant outcomes, then the actions from which the DM can learn are also the ambiguous actions he tends to avoid. In particular, we illustrate with a 3-color urn example that higher ambiguity aversion may make it more likely that the agent falls into a "certainty trap" whereby he keeps choosing an unambiguous action from which he cannot learn, which prevents him from finding out the objectively optimal action (see Examples 6 (Act III) and 7 (Act IV)).

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

[29] 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.

1993). The main difference is that Fudenberg and Levine consider an overlapping generation 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 generation model (cf. Blanchard 1985).

Our analysis of the more general case of an exogenous process of states of nature may shed some light on learning in a non-steady state environment, where the statistics of the populations from which co-players are drawn change over time due to their learning.

7 Appendix: proofs and related material

7.1 Jessen's Theorem

Throughout the appendix, to simplify notation we denote by Ω the set S^∞ and by $(\mathcal{F}_t)$ the natural filtration, where each $\mathcal{F}_t$ is induced by the elementary cylinders

$$s^t = (s_1, ..., s_t) = \{s_1\} \times \cdots \times \{s_t\} \times S \times \cdots ,$$

where $s^t = \mathrm{p}^t(\omega) = (\mathrm{p}_1(\omega), \ldots, \mathrm{p}_t(\omega)) = (s_1, \ldots, s_t)$ given the projections $\mathrm{p}^t : \Omega \to S^t$ and $\mathrm{p}_t : \Omega \to S$. We also consider a coarser filtration $(\mathcal{G}_t)$ where $\mathcal{G}_t \subseteq \mathcal{F}_t$ for all t, and we denote by $\mathcal{E}_t$ the (finite atomic) partition of Ω that generates $\mathcal{G}_t$. Note that a probability on $\mathcal{E}_t$ extends in a unique way to $\mathcal{G}_t$ for all t.

Remark 2. It is immediate to see that $\mathcal{E}_t$ is coarser than the partition induced by the elementary cylinders $\{\{s_1\} \times \cdots \times \{s_t\} \times S \times \cdots : (s_1, \ldots, s_t) \in S^t\}$ for all t.

We define $\mathcal{F}_\infty = \sigma(\cup_t \mathcal{F}_t)$ and $\mathcal{G}_\infty = \sigma(\cup_t \mathcal{G}_t)$. Let $\Delta(\Omega, \mathcal{F}_\infty)$ be the set of probability measures on $(\Omega, \mathcal{F}_\infty)$, and let $p, q \in \Delta(\Omega, \mathcal{F}_\infty)$. We denote by p_t and q_t the restrictions $p_{|\mathcal{G}_t}$ and $q_{|\mathcal{G}_t}$ of p and q to $\mathcal{G}_t$ for all t. Similarly, we denote by p_∞ and q_∞ the restrictions $p_{|\mathcal{G}_\infty}$ and $q_{|\mathcal{G}_\infty}$ of p and q to $\mathcal{G}_\infty$. For each t, the absolutely continuous part of q_t with respect to p_t is given by

$$q_{t,a}(E) = \begin{cases} q_t(E) & \text{if } p_t(E) > 0, \\ 0 & \text{else,} \end{cases} = \begin{cases} q(E) & \text{if } p(E) > 0, \\ 0 & \text{else,} \end{cases}$$

for each $E \in \mathcal{E}_t$, and (a version of) the Radon-Nikodym derivative of $q_{t,a}$ with respect to p_t is:

$$\lambda_t = \sum_{E \in \mathcal{E}_t : p_t(E) > 0} \frac{q_t(E)}{p_t(E)} \mathbf{1}_E = \sum_{E \in \mathcal{E}_t : p(E) > 0} \frac{q(E)}{p(E)} \mathbf{1}_E,$$

where $\mathbf{1}_E$ is the indicator function for event E. In particular, for each $\omega \in \Omega$, letting $E_t(\omega)$ denote the only element in $\mathcal{E}_t$ containing ω,[30]

$$\lambda_t(\omega) = \begin{cases} \frac{q_t(E_t(\omega))}{p_t(E_t(\omega))} & \text{if } p(E_t(\omega)) > 0, \\ 0 & \text{else,} \end{cases} = \begin{cases} \frac{q(E_t(\omega))}{p(E_t(\omega))} & \text{if } p(E_t(\omega)) > 0, \\ 0 & \text{else.} \end{cases}$$

[30] Given Remark 2, observe that if $(\omega_1, \ldots, \omega_t) = (\omega'_1, \ldots, \omega'_t)$ then $E_t(\omega) = E_t(\omega')$.

For each t, the set

$$C(p_t) := \bigcup_{E \in \mathcal{E}_t : p_t(E) > 0} E = \{\omega \in \Omega : p(E_t(\omega)) > 0\} \in \mathcal{G}_t$$

is the *minimal carrier of* p_t *in* $\mathcal{G}_t$; that is, it is the smallest event F in $\mathcal{G}_t$ such that $p_t(F) = p_\infty(F) = p(F) = 1$. Then, for each t,

$$\lambda_t(\omega) = \frac{q(E_t(\omega))}{p(E_t(\omega))} \mathbf{1}_{C(p_t)}(\omega), \quad \forall \omega \in \Omega.$$

We introduce one last object, the set

$$C(p_\infty) := \bigcap_{t=1}^{\infty} C(p_t) = \{\omega \in \Omega : p(E_t(\omega)) > 0 \ \ \forall t\} \in \mathcal{G}_\infty.$$

Clearly, we have that $p_\infty(C(p_\infty)) = 1$, and

$$\forall \omega \in C(p_\infty), \forall t, \ \lambda_t(\omega) = \frac{q(E_t(\omega))}{p(E_t(\omega))} = \frac{q_\infty(E_t(\omega))}{p_\infty(E_t(\omega))}.$$

The next theorem is credited to Jessen by Stroock (1993).

Theorem 1 (Jessen)**.** *If* $p, q \in \Delta(\Omega, \mathcal{F}_\infty)$*,* $q_\infty \perp p_\infty$ *if and only if* $\lambda_t \to 0$ p_∞*-a.s.*

Let μ be a *prior* on $\Delta(\Omega, \mathcal{F}_\infty)$ with finite support $\operatorname{supp}\mu$. The predictive measure p_μ is:

$$\forall E \in \mathcal{F}_\infty, \ p_\mu(E) = \sum_{q \in \operatorname{supp}\mu} q(E)\mu(q).$$

For each t and $E \in \mathcal{E}_t$, the posterior distribution of μ given E is defined by:

$$\mu(q \mid E) = \begin{cases} \frac{\mu(q)q(E)}{p_\mu(E)} & \text{if } p_\mu(E) > 0, \\ \frac{\mathbf{1}_{\operatorname{supp}\mu}(q)}{|\operatorname{supp}\mu|} & \text{otherwise.} \end{cases}$$

Note that $\mu(q \mid E) = 0$ for all $q \notin \operatorname{supp}\mu$, for all $E \in \mathcal{E}_t$ and all t. For each $q \in \operatorname{supp}\mu$ and t, consider the function $\mu^t(q \mid \cdot) : \Omega \to [0,1]$ given by:

$$\begin{aligned} \mu^t(q \mid \omega) &:= \mu(q \mid E_t(\omega)) \\ &= \tfrac{\mu(q)q(E_t(\omega))}{p_\mu(E_t(\omega))} \mathbf{1}_{C(p_{\mu,t})}(\omega) + \tfrac{\mathbf{1}_{\operatorname{supp}\mu}(q)}{|\operatorname{supp}\mu|}\left(1 - \mathbf{1}_{C(p_{\mu,t})}(\omega)\right) \\ &= \tfrac{\mu(q)q(E_t(\omega))}{p_\mu(E_t(\omega))} \mathbf{1}_{C(q_t)} \quad q_\infty\text{-a.s.}; \end{aligned}$$

for each $q \in \operatorname{supp}\mu$ and t, $\mu^t(q \mid \cdot)$ is an element of $L^1_+(\Omega, \mathcal{G}_t, q_t)$.

For each t, the map $t \mapsto \mu^t(\cdot|\cdot)$ has a natural extension on $\sigma(\Delta(\Omega, \mathcal{F}_\infty)) \times \Omega$:

$$\mu^t(D \mid \omega) := \sum_{q \in \operatorname{supp}\mu \cap D} \mu^t(q \mid \omega) \quad \forall (D, \omega) \in \sigma(\Delta(\Omega, \mathcal{F}_\infty)) \times \Omega.$$

Without loss of generality, we use the same symbol to denote both the original function and this extension.

The function μ^t is the posterior of μ given $\mathcal{G}_t$. Note that, for each $D \in \sigma\left(\Delta\left(\Omega, \mathcal{F}_\infty\right)\right)$, the function $\mu^t\left(D \mid \cdot\right)$ is $\mathcal{G}_t$-measurable.

Consider the following measurable equivalence relation in $\Delta\left(\Omega, \mathcal{F}_\infty\right)$:

$$p \sim q \Leftrightarrow p_\infty = q_\infty.$$

We denote by $[p]$ the equivalence class, with respect to $\sim$, that contains p. It is easy to show that $[p] \in \sigma\left(\Delta\left(\Omega, \mathcal{F}_\infty\right)\right)$. The next result follows from Theorem 1.

Proposition 8. *Let μ be a prior on $\Delta\left(\Omega, \mathcal{F}_\infty\right)$ with finite support. If $\bar{p} \in \Delta\left(\Omega, \mathcal{F}_\infty\right)$ is such that $[\bar{p}] \cap \operatorname{supp} \mu \neq \emptyset$ and $q_\infty \perp \bar{p}_\infty$ for all $q \in \operatorname{supp} \mu \setminus [\bar{p}]$, then there exists a set $D\left([\bar{p}]\right) \in \mathcal{G}_\infty$ such that:*

(i) $\bar{p}\left(D\left([\bar{p}]\right)\right) = \bar{p}_\infty\left(D\left([\bar{p}]\right)\right) = 1$*; and,*

(ii) for each $\omega \in D\left([\bar{p}]\right)$, $\lim_{t\to\infty} \mu^t\left([\bar{p}] \mid \omega\right) = 1$.

Proof. Set $Q = \operatorname{supp}\mu$ and $\bar{Q} = \operatorname{supp} \mu \setminus [\bar{p}]$. Consider $\check{p} \in [\bar{p}] \cap \operatorname{supp} \mu$. For each $q \in \bar{Q}$, $q_\infty \perp \check{p}_\infty$; By Theorem 1, we can produce a set $B\left(q\right) \in \mathcal{G}_\infty$ such that:

(a) $B\left(q\right) \subseteq C\left(\check{p}_\infty\right) = C\left(\bar{p}_\infty\right)$;

(b) $\dfrac{q_\infty\left(E_t\left(\omega\right)\right)}{\bar{p}_\infty\left(E_t\left(\omega\right)\right)} = \dfrac{q_\infty\left(E_t\left(\omega\right)\right)}{\check{p}_\infty\left(E_t\left(\omega\right)\right)} \to 0$ for all $\omega \in B\left(q\right)$;

(c) $\bar{p}_\infty\left(B\left(q\right)\right) = \check{p}\left(B\left(q\right)\right) = 1$.

If we define $D\left([\bar{p}]\right) := \cap_{q\in\bar{Q}} B\left(q\right) \subseteq C\left(\bar{p}_\infty\right)$, it follows that $\bar{p}_\infty\left(D\left([\bar{p}]\right)\right) = 1$ —which gives point (i) — and

$$\frac{q_\infty\left(E_t\left(\omega\right)\right)}{\bar{p}_\infty\left(E_t\left(\omega\right)\right)} \to 0$$

for all $q \in \bar{Q}$ and all $\omega \in D\left([\bar{p}]\right)$. Consider $\omega \in D\left([\bar{p}]\right) \subseteq C\left(\bar{p}_\infty\right)$. It follows that $\check{p}\left(E_t\left(\omega\right)\right) = \bar{p}\left(E_t\left(\omega\right)\right) > 0$, hence $p_\mu\left(E_t\left(\omega\right)\right) > 0$ for all t. We conclude that:

$$\begin{aligned}
\mu^t\left([\bar{p}] \mid \omega\right) = \mu^t\left([\bar{p}] \cap \operatorname{supp} \mu \mid \omega\right) &= \textstyle\sum_{p\in Q\setminus\bar{Q}} \frac{\mu(p)p(E_t(\omega))}{\sum_{q\in Q} q(E_t(\omega))\mu(q)} \\
&= \tfrac{\sum_{p\in Q\setminus\bar{Q}} \mu(p)p(E_t(\omega))}{\sum_{p\in Q\setminus\bar{Q}} \mu(p)p(E_t(\omega))+\sum_{q\in\bar{Q}} q(E_t(\omega))\mu(q)} \\
&= \tfrac{\sum_{p\in Q\setminus\bar{Q}} \mu(p)p_\infty(E_t(\omega))}{\sum_{p\in Q\setminus\bar{Q}} p_\infty(E_t(\omega))\mu(p)+\sum_{q\in\bar{Q}} q_\infty(E_t(\omega))\mu(q)} \\
&= \tfrac{\bar{p}_\infty(E_t(\omega)) \sum_{p\in Q\setminus\bar{Q}} \mu(p)}{\bar{p}_\infty(E_t(\omega)) \sum_{p\in Q\setminus\bar{Q}} \mu(p)+\sum_{q\in\bar{Q}} q_\infty(E_t(\omega))\mu(q)} \\
&= \tfrac{\sum_{p\in Q\setminus\bar{Q}} \mu(p)}{\sum_{p\in Q\setminus\bar{Q}} \mu(p)+\sum_{q\in\bar{Q}} \frac{q_\infty(E_t(\omega))}{\bar{p}_\infty(E_t(\omega))}\mu(q)} \\
&= \tfrac{\sum_{p\in Q\setminus\bar{Q}} \mu(p)}{\sum_{p\in Q\setminus\bar{Q}} \mu(p)+\sum_{q\in\bar{Q}} \lambda_t(\omega)\mu(q)} \\
&\to \tfrac{\sum_{p\in Q\setminus\bar{Q}} \mu(p)}{\sum_{p\in Q\setminus\bar{Q}} \mu(p)} = 1,
\end{aligned}$$

proving point (ii). ■

Given $p \in \Delta(\Omega, \mathcal{F}_\infty)$ and the algebra $\mathcal{G}_t$, we denote by $p^t(\cdot \mid \cdot) : \mathcal{F}_\infty \times \Omega \to [0,1]$ any function with the following properties:

(i) $p^t(\cdot \mid \omega)$ is a probability measure on $\mathcal{F}_\infty$ for each $\omega \in \Omega$;

(ii) $p^t(B \mid \cdot)$ is a version of the conditional probability of B given $\mathcal{G}_t$.

We call p^t the regular conditional probability of p given $\mathcal{G}_t$. Since $\mathcal{G}_t$ is finite, then by Gray (2009, Lemma 6.7), the existence of p^t is guaranteed for all t. We also have that, for each t and for each $\omega' \in \Omega$, if $p(E_t(\omega')) > 0$, then

$$\forall F \in \mathcal{G}_\infty,\ p^t(F \mid \omega') = \frac{p(F \cap E_t(\omega'))}{p(E_t(\omega'))}. \tag{16}$$

For each $\omega' \in \Omega$ and for each t, $p \sim q$ and $p(E_t(\omega')) > 0$ imply $p^t(\cdot \mid \omega') = q^t(\cdot \mid \omega')$.

Corollary 4. *Suppose that*

- μ, $\bar{p}$ *and* $D([\bar{p}])$ *are as in Proposition 8;*
- (v_t) *is a uniformly bounded process such that* v_t *is* $\mathcal{G}_\infty$*-measurable for all* t*; in particular, there exists* $a, b \in \mathbb{R}$ *such that* $v_t(\omega) \in [a,b]$ *for all* $(t,\omega) \in \mathbb{N} \times \Omega$*;*
- *and* $\phi : [a,b] \to \mathbb{R}$ *is a strictly increasing and continuous function;*

Then, for all $\omega' \in D([\bar{p}])$*,*

$$\left| \phi^{-1}\left(\int_{\Delta(\Omega,\mathcal{F}_\infty)} \phi\left(\int_\Omega v_t(\omega)\, q^t(\mathrm{d}\omega \mid \omega') \right) \mu^t(\mathrm{d}q \mid \omega') \right) - \int_\Omega v_t(\omega)\, \bar{p}^t(\mathrm{d}\omega \mid \omega') \right| \to 0$$

Proof For each t and for each $\omega' \in \Omega$ such that $p_\mu(E_t(\omega')) > 0$, we have that $\mu^t(q \mid \omega') = \frac{\mu(q) q(E_t(\omega'))}{p_\mu(E_t(\omega'))} > 0$ if and only if $q \in \operatorname{supp}\mu$ and $q(E_t(\omega')) > 0$, namely, $\operatorname{supp}\mu^t(\cdot \mid \omega') = \{q \in \operatorname{supp}\mu : q(E_t(\omega')) > 0\}$. Thus, for each t and each $\omega' \in \Omega$ such that $p_\mu(E_t(\omega')) > 0$, it holds that

$$\begin{aligned} &\int_{\Delta(\Omega,\mathcal{F}_\infty)} \left(\int_\Omega v_t(\omega)\, q^t(\mathrm{d}\omega \mid \omega') \right) \mu^t(\mathrm{d}q \mid \omega') \\ = &\sum_{q \in \operatorname{supp}\mu^t(\cdot \mid \omega')} \left(\int_\Omega v_t(\omega)\, q^t(\mathrm{d}\omega \mid \omega') \right) \mu^t(q \mid \omega'). \end{aligned}$$

Let $\omega' \in D([\bar{p}]) \subseteq C(\bar{p}_\infty)$. Since there exists $\check{p} \in [\bar{p}] \cap \operatorname{supp}\mu$, we have $\check{p}(E_t(\omega')) = \bar{p}(E_t(\omega')) > 0$ for all t, implying that $p_\mu(E_t(\omega')) > 0$. Then, letting $K > 0$ be a bound on ϕ (which exists because ϕ is a continuous function defined on a compact domain),

$$\begin{aligned}
&\left|\int_{\Delta(\Omega,\mathcal{F}_\infty)} \phi\left(\int_\Omega v_t(\omega)\, q^t(\mathrm{d}\omega \mid \omega')\right)\mu^t(\mathrm{d}q \mid \omega') - \phi\left(\int_\Omega v_t(\omega)\, \bar{p}^t(\mathrm{d}\omega \mid \omega')\right)\right| \\
&= \left|\int_{\Delta(\Omega,\mathcal{F}_\infty)} \phi\left(\int_\Omega v_t(\omega)\, q^t(\mathrm{d}\omega \mid \omega')\right)\mu^t(\mathrm{d}q \mid \omega') - \phi\left(\int_\Omega v_t(\omega)\, \check{p}^t(\mathrm{d}\omega \mid \omega')\right)\right| \\
&\leq \left|\sum_{\substack{q\in\operatorname{supp}\mu \\ q(E_t(\omega'))>0,\, q\notin[\bar{p}]}} \phi\left(\int_\Omega v_t(\omega)\, q^t(\mathrm{d}\omega \mid \omega')\right)\mu^t(q \mid \omega')\right| \\
&+ \left|\sum_{\substack{q\in\operatorname{supp}\mu \\ q(E_t(\omega'))>0,\, q\in[\bar{p}]}} \phi\left(\int_\Omega v_t(\omega)\, q^t(\mathrm{d}\omega \mid \omega')\right)\mu^t(q \mid \omega') - \phi\left(\int_\Omega v_t(\omega)\, \check{p}^t(\mathrm{d}\omega \mid \omega')\right)\right| \\
&\leq K \sum_{\substack{q\in\operatorname{supp}\mu \\ q(E_t(\omega'))>0,\, q\notin[\bar{p}]}} \mu^t(q \mid \omega') \\
&+ \left|\sum_{\substack{q\in\operatorname{supp}\mu \\ q(E_t(\omega'))>0,\, q\in[\bar{p}]}} \phi\left(\int_\Omega v_t(\omega)\, \check{p}^t(\mathrm{d}\omega \mid \omega')\right)\mu^t(q \mid \omega') - \phi\left(\int_\Omega v_t(\omega)\, \check{p}^t(\mathrm{d}\omega \mid \omega')\right)\right| \\
&\leq K \sum_{\substack{q\in\operatorname{supp}\mu \\ q(E_t(\omega'))>0, q\notin[\bar{p}]}} \mu^t(q \mid \omega') + K\left|\sum_{\substack{q\in\operatorname{supp}\mu \\ q(E_t(\omega'))>0, q\in[\bar{p}]}} \mu^t(q \mid \omega') - 1\right| \\
&= K\left[\mu^t([\bar{p}]^c \mid \omega') - (1 - \mu^t([\bar{p}] \mid \omega'))\right].
\end{aligned}$$

By Proposition 8, the last expression converges to 0. Finally, for each $(t, \omega') \in \mathbb{N} \times D([\bar{p}])$, set

$$\begin{aligned}
z_t(\omega') &= \int_{\Delta(\mathcal{F}_\infty)} \phi\left(\int_\Omega v_t(\omega)\, q^t(\mathrm{d}\omega \mid \omega')\right)\mu^t(\mathrm{d}q \mid \omega'); \\
y_t(\omega') &= \phi\left(\int_\Omega v_t(\omega)\, \bar{p}^t(\mathrm{d}\omega \mid \omega')\right).
\end{aligned}$$

Note that $z_t(\omega'), y_t(\omega') \in [\phi(a), \phi(b)]$ and that $\phi^{-1} : [\phi(a), \phi(b)] \to [a, b]$ is uniformly continuous (since ϕ is continuous and strictly increasing on a closed and bounded interval, and so is its inverse). Therefore, for each $\varepsilon > 0$ there exists $\delta = \delta(\varepsilon) > 0$, such that $\left|\phi^{-1}(z) - \phi^{-1}(y)\right| \leq \varepsilon$ for all $y, z \in [\phi(a), \phi(b)]$ such that $|z - y| \leq \delta$. Fix $\omega' \in D([\bar{p}])$. By the previous part of the proof, for each $\varepsilon > 0$, there exists $n = n(\varepsilon)$ such that $|z_t(\omega') - y_t(\omega')| \leq \delta(\varepsilon)$ for all $t \geq n$, proving the statement. ■

7.2 Further proofs

Proof of Lemma 1 Define the correspondence $\mathcal{S}_t^\alpha : S^{t-1} \to 2^{S^\infty}$ by $\mathcal{S}_t^\alpha(s^{t-1}) = \iota_t^\alpha(s^{t-1}) \times S^\infty$; for each finite state history s^{t-1}, $\mathcal{S}_t^\alpha(s^{t-1})$ is the set of infinite state histories that yield the same information history up to t under α as s^{t-1}. Thus, $\mathcal{S}_t^\alpha$ is the correspondence of observationally-equivalent infinite state histories under α. Fix s^t with $\bar{p}(s^t) > 0$. Note that $s^t \in \mathcal{S}_{t+1}^\alpha(s^t) \subseteq \mathcal{S}_t^\alpha(s^{t-1})$; thus, $\bar{p}(\mathcal{S}_t^\alpha(s^{t-1})) \geq \bar{p}\left(\mathcal{S}_{t+1}^\alpha(s^t)\right) \geq \bar{p}(s^t) > 0$. Let $p \in \hat{P}_t^{\alpha,\mu}(\bar{p})(s^{t-1})$; by definition, $p(\mathcal{S}_t^\alpha(s^{t-1})) > 0$. We want to show that $p \in \hat{P}_{t+1}^{\alpha,\mu}(\bar{p})(s^t)$. To this end, fix $E \in \sigma(\mathbf{h}^\alpha)$ with $E \subseteq \mathcal{S}_{t+1}^\alpha(s^t)$. Since $p \in \hat{P}_t^{\alpha,\mu}(\bar{p})(s^{t-1})$, then

$$\frac{p(E)}{p(\mathcal{S}_t^\alpha(s^{t-1}))} = \frac{\bar{p}(E)}{\bar{p}(\mathcal{S}_t^\alpha(s^{t-1}))}; \quad \frac{p\left(\mathcal{S}_{t+1}^\alpha(s^t)\right)}{p(\mathcal{S}_t^\alpha(s^{t-1}))} = \frac{\bar{p}\left(\mathcal{S}_{t+1}^\alpha(s^t)\right)}{\bar{p}(\mathcal{S}_t^\alpha(s^{t-1}))}.$$

The second equality implies $p\left(\mathcal{S}_{t+1}^{\alpha}\left(s^{t}\right)\right)>0$. Since

$$\frac{p\left(E\right)}{p\left(\mathcal{S}_{t+1}^{\alpha}(s^{t})\right)}\frac{p\left(\mathcal{S}_{t+1}^{\alpha}(s^{t})\right)}{p\left(\mathcal{S}_{t}^{\alpha}(s^{t-1})\right)}=\frac{\bar{p}\left(E\right)}{\bar{p}\left(\mathcal{S}_{t+1}^{\alpha}(s^{t})\right)}\frac{\bar{p}\left(\mathcal{S}_{t+1}^{\alpha}(s^{t})\right)}{\bar{p}\left(\mathcal{S}_{t}^{\alpha}(s^{t-1})\right)},$$

it follows that

$$\frac{p\left(E\right)}{p\left(\mathcal{S}_{t+1}^{\alpha}(s^{t})\right)}=\frac{\bar{p}\left(E\right)}{\bar{p}\left(\mathcal{S}_{t+1}^{\alpha}(s^{t})\right)}.$$

Hence, $p^{\alpha}\left(\cdot\mid\mathbf{h}_{t+1}^{\alpha}\left(s^{t}\right)\right)=\bar{p}^{\alpha}\left(\cdot\mid\mathbf{h}_{t+1}^{\alpha}\left(s^{t}\right)\right)$. Since $p\in\operatorname{supp}\mu\left(\cdot\mid\mathbf{h}_{t}^{\alpha}\left(s^{t-1}\right)\right)$ and $p\left(\mathcal{S}_{t+1}^{\alpha}\left(s^{t}\right)\right)>0$, it follows that $p\in\operatorname{supp}\mu\left(\cdot\mid\mathbf{h}_{t+1}^{\alpha}\left(s^{t}\right)\right)$, hence $p\in\hat{P}_{t+1}^{\alpha,\mu}\left(\bar{p}\right)\left(s^{t}\right)$. ■

Consider the coarser filtration $(\mathcal{G}_t)$ where $\mathcal{G}_t=\sigma\left(\mathbf{h}_{t+1}^{\alpha}\right)\subseteq\mathcal{F}_t$ for all t. Note that $\mathcal{G}_\infty=\sigma\left(\mathbf{h}^{\alpha}\right)$ and, given $\bar{p}\in\Delta\left(\Omega,\mathcal{F}_\infty\right)$, it holds that $[\bar{p}]\cap\operatorname{supp}\mu=\hat{P}_{t}^{\alpha,\mu}\left(\bar{p}\right)$. Given a prior μ, we denote its posterior μ^t given by $\mathcal{G}_t$ alternatively as $\mu\left(\cdot\mid\mathbf{h}_{t+1}^{\alpha}\left(\cdot\right)\right)$, that is, $\mu\left(D\mid\mathbf{h}_{t+1}^{\alpha}\left(\omega\right)\right)=\mu^{t}\left(D\mid\omega\right)$ for all $(D,\omega)\in\sigma\left(\Delta\left(\mathcal{F}_\infty\right)\right)\times\Omega$. As the posterior is $\mathcal{G}_t$-measurable with respect to the second component and $\mathcal{G}_t\subseteq\mathcal{F}_t$, if ω and $\bar{\omega}$ are such that $(s_1,...,s_t)=(\bar{s}_1,...,\bar{s}_t)$, then $\mu\left(D\mid\mathbf{h}_{t+1}^{\alpha}\left(\omega\right)\right)=\mu\left(D\mid\mathbf{h}_{t+1}^{\alpha}\left(\bar{\omega}\right)\right)$. Thus, with a slight abuse of notation, the second argument ω can be replaced by the finite history $(s_1,...,s_t)$. Finally, given $p\in\Delta\left(\Omega,\mathcal{F}_\infty\right)$ and the algebra $\sigma\left(\mathbf{h}_{t+1}^{\alpha}\right)$, we denote the regular conditional probability either by $p^{t}\left(\cdot\mid\cdot\right)$ or by $p\left(\cdot\mid\mathbf{h}_{t+1}^{\alpha}(\cdot)\right)$.

Proof of Proposition 1 By Proposition 8 and the above notation, there exists a set $D\left([\bar{p}]\right)\in\mathcal{G}_\infty$ such that $\bar{p}_\infty\left(D\left([\bar{p}]\right)\right)=1$ and, for each $\omega\in D\left([\bar{p}]\right)$, $\lim_{t\to\infty}\mu^{t}\left([\bar{p}]\mid\omega\right)=1$. Moreover, we know that if $\mu(p)=0$, then $\mu^{t}\left(p\mid\omega\right)=0$ for every $\omega\in\Omega$. This proves that, if $T=1$, then $\mu\left([\bar{p}]\cap\operatorname{supp}\mu\mid\mathbf{h}_{t}^{\alpha}(\cdot)\right)=\mu\left(\hat{P}_{1}^{\alpha,\mu}\left(\bar{p}\right)\mid\mathbf{h}_{t}^{\alpha}(\cdot)\right)\to 1$ $\bar{p}^{\alpha}$-a.s. Let $(\alpha,\mu,\bar{p})$ be consistent from period $T>1$. For each $h_T=(a^{T-1},m^{T-1})$ consistent with α and such that $\bar{p}\left(I(h_T)\right)>0$, we can look at the triple $\left(\alpha_{h_T},\mu_{h_T},\bar{p}_{h_T}\right)$ obtained from $(\alpha,\mu,\bar{p})$ by initializing the strategy and the processes at information history h_T:

- $\alpha_{h_T}\left(a^0,m^0\right):=\alpha(a^{T-1},m^{T-1})$, $\alpha_{h_T}(a^k,m^k):=\alpha\left((a^{T-1},a^k),(m^{T-1},m^k)\right)$;
- $\mu_{h_T}(p_{h_T}):=\mu([p_{h_T}]\,|h_T)$ where

$$[p_{h_T}]:=\left\{p\in\Delta\left(S\right):\forall E\in\sigma(S^\infty),\,p_{h_T}(E):=p\left(I(h_T)\times E|I(h_T)\right)\right\};$$

- $\bar{p}_{h_T}(E):=\bar{p}\left(I(h_T)\times E|I(h_T)\right)$ for each $E\in\sigma(S^\infty)$.

Then $\left(\alpha_{h_T},\mu_{h_T},\bar{p}_{h_T}\right)$ is consistent from period 1, so

$$\lim_{t\to\infty}\mu_{h_T}\left(\hat{P}_{1}^{\alpha_{h_T},\mu_{h_T}}\left(\bar{p}_{h_T}\right)\middle|\mathbf{h}_{t}^{\alpha_{h_T}}\left(\cdot\right)\right)=1\;\;\bar{p}_{h_T}^{\alpha_{h_T}}-a.s.,$$

which in turn implies that

$$\lim_{t\to\infty}\mu\left(\hat{P}_{T}^{\alpha,\mu}\left(\bar{p}\right)|\mathbf{h}_{t}^{\alpha}(\cdot)\right)=1\;\;\bar{p}^{\alpha}(\cdot|I(h_T))-a.s.$$

Since this holds for each h_T consistent with α such that $\bar{p}^{\alpha}(I(h_T)) > 0$, the thesis follows. ■

Proof of Corollary 1 Since $\mathcal{G}_t = \mathcal{F}_t$ for all $t > 1$, it follows that $\hat{P}_1^{\alpha}(\bar{p}) = \{\bar{p}\}$ and $\sigma(\mathbf{h}^{\alpha}) = \mathcal{F}_{\infty}$. By Proposition 1, the statement follows. ■

Proof of Corollary 2 Consider the process (v_t) such that $v_t(\omega) = \mathbf{1}_{B_t}(\omega)$ for all $\omega \in \Omega$ and for all t. It follows that (v_t) is $\sigma(\mathbf{h}^{\alpha})$-measurable and clearly uniformly bounded. By Corollary 4 it follows that,

$$\begin{aligned}&\left|p_{\mu}\left(B_t \mid \mathbf{h}_t^{\alpha}(\omega')\right) - \bar{p}\left(B_t \mid \mathbf{h}_t^{\alpha}(\omega')\right)\right| \\ &= \left|\int_{\Delta(\Omega,\mathcal{F}_{\infty})} \phi\left(\int_{\Omega} v_t(\omega)\, q^t(\mathrm{d}\omega \mid \omega')\right) \mu^t(\mathrm{d}q \mid \omega') - \phi\left(\int_{\Omega} v_t(\omega)\, \bar{p}^t(\mathrm{d}\omega \mid \omega')\right)\right| \to 0,\end{aligned}$$

for all $\omega' \in D([\bar{p}])$. Let ϕ be the identity function, since $\bar{p}(D([\bar{p}])) = 1$ and $D([\bar{p}]) \in \sigma(\mathbf{h}^{\alpha})$, the statement follows. ■

Proof of Lemma 2 First note that since $p_{\pi_\nu}(I(h_t))$ and $p_{\pi_\nu}(I(h_{t'}))$ are strictly positive, then

$$\frac{\nu(\pi)\, p_{\pi}(I(h_{t'}))}{p_{\pi_\nu}(I(h_{t'}))} = \nu(\pi \mid h_{t'}) = \nu(\pi \mid h_t) = \frac{\nu(\pi)\, p_{\pi}(I(h_t))}{p_{\pi_\nu}(I(h_t))}.$$

In particular,

$$\nu(\pi \mid h_{t'}) = \nu(\pi \mid h_t) > 0 \;\Rightarrow p_{\pi}(I(h_t)) > 0, \text{ and } p_{\pi}(I(h_{t'})) > 0.$$

That is, the models in the support of the $\nu(\cdot \mid h_{t'}) = \nu(\pi \mid h_t)$ assign positive probability to the two conditioning events. In turn, this implies that $p_{\pi}(\cdot|h_t)$ is well defined by (16). Hence we have:

$$\begin{aligned}&V(\alpha, \nu \mid h_t) \\ =\;& \sum_{\tau=t}^{\infty} \delta^{\tau-t} \phi^{-1}\left(\int_{\operatorname{supp}\nu(\cdot|h_t)} \phi\left(\sum_{s^{\tau}\in S^{\tau}} r\left(\mathbf{a}_{\tau}^{\alpha}\left(s^{\tau-1}\right), s_{\tau}\right) p_{\pi}(s^{\tau}|h_t)\right) \nu(\mathrm{d}\pi \mid h_t)\right) \quad (17)\end{aligned}$$

To show our result, we will prove that for every n in $\mathbb{N}_0$,

$$\begin{aligned}&\phi^{-1}\left(\int_{\operatorname{supp}\nu(\cdot|h_t)} \phi\left(\sum_{s^{t+n}:\mathbf{h}_t^{\alpha}(s^{t-1})=h_t} r\left(\mathbf{a}_{t+n}^{\alpha}\left(s^{t+n-1}\right), s_{t+n}\right) p_{\pi}(s^{t+n}|h_t)\right) \nu(\mathrm{d}\pi \mid h_t)\right) \\ =\;& \phi^{-1}\left(\int_{\operatorname{supp}\nu(\cdot|h_{t'})} \phi\left(\sum_{s^{t'+n}:\mathbf{h}_{t'}^{\alpha}(s^{t'-1})=h_{t'}} r\left(\mathbf{a}_{t'+n}^{\alpha}\left(s^{t'+n-1}\right), s_{t'+n}\right) p_{\pi}(s^{t'+n}|h_{t'})\right) \nu(\mathrm{d}\pi \mid h_{t'})\right).\end{aligned}$$

Since $V(\alpha, \nu \mid h_t)$ and $V(\alpha, \nu \mid h_{t'})$ are defined as the sum from $n = 0$ to infinity of, respectively, the first and second line above, the statement will follow.

Choose arbitrarily $n \in \mathbb{N}_0$, and $\pi \in \operatorname{supp} \nu(\cdot \mid h_t) = \operatorname{supp} \nu(\cdot \mid h_{t'})$. Define

$$K_{(k_0,...,k_n)} := \left\{ s^{t+n} | s_t = k_0, ..., s_{t+n} = k_n \right\} \cap \left\{ s^{t+n} : \mathbf{h}_t^\alpha(s^{t-1}) = h_t \right\}$$

and

$$K'_{(k_0,...,k_n)} := \left\{ s^{t'+n} | s_{t'} = k_0, ..., s_{t'+n} = k_n \right\} \cap \left\{ s^{t'+n} : \mathbf{h}_{t'}^\alpha(s^{t'-1}) = h_{t'} \right\}$$

for some $(k_0, ..., k_n)$ such that $\pi(k_i) \neq 0$ for every i in $\{1, ..., n\}$. Note that, since P is composed by i.i.d. models, then

$$p_\pi(K_{(k_0,...,k_n)} | h_t) = \prod_{i=0}^{n} \pi(k_n) = p_\pi(K'_{(k_0,...,k_n)} | h_{t'}).$$

To ease notation, fix $(k_0, ..., k_n)$ and let $K = K_{(k_0,...,k_n)}$ and $K' = K'_{(k_0,...,k_n)}$. Note that

$$r\left(\mathbf{a}_{t+n}^\alpha\left(s^{t+n-1}\right), s_{t+n}\right)$$

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

$$\mathbf{a}_t^\alpha\left(s^{t-1}\right) = \alpha\left(\nu\left(\cdot | h_t\right)\right).$$

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

$$\mathbf{h}_{t+j}^\alpha(s^{t+j-1}) = \left(h_t, \mathbf{a}_t^\alpha\left(s^{t-1}\right), f\left(\mathbf{a}_t^\alpha\left(s^{t-1}\right), s_t\right), ..., f\left(\mathbf{a}_{t+j-1}^\alpha\left(s^{t+j-1}\right), s_{t+j-1}\right)\right)$$

that, by definition of K and by the inductive hypothesis, is costant on K. But then it follows that

$$\mathbf{a}_{t+j}^\alpha\left(s^{t+j-1}\right) = \alpha\left(\nu\left(\cdot | \mathbf{h}_{t+j}^\alpha(s^{t+j-1})\right)\right)$$

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

$$\nu\left(\cdot | \mathbf{h}_{t+j}^\alpha(s^{t+j-1})\right) = \nu\left(\cdot | \mathbf{h}_{t'+j}^\alpha(s^{t'+j-1})\right).$$

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

$$\begin{aligned} \mathbf{a}_{t+j-1}^\alpha\left(s^{t+j-2}\right) &= \alpha\left(\nu\left(\cdot | \mathbf{h}_{t+j-1}^\alpha(s^{t+j-2})\right)\right) \\ &= \alpha\left(\nu\left(\cdot | \mathbf{h}_{t'+j-1}^\alpha(s^{t'+j-2})\right)\right) \\ &= \mathbf{a}_{t'+j-1}^\alpha\left(s^{t'+j-2}\right). \end{aligned}$$

Therefore:

$$\begin{aligned}
& \nu\left(\pi|\mathbf{h}^\alpha_{t+j}(s^{t+j-1})\right) \\
= \;& \frac{\nu\left(\cdot|\mathbf{h}^\alpha_{t+j-1}(s^{t+j-2})\right)\pi\left(f^{-1}_{\mathbf{a}^\alpha_{t+j-1}(s^{t+j-2})}\left(f\left(\mathbf{a}^\alpha_{t+j-1}\left(s^{t+j-2}\right),k_{j-1}\right)\right)\right)}{\pi_{\nu\left(\cdot|\mathbf{h}^\alpha_{t+j-1}(s^{t+j-2})\right)}\left(f^{-1}_{\mathbf{a}^\alpha_{t+j-1}(s^{t+j-2})}\left(f\left(\mathbf{a}^\alpha_{t+j-1}\left(s^{t+j-2}\right),k_{j-1}\right)\right)\right)} \\
= \;& \frac{\nu\left(\cdot|\mathbf{h}^\alpha_{t'+j-1}(s^{t'+j-2})\right)\pi\left(f^{-1}_{\mathbf{a}^\alpha_{t'+j-1}\left(s^{t'+j-2}\right)}\left(f\left(\mathbf{a}^\alpha_{t'+j-1}\left(s^{t'+j-2}\right),k_{j-1}\right)\right)\right)}{\pi_{\nu\left(\cdot|\mathbf{h}^\alpha_{t'+j-1}(s^{t'+j-2})\right)}\left(f^{-1}_{\mathbf{a}^\alpha_{t'+j-1}\left(s^{t'+j-2}\right)}\left(f\left(\mathbf{a}^\alpha_{t'+j-1}\left(s^{t'+j-2}\right),k_{j-1}\right)\right)\right)} \\
= \;& \nu\left(\pi|\mathbf{h}^\alpha_{t'+j-1}(s^{t'+j-1})\right).
\end{aligned}$$

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

$$\begin{aligned}
r\left(\mathbf{a}^\alpha_{t+n}\left(s^{t+n-1}\right),s_{t+n}\right) &= r\left(\alpha\left(\nu\left(\cdot|\mathbf{h}^\alpha_{t+n}(s^{t+n-1})\right)\right),k_n\right) \\
&= r\left(\alpha\left(\nu\left(\cdot|\mathbf{h}^\alpha_{t'+n}(s^{t'+n-1})\right)\right),k_n\right) \qquad (18) \\
&= r\left(\mathbf{a}^\alpha_{t'+n}\left(s^{t'+n-1}\right),s_{t'+n}\right).
\end{aligned}$$

Now, we restart to explicitly highlight the dependence on $(k_0, ..., k_n)$ of K. Moreover, for every $n \in \mathbb{N}_0$ and for every $(k_1, ..., k_n) \in S^n$, let $r(k_0, ..., k_n) = r\left(\mathbf{a}^\alpha_{t+n}\left(s^{t+n-1}\right),s_{t+n}\right) = r\left(\mathbf{a}^\alpha_{t'+n}\left(s^{t'+n-1}\right),s_{t'+n}\right)$, where $s^{t+n} \in K_{k_0,...,k_n}$ and $s^{t'+n} \in K'_{k_0,...,k_n}$. By (18) this quantity is well defined. We have:

$$\begin{aligned}
& \sum_{s^{t+n}:\mathbf{h}^\alpha_t(s^{t+n})=h_t} r\left(\mathbf{a}^\alpha_{t+n}\left(s^{t+n-1}\right),s_{t+n}\right)p_\pi(s^{t+n}|h_t) \\
= \;& \sum_{(k_1,...,k_n):\prod_{i=0}^n \pi(k_i)\neq 0} r(k_0,...,k_n)\,p_\pi(K_{k_0,...,k_n}|h_t) \\
= \;& \sum_{(k_1,...,k_n):\prod_{i=0}^n \pi(k_i)\neq 0} r(k_0,...,k_n)\,p_\pi(K'_{k_0,...,k_n}|h_{t'}) \\
= \;& \sum_{s^{t'+n}:\mathbf{h}^\alpha_{t'}(s^{t'+n})=h_{t'}} r\left(\mathbf{a}^\alpha_{t'+n}\left(s^{t'+n-1}\right),s_{t'+n}\right)p_\pi(s^{t'+n}|h_{t'}).
\end{aligned}$$

Finally, since we have $\nu(\cdot \mid h_t) = \nu(\cdot \mid h_{t'})$, this implies that:

$$\begin{aligned}
& \phi^{-1}\left(\int_{\operatorname{supp}\nu(\cdot|h_t)}\phi\left(\sum_{s^{t+n}:\mathbf{h}^\alpha_t(s^{t-1})=h_t} r\left(\mathbf{a}^\alpha_{t+n}\left(s^{t+n-1}\right),s_{t+n}\right)p_\pi(s^{t+n}|h_t)\right)\nu\left(\mathrm{d}\pi \mid h_t\right)\right) \\
= \;& \phi^{-1}\left(\int_{\operatorname{supp}\nu(\cdot|h_{t'})}\phi\left(\sum_{s^{t'+n}:\mathbf{h}^\alpha_{t'}(s^{t'-1})=h_{t'}} r\left(\mathbf{a}^\alpha_{t'+n}\left(s^{t'+n-1}\right),s_{t'+n}\right)p_\pi(s^{t'+n}|h_{t'})\right)\nu\left(\mathrm{d}\pi \mid h_{t'}\right)\right)
\end{aligned}$$

and the thesis follows. ■

Proof of Proposition 2 By Proposition 1, we know that there exist $G \in \sigma(\mathbf{h}^{\alpha})$ such that $\overline{p}^{\alpha}(G) = 1$ and for all $s^{\infty} \in G$,

$$\lim_{t\to\infty} \mu(\hat{P}^{\alpha,\mu}_{T}(\overline{p}) \left(s^{T-1}\right) |\mathbf{h}^{\alpha}_{t+1} \left(s^{t}\right)) = 1. \tag{19}$$

Define G' as $G \cap \bigcap_{n\in\mathbb{N}} \{s^{\infty} \in S^{\infty} : \ \overline{p}^{\alpha}(\iota^{\alpha}_{n} \left(s^{n-1}\right)) > 0\}$. Note that, by definition of $\sigma(\mathbf{h}^{\alpha}_{n})$ and $\sigma(\mathbf{h}^{\alpha})$, it holds that $\{s^{\infty} \in G : \ \overline{p}^{\alpha}(\iota^{\alpha}_{n} \left(s^{n-1}\right)) > 0\} \in \sigma(\mathbf{h}^{\alpha}_{n}) \subseteq \sigma(\mathbf{h}^{\alpha})$ for every $n \in \mathbb{N}$. Moreover, it is immediate to see that:

$$\overline{p}^{\alpha} \left(s^{\infty} \in G : \ \overline{p}^{\alpha}(\iota^{\alpha}_{n} \left(s^{n-1}\right)) > 0\right) = 1.$$

Hence, $\overline{p}^{\alpha}(G') = 1$. Fix any path $s^{\infty} \in G'$. Since $s^{\infty} \in G' \subseteq G$, we have that (19) holds. Therefore, for every $\varepsilon > 0$, there exists $t_{\varepsilon,s^{\infty}} > T$ such that, for every $t > t_{\varepsilon,s^{\infty}}$

$$\mu(\hat{P}^{\alpha,\mu}_{T}(\overline{p}) \left(s^{T-1}\right) |\mathbf{h}^{\alpha}_{t+1} \left(s^{t}\right)) > 1 - \varepsilon.$$

We have that

$$\begin{aligned} & \mu_{\mathbf{h}^{\alpha}_{t+1}(s_t)} \left(p \in \Delta(S^{\infty}) : p^{\alpha_{\mathbf{h}^{\alpha}_{t+1}(s^t)}} = \overline{p}^{\alpha_{\mathbf{h}^{\alpha}_{t+1}(s^t)}}\right) \\ = \ & \mu \left(\hat{P}^{\alpha,\mu}_{t}(\overline{p}) \left(s^{t-1}\right) |\mathbf{h}^{\alpha}_{t+1} \left(s^{t}\right)\right) \\ \geq \ & \mu(\hat{P}^{\alpha,\mu}_{T}(\overline{p}) \left(s^{T-1}\right) |\mathbf{h}^{\alpha}_{t+1} \left(s^{t}\right)) > 1 - \varepsilon, \end{aligned}$$

where the equality holds by definition of $\hat{P}^{\alpha,\mu}_{t}(\overline{p}) \left(s^{t-1}\right)$, and the weak inequality follows from Lemma 1. Therefore the first condition of ε-self-confirming equilibrium holds. For the second one, let h_{τ} be a generic information history such that $p_{\mu_{\mathbf{h}^{\alpha}_{t+1}(s^t)}}(I(h_{\tau})) > 0$. By definition of G', we have $p_{\mu}(\iota^{\alpha}_{t} \left(s^{t-1}\right)) > 0$, and so:

$$p_{\mu}(\iota^{\alpha}_{t+1} \left(s^{t}\right) \cap I(h_{\tau})) = p_{\mu}(\iota^{\alpha}_{t+1} \left(s^{t}\right)) p_{\mu_{\mathbf{h}^{\alpha}_{t+1}(s^t)}}(I(h_{\tau})) > 0.$$

Next note that, for every a in A,

$$\begin{aligned} & V \left(\alpha_{\mathbf{h}^{\alpha}_{t+1}(\bar{s}_t)}, \mu_{\mathbf{h}^{\alpha}_{t+1}(\bar{s}_t)} \mid h_{\tau}\right) = V \left(\alpha, \mu \mid \left(\mathbf{h}^{\alpha}_{t+1} \left(s^{t}\right), h_{\tau}\right)\right) \\ \geq \ & V \left(\alpha / \left(\left(\mathbf{h}^{\alpha}_{t+1} \left(s^{t}\right), h_{\tau}\right), a\right), \mu \mid \left(\mathbf{h}^{\alpha}_{t+1} \left(s^{t}\right), h_{\tau}\right)\right) \\ = \ & V \left(\alpha_{\mathbf{h}^{\alpha}_{t+1}(s^t)} / \left(h_t, a\right), \mu_{\mathbf{h}^{\alpha}_{t+1}(s^t)} \mid h_{\tau}\right), \end{aligned}$$

and therefore also the second condition holds. In other words, for every $\varepsilon > 0$, there exists $t_{\varepsilon,s^{\infty}} > T$ such that the reinitialized triple is an ε-self-confirming equilibrium. Since the result holds for every $s^{\infty} \in G'$ and $p^{\alpha}(G') = \overline{p}^{\alpha}(G') = 1$, we have proved the statement. ■

Proof of Proposition 3 Let G' be as in the proof of Proposition 2. Choose arbitrarily $s^\infty \in G'$ and $\varepsilon' > 0$; Proposition 1 yields the existence of $t_{\varepsilon', s^\infty} > T$ such that for every $t > t_{\varepsilon', s^\infty}$ it holds that

$$\mu\left(\hat{P}_T^{\alpha,\mu}(\bar{p})(s^\infty)|\mathbf{h}_{t+1}^\alpha(s^t)\right) \geq 1-\varepsilon$$

Therefore, for every $t > t_{\varepsilon', s^\infty}$ and for every $E \in \sigma(\mathbf{h}^\alpha)$

$$\begin{aligned}
& |p_{\mu\left(\cdot|\mathbf{h}_{t+1}^\alpha(s^t)\right)}\left(E|\mathbf{h}_{t+1}^\alpha(s^t)\right) - \bar{p}^\alpha\left(E|\mathbf{h}_{t+1}^\alpha(s^t)\right)| \\
= & \left|\int_{\Delta(S^\infty)} p(E|\mathbf{h}_{t+1}^\alpha(s^t))\mu(\mathrm{d}p|\mathbf{h}_{t+1}^\alpha(s^t)) - \bar{p}^\alpha\left(E|\mathbf{h}_{t+1}^\alpha(s^t)\right)\right| \\
= & \left|\int_{\hat{P}_T^{\alpha,\mu}(\bar{p})(s^\infty)} p(E|\mathbf{h}_{t+1}^\alpha(s^t))\mu(\mathrm{d}p|\mathbf{h}_{t+1}^\alpha(s^t))\right. \\
& \left. + \int_{\Delta(S^\infty)/\hat{P}_T^{\alpha,\mu}(\bar{p})(s^\infty)} p(E|\mathbf{h}_{t+1}^\alpha(s^t))\mu(\mathrm{d}p|\mathbf{h}_{t+1}^\alpha(s^t)) - \bar{p}^\alpha\left(E|\mathbf{h}_{t+1}^\alpha(s^t)\right)\right| \\
= & \left|\mu(\hat{P}_T^{\alpha,\mu}(\bar{p})(s^\infty)|\mathbf{h}_{t+1}^\alpha(s^t))\bar{p}^\alpha\left(E|\mathbf{h}_{t+1}^\alpha(s^t)\right) - \bar{p}^\alpha\left(E|\mathbf{h}_{t+1}^\alpha(s^t)\right)\right. \\
& \left. + \int_{\Delta(S^\infty)/\hat{P}_T^{\alpha,\mu}(\bar{p})(s^\infty)} p(E|\mathbf{h}_{t+1}^\alpha(s^t))\mu(\mathrm{d}p|\mathbf{h}_{t+1}^\alpha(s^t))\right| \\
= & \left|\int_{\Delta(S^\infty)/\hat{P}_T^{\alpha,\mu}(\bar{p})(s^\infty)} p(E|\mathbf{h}_{t+1}^\alpha(s^t))\mu(\mathrm{d}p|\mathbf{h}_{t+1}^\alpha(s^t))\right. \\
& \left. - \mu\left(\Delta(S^\infty)/\hat{P}_T^{\alpha,\mu}(\bar{p})(s^\infty)|\mathbf{h}_{t+1}^\alpha(s^t)\right)\bar{p}^\alpha\left(E|\mathbf{h}_{t+1}^\alpha(s^t)\right)\right| \\
\leq & \ \varepsilon'.
\end{aligned}$$

Summing up, for every $t > t_{\varepsilon', s^\infty}$

$$\forall E \in \sigma(\mathbf{h}^\alpha) \ \ |p_{\mu\left(\cdot|\mathbf{h}_{t+1}^\alpha(s^t)\right)}\left(E|\mathbf{h}_{t+1}^\alpha(s^t)\right) - \bar{p}^\alpha\left(E|\mathbf{h}_{t+1}^\alpha(s^t)\right)| < \varepsilon'. \tag{20}$$

Moreover, by Proposition 2 of Kalai and Lehrer (1994), we know that for every $\varepsilon > 0$, there exists $\varepsilon' > 0$ such that (20) implies (12). Therefore, we obtain the desired result. ■

Proof of Proposition 4 Since A and S are finite, there exists K such that, for every $a, a' \in A$, for every $s, s' \in S$,

$$|u(a,s) - u(a',s')| \leq K.$$

Moreover, for every $\varepsilon > 0$, there exists $n \in \mathbb{N}$ such that

$$\frac{\delta^n}{1-\delta}K < \varepsilon/2.$$

Let G' be as in the proof of Proposition 2; for every $\bar{s}^\infty$ in G', for every $t \in \mathbb{N}$, we have that

$$\begin{aligned}
& \left| \begin{array}{c} \sum_{\tau=t+n}^{\infty} \delta^{\tau-t} \phi^{-1} \left(\int_{\Delta(S^\infty)} \phi \left(\sum_{s^\tau \in S^\tau} r \left(\mathbf{a}_\tau^\alpha \left(s^{\tau-1} \right), s_\tau \right) p \left(s^\tau \mid \mathbf{h}_t^\alpha \left(\bar{s}^{t-1} \right) \right) \right) \mu \left(\mathrm{d}p \mid \mathbf{h}_t^\alpha \left(\bar{s}^{t-1} \right) \right) \right) \\ - \sum_{\tau=t+n}^{\infty} \delta^{\tau-t} \sum_{s^\tau \in S^\tau} r \left(\mathbf{a}_\tau^\alpha \left(s^{\tau-1} \right), s_\tau \right) \bar{p} \left(s^\tau \mid \mathbf{h}_t^\alpha \left(\bar{s}^{t-1} \right) \right) \end{array} \right| \\
= & \left| \sum_{\tau=t+n}^{\infty} \delta^{\tau-t} \left(\begin{array}{c} \phi^{-1} \left(\int_{\Delta(S^\infty)} \phi \left(\sum_{s^\tau \in S^\tau} r \left(\mathbf{a}_\tau^\alpha \left(s^{\tau-1} \right), s_\tau \right) p \left(s^\tau \mid \mathbf{h}_t^\alpha \left(\bar{s}^{t-1} \right) \right) \right) \mu \left(\mathrm{d}p \mid \mathbf{h}_t^\alpha \left(\bar{s}^{t-1} \right) \right) \right) \\ - \sum_{s^\tau \in S^\tau} r \left(\mathbf{a}_\tau^\alpha \left(s^{\tau-1} \right), s_\tau \right) \bar{p} \left(s^\tau \mid \mathbf{h}_t^\alpha \left(\bar{s}^{t-1} \right) \right) \end{array} \right) \right| \\
< & \sum_{\tau=t+n}^{\infty} \delta^{\tau-t} K = \frac{\delta^n}{1-\delta} K < \varepsilon/2.
\end{aligned}$$

Now, consider the process $v_t(s^\infty) = r\left(\mathbf{a}_{t+k}^\alpha\left(s^{t+k-1}\right), s_{t+k}\right)$, with $k \in \{0, ..., n\}$. Again, since the spaces of actions and states are finite, this process is uniformly bounded. Moreover, from observable consequences (3) and since $r = u \circ \rho$, v_t is $\sigma(\mathbf{h}_{t+k+1}^\alpha)$-measurable. Since, by definition, $\sigma(\mathbf{h}^\alpha) = \sigma\left(\cup_t \sigma\left(\mathbf{h}_t^\alpha\right)\right)$, we have that for every t, v_t is also $\sigma(\mathbf{h}^\alpha)$-measurable. Hence, we can apply Corollary 4 to v_t, and we obtain that, for every $\varepsilon > 0$, there exists $t_{\varepsilon,k}$, for every $\tau \geq t_{\varepsilon,k}$,

$$\begin{aligned}
& \left| \begin{array}{c} \phi^{-1} \left(\int_{\Delta(S^\infty)} \phi \left(\sum_{s^{\tau+k} \in S^{\tau+k}} r \left(\mathbf{a}_{\tau+k}^\alpha \left(s^{\tau+k-1} \right), s_{\tau+k} \right) p \left(s^{\tau+k} \mid \mathbf{h}_t^\alpha \left(\bar{s}^{t-1} \right) \right) \right) \mu \left(\mathrm{d}p \mid \mathbf{h}_t^\alpha \left(\bar{s}^{t-1} \right) \right) \right) \\ - \sum_{s^{\tau+k} \in S^{\tau+k}} r \left(\mathbf{a}_{\tau+k}^\alpha \left(s^{\tau+k-1} \right), s_{\tau+k} \right) \bar{p} \left(s^{\tau+k} \mid \mathbf{h}_t^\alpha \left(\bar{s}^{t-1} \right) \right) \end{array} \right| \\
< & \frac{\varepsilon}{2n}.
\end{aligned}$$

Let $t_\varepsilon^* := \max_{k \in \{0,...,n\}} t_{\varepsilon,k}$. We have that, for every $t > t_\varepsilon^*$

$$\begin{aligned}
& \left| \begin{array}{c} V\left(\alpha, \mu | \mathbf{h}_t^\alpha \left(\bar{s}^{t-1} \right) \right) \\ - \sum_{\tau=t}^{\infty} \delta^{\tau-t} \sum_{s^\tau \in S^\tau} r \left(\mathbf{a}_\tau^\alpha \left(s^{\tau-1} \right), s_\tau \right) \bar{p} \left(s^\tau \mid \mathbf{h}_t^\alpha \left(\bar{s}^{t-1} \right) \right) \end{array} \right| \\
= & \left| \sum_{\tau=t}^{\infty} \delta^{\tau-t} \left(\begin{array}{c} \phi^{-1} \left(\int_{\Delta(S^\infty)} \phi \left(\sum_{s^\tau \in S^\tau} r \left(\mathbf{a}_\tau^\alpha \left(s^{\tau-1} \right), s_\tau \right) p \left(s^\tau \mid \mathbf{h}_t^\alpha \left(\bar{s}^{t-1} \right) \right) \right) \mu \left(\mathrm{d}p \mid \mathbf{h}_t^\alpha \left(\bar{s}^{t-1} \right) \right) \right) \\ - \sum_{s^\tau \in S^\tau} r \left(\mathbf{a}_\tau^\alpha \left(s^{\tau-1} \right), s_\tau \right) \bar{p} \left(s^\tau \mid \mathbf{h}_t^\alpha \left(\bar{s}^{t-1} \right) \right) \end{array} \right) \right| \\
\leq & \left| \sum_{\tau=t}^{t+n-1} \delta^{\tau-t} \left(\begin{array}{c} \phi^{-1} \left(\int_{\Delta(S^\infty)} \phi \left(\sum_{s^\tau \in S^\tau} r \left(\mathbf{a}_\tau^\alpha \left(s^{\tau-1} \right), s_\tau \right) p \left(s^\tau \mid \mathbf{h}_t^\alpha \left(\bar{s}^{t-1} \right) \right) \right) \mu \left(\mathrm{d}p \mid \mathbf{h}_t^\alpha \left(\bar{s}^{t-1} \right) \right) \right) \\ - \sum_{s^\tau \in S^\tau} r \left(\mathbf{a}_\tau^\alpha \left(s^{\tau-1} \right), s_\tau \right) \bar{p} \left(s^\tau \mid \mathbf{h}_t^\alpha \left(\bar{s}^{t-1} \right) \right) \end{array} \right) \right| \\
& + \left| \sum_{\tau=t+n}^{\infty} \delta^{\tau-t} \left(\begin{array}{c} \phi^{-1} \left(\int_{\Delta(S^\infty)} \phi \left(\sum_{s^\tau \in S^\tau} r \left(\mathbf{a}_\tau^\alpha \left(s^{\tau-1} \right), s_\tau \right) p \left(s^\tau \mid \mathbf{h}_t^\alpha \left(s^{t-1} \right) \right) \right) \mu \left(\mathrm{d}p \mid \mathbf{h}_t^\alpha \left(s^{t-1} \right) \right) \right) \\ - \sum_{s^\tau \in S^\tau} r \left(\mathbf{a}_\tau^\alpha \left(s^{\tau-1} \right), s_\tau \right) \bar{p} \left(s^\tau \mid \mathbf{h}_t^\alpha \left(s^{t-1} \right) \right) \end{array} \right) \right|
\end{aligned}$$

$$
\begin{aligned}
&\leq \left| \sum_{\tau=t}^{t+n-1} \delta^{\tau-t} \left(\begin{array}{c} \phi^{-1}\left(\int_{\Delta(S^{\infty})} \phi\left(\sum_{s^{\tau}\in S^{\tau}} r\left(\mathbf{a}_{\tau}^{\alpha}\left(s^{\tau-1}\right), s_{\tau}\right) p\left(s^{\tau} \mid \mathbf{h}_{t}^{\alpha}\left(\bar{s}^{t-1}\right)\right)\right) \mu\left(\mathrm{d}p \mid \mathbf{h}_{t}^{\alpha}\left(\bar{s}^{t-1}\right)\right)\right) \\ -\sum_{s^{\tau}\in S^{\tau}} r\left(\mathbf{a}_{\tau}^{\alpha}\left(s^{\tau-1}\right), s_{\tau}\right) \bar{p}\left(s^{\tau} \mid \mathbf{h}_{t}^{\alpha}\left(\bar{s}^{t-1}\right)\right) \end{array} \right) \right| \\
&\quad +\frac{\varepsilon}{2} \\
&\leq \sum_{\tau=t}^{t+n-1} \left| \delta^{\tau-t} \left(\begin{array}{c} \phi^{-1}\left(\int_{\Delta(S^{\infty})} \phi\left(\sum_{s^{\tau}\in S^{\tau}} r\left(\mathbf{a}_{\tau}^{\alpha}\left(s^{\tau-1}\right), s_{\tau}\right) p\left(s^{\tau} \mid \mathbf{h}_{t}^{\alpha}\left(\bar{s}^{t-1}\right)\right)\right) \mu\left(\mathrm{d}p \mid \mathbf{h}_{t}^{\alpha}\left(\bar{s}^{t-1}\right)\right)\right) \\ -\sum_{s^{\tau}\in S^{\tau}} r\left(\mathbf{a}_{\tau}^{\alpha}\left(s^{\tau-1}\right), s_{\tau}\right) \bar{p}\left(s^{\tau} \mid \mathbf{h}_{t}^{\alpha}\left(\bar{s}^{t-1}\right)\right) \end{array} \right) \right| \\
&\quad +\frac{\varepsilon}{2} \\
&\leq \sum_{\tau=t}^{t+n-1} \left| \delta^{\tau-t} \left(\frac{\varepsilon}{2n}\right) \right| + \frac{\varepsilon}{2} < \varepsilon.
\end{aligned}
$$

This, by definition of limit implies that

$$
\lim_{t\to\infty} \left| \begin{array}{c} V\left(\alpha, \mu | \mathbf{h}_{t}^{\alpha}\left(\bar{s}^{t-1}\right)\right) \\ -\sum_{\tau=t}^{\infty} \delta^{\tau-t} \sum_{s^{\tau}\in S^{\tau}} r\left(\mathbf{a}_{\tau}^{\alpha}\left(s^{\tau-1}\right), s_{\tau}\right) \bar{p}\left(s^{\tau} \mid \mathbf{h}_{t}^{\alpha}\left(\bar{s}^{t-1}\right)\right) \end{array} \right| = 0.
$$

Summing up the subjective continuation value of strategy α converges to the objective one almost surely. ■

In order to prove Proposition 5, for the sake of completeness, we show that Bayesian beliefs satisfy the Martingale property.

Denote the period-$(t+1)$ Bayesian map $\beta_{t+1} : \Delta(\Delta(S)) \times A \to \Delta_0(\Delta(S))$, and define it in the the following way for every $B \in \mathcal{B}(\cdot(S))$

$$
\beta_{t+1}(\nu^t(\cdot), a)(B) := \begin{cases} \frac{\int_B q(f_a^{-1}(m_1))\nu^t(\mathrm{d}\pi)}{p_{\nu^t}(f_a^{-1}(m_1))} & \text{with probability} \quad \int_P \widehat{\pi}(f_a^{-1}(m_1))\nu^t(\mathrm{d}\widehat{\pi}) \\ \ldots & \ldots \qquad\qquad \ldots \\ \frac{\int_B q(f_a^{-1}(m_{|M|}))\nu^t(\mathrm{d}\pi)}{p_{\nu^t}(f_a^{-1}(m_{|M|}))} & \text{with probability} \quad \int_P \widehat{\pi}(f_a^{-1}(m_{|M|}))\nu^t(\mathrm{d}\widehat{\pi}). \end{cases}
$$

The distribution $\beta_{t+1}(\nu, a,)$ is the distribution over period $t+1$ beliefs consistent with holding belief ν and taking action a at time t.

Lemma 4. *In an i.i.d. environment, for every Borel subset B in $\mathcal{B}(\Delta(S))$,*

$$
\mathbb{E}_{\beta(\nu^t(\cdot),a)}(\nu^{t+1}(B)) = \nu^t(B).
$$

Proof of Lemma 4 Note that the subjective probability assigned to the models in B after having observed m and having played a is

$$
\frac{\int_B q(f_a^{-1}(m))\nu^t(\mathrm{d}\pi)}{p_{\nu^t}(f_a^{-1}(m))},
$$

whereas the subjective probability of observing message m when a is played is

$$
\int_{\Delta(S)} \widehat{\pi}(f_a^{-1}(m))\nu^t(\mathrm{d}\widehat{\pi}).
$$

By definition of Bayesian map, we have:

$$\begin{aligned}
\mathbb{E}_{\beta(\nu(\cdot),a)}(\nu^{t+1}(B)) &= \sum_{m:p_{\nu^t}(f_a^{-1}(m))>0} \int_P \left(\frac{\int_B q(f_a^{-1}(m))\nu^t(\mathrm{d}\pi)}{p_{\nu^t}(f_a^{-1}(m))} \right) \widehat{\pi}(f_a^{-1}(m))\nu^t(\mathrm{d}\widehat{\pi}) \\
&= \sum_{m:p_{\nu^t}(f_a^{-1}(m))>0} \frac{\int_B \pi(f_a^{-1}(m))\nu^t(\mathrm{d}\pi)}{p_{\nu^t}(f_a^{-1}(m))} p_{\nu^t}(f_a^{-1}(m)) \\
&= \sum_{m:p_{\nu^t}(f_a^{-1}(m))>0} \int_B \pi(f_a^{-1}(m))\nu^t(\mathrm{d}\pi) \\
&= \int_B \sum_{m:p_{\nu^t}(f_a^{-1}(m))>0} \pi(f_a^{-1}(m))\nu^t(\mathrm{d}\pi) \\
&= \nu^t(B)
\end{aligned}$$

Note that the result holds also if $\nu^t(B) = 0$. ■

Proof of Proposition 5 It is immediate to see that condition (i) of SCE implies condition (ii) of static SCE. Now, we show that an SCE features myopic best reply on path, that is, $(\alpha(\nu), \nu, \overline{\pi})$ satisfies property (i) of static SCE. By way of contradiction, suppose there is $a \in A$ such that

$$\phi^{-1}\left(\int_{\Delta(S)} \phi\left(\sum_{s\in S} r(a,s)\pi(s)\right)\nu(\mathrm{d}\pi)\right) > \phi^{-1}\left(\int_{\Delta(S)} \phi\left(\sum_{s\in S} r(\alpha(\nu),s)\pi(s)\right)\nu(\mathrm{d}\pi)\right).$$

By definition of SCE it must be the case that $V(\alpha/a, \nu) \leq V(\alpha, \nu)$. However, we have:

$$
\begin{aligned}
& V(\alpha, \nu) \\
= \; & \sum_{s\in S} r(\alpha(\nu), s)\overline{\pi}(s) + \delta V(\alpha, \nu) \\
\leq \; & \sum_{s\in S} r(\alpha(\nu), s)\overline{\pi}(s) + \delta \min_{m:\pi_\nu(f_a^{-1}(m))>0} V(\alpha, \nu(\cdot|(a, m)) \\
< \; & \phi^{-1}\left(\int_{\Delta(S)} \phi\left(\sum_{s\in S} r(a, s)\pi(s)\right) \nu(\mathrm{d}\pi)\right) + \\
& \delta \min_{m:\pi_\nu(f_a^{-1}(m))>0} \left(\sum_{\tau=t}^{\infty} \delta^{\tau-t}\phi^{-1}\left(\int_{\Delta(S)} \phi\left(\sum_{s^\tau\in S^\tau} r\left(\mathbf{a}_\tau^\alpha\left(s^{\tau-1}\right), s_\tau\right) p_\pi(s^\tau|h_t)\right) \nu\left(\mathrm{d}\pi \mid (a, m)\right)\right)\right) \\
= \; & \phi^{-1}\left(\int_{\Delta(S)} \phi\left(\sum_{s\in S} r(a, s)\pi(s)\right) \nu(\mathrm{d}\pi)\right) + \\
& \delta \left(\sum_{\tau=t}^{\infty} \delta^{\tau-t}\phi^{-1}\left(\min_{m:\pi_\nu(f_a^{-1}(m))>0}\left(\int_{\Delta(S)} \phi\left(\sum_{s^\tau\in S^\tau} r\left(\mathbf{a}_\tau^\alpha\left(s^{\tau-1}\right), s_\tau\right) p_\pi(s^\tau|h_t)\right) \nu\left(\mathrm{d}\pi \mid (a, m)\right)\right)\right)\right) \\
\leq \; & \phi^{-1}\left(\int_{\Delta(S)} \phi\left(\sum_{s\in S} r(a, s)\pi(s)\right) \nu(\mathrm{d}\pi)\right) + \\
& \delta \left(\sum_{\tau=t}^{\infty} \delta^{\tau-t}\phi^{-1}\left(\int_{\Delta(S)} \phi\left(\sum_{s^\tau\in S^\tau} r\left(\mathbf{a}_\tau^\alpha\left(s^{\tau-1}\right), s_\tau\right) p_\pi(s^\tau|h_t)\right) \mathbb{E}_{\beta(\nu(\cdot),a)}(\nu(\mathrm{d}\pi|(a, m)))\right)\right) \\
= \; & \phi^{-1}\left(\int_{\Delta(S)} \phi\left(\sum_{s\in S} r(a, s)\pi(s)\right) \nu(\mathrm{d}\pi)\right) + \\
& \delta \left(\sum_{\tau=t}^{\infty} \delta^{\tau-t}\phi^{-1}\left(\int_{\Delta(S)} \phi\left(\sum_{s^\tau\in S^\tau} r\left(\mathbf{a}_\tau^\alpha\left(s^{\tau-1}\right), s_\tau\right) p_\pi(s^\tau|h_t)\right) \nu\left(\mathrm{d}\pi\right)\right)\right) \\
= \; & \sum_{\tau=t}^{\infty} \delta^{\tau-t}\phi^{-1}\left(\int_{\Delta(S)} \phi\left(\sum_{s^\tau\in S^\tau} r\left(\mathbf{a}_\tau^{\alpha/a}\left(s^{\tau-1}\right), s_\tau\right) p_\pi(s^\tau|h_t)\right) \nu\left(\mathrm{d}\pi\right)\right) \\
= \; & V(\alpha/a, \nu),
\end{aligned}
$$

where the first equality comes from by the confirmed beliefs property of SCE, the strict inequality comes from hypothesis, the second equality comes from the fact that ϕ is strictly increasing, the third equality by Lemma 4**,** and the fourth and fifth equalities by the definition of α/a. Note that we will be done as soon as we prove the first weak inequality, that is

$$V(\alpha, \nu) \leq \min_{m:\pi_\nu(f_a^{-1}(m))>0} V(\alpha, \nu(\cdot|(a, m)).$$

Indeed, it would follow that $V(\alpha, \nu) < V(\alpha/a, \nu)$, a contradiction with the fact that $(\alpha, \nu\,, \overline{\pi})$ is an SCE.

Suppose that there exist m such that $\pi_\nu(f_a^{-1}(m)) > 0$ with $V(\alpha, \nu(\cdot|(a, m)) <$

$V(\alpha,\nu)$. The fact that $\pi_\nu(f_a^{-1}(m)) > 0$ implies that $\pi_\nu(I(a,m)) > 0$. On the other hand, since $(\alpha,\nu,\overline{\pi})$ is an SCE, $\nu\left(\pi \in \Delta(S) : \pi^\alpha = \overline{\pi}^\alpha\right) = 1$, and in particular

$$\nu\left(\pi \in \Delta(S) : \pi \circ f_{\alpha(\nu)}^{-1} = \overline{\pi} \circ f_{\alpha(\nu)}^{-1}\right) = 1.$$

Then, let $B = \left\{\pi \in \Delta(S) : \pi \circ f_{\alpha(\nu)}^{-1} = \overline{\pi} \circ f_{\alpha(\nu)}^{-1}\right\}$. By Bayes rule, we have that

$$\nu\left(B|(a,m)\right) = \nu\left(B\right) \frac{\int_B \pi(f_a^{-1}(m))\nu(\mathrm{d}\pi)}{\pi_\nu(f_a^{-1}(m))} = \nu\left(B\right) \frac{\int_{\Delta(S)} \pi(f_a^{-1}(m))\nu(\mathrm{d}\pi)}{\pi_\nu(f_a^{-1}(m))} = \nu\left(B\right) = 1.$$

But then it follows that

$$\begin{aligned}
V(\alpha,\nu(\cdot|(a,m)) &< (1-\delta)V(\alpha,\nu) + \delta V(\alpha,\nu(\cdot|(a,m)) \\
&= \phi^{-1}\left(\int_{\Delta(S)} \phi\left(\sum_{s\in S} r(\alpha(\nu),s)\pi(s)\right)\nu(\mathrm{d}\pi)\right) + \delta V(\alpha,\nu(\cdot|(a,m)) \\
&= \phi^{-1}\left(\int_{\Delta(S)} \phi\left(\sum_{s\in S} r(\alpha(\nu),s)\pi(s)\right)\nu(\mathrm{d}\pi|(a,m))\right) + \delta V(\alpha,\nu(\cdot|(a,m)) \\
&= \sum_{\tau=t}^{\infty} \delta^{\tau-t}\phi^{-1}\left(\int_{\Delta(S)} \phi\left(\sum_{s^\tau \in S^\tau} r\left(\mathbf{a}_\tau^{\alpha/\alpha(\nu)}\left(s^{\tau-1}\right), s_\tau\right) p_\pi(s^\tau|h_t)\right)\nu\left(\mathrm{d}\pi|(a,m)\right)\right) \\
&= V(\alpha/(\alpha(\nu),\nu(\cdot|(a,m)).
\end{aligned}$$

But this contradicts the fact that $(\alpha,\nu\,,\overline{\pi})$ is an SCE. ■

Proof of Lemma 3 Since the assumptions of Proposition 2 are satisfied, we can consider G' as defined in the corresponding proof. Fix any path $s^\infty \in G'$. By definition of $\hat{P}_T^{\alpha,\mu}(\overline{p})\left(s^{T-1}\right)$, we have that for every t larger than T and, for every p'_π,p''_π in $\hat{P}_T^{\alpha,\mu}(p_{\overline{\pi}})\left(s^{T-1}\right)$

$$p'_\pi(\mathbf{h}_t^\alpha\left(s^{t-1}\right)|\mathbf{h}_T^\alpha\left(s^{T-1}\right)) = p''_\pi(\mathbf{h}_t^\alpha\left(s^{t-1}\right)|\mathbf{h}_T^\alpha\left(s^{T-1}\right)).$$

Hence, since property (i) of pre self-confirming equilibria implies that the agent does not reach information histories that are subjectively unreachable. It follows from the chain rule that:

$$\frac{\nu(\pi'|\mathbf{h}_t^\alpha\left(s^{t-1}\right))}{\nu(\pi''|\mathbf{h}_t^\alpha\left(s^{t-1}\right))} = \frac{\nu(\pi'|\mathbf{h}_T^\alpha\left(s^{T-1}\right))}{\nu(\pi''|\mathbf{h}_T^\alpha\left(s^{T-1}\right))}.$$

This assures that

$$\nu_{s^\infty}^\alpha := \lim_{t\to\infty} \nu(\cdot|\mathbf{h}_t^\alpha\left(s^{t-1}\right))$$

is well defined.

Indeed, the relative probabilities of the models in $\hat{P}_T^{\alpha,\mu}(p_{\overline{\pi}})\left(s^{T-1}\right)$ remain constant and (19) holds, then:

$$\nu_{s^\infty}^\alpha(\pi) = \begin{cases} \frac{\nu(\pi|\mathbf{h}_T^\alpha\left(s^{T-1}\right))}{\nu(\hat{P}_T^{\alpha,\mu}(p_{\overline{\pi}})(s^{T-1})|\mathbf{h}_T^\alpha(s^{T-1}))} & \text{if } p_\pi \in \hat{P}_T^{\alpha,\mu}(p_{\overline{\pi}})\left(s^{T-1}\right), \\ 0 & \text{if } p_\pi \notin \hat{P}_T^{\alpha,\mu}(p_{\overline{\pi}})\left(s^{T-1}\right). \end{cases}$$

and the thesis follows. ■

Proof of Proposition 6 First, we have that the hypothesis of Proposition 2 are satisfied, so let G' be as in the corresponding proof. Second, note that the value function (10) is continuos in beliefs ν. Then, by Proposition 3**,** for every a in A,

$$\lim_{t\to\infty} V(\alpha/a, \nu(\cdot|\mathbf{h}_t^\alpha\left(s^{t-1}\right))) = V(\alpha/a, \nu_{s^\infty}^\alpha),$$

where α/a is the strategy that prescribes a in the first period to come and coincide with α otherwise.

Let $A_\infty := \arg\max_{a\in A} V(\alpha/a, \nu_{s^\infty}^\alpha)$. Note that, in general, the definition of pre self-confirming equilibrium does not require that

$$\alpha(\nu_{s^\infty}^\alpha) \in \arg\max_{a\in A} V(\alpha/a, \nu_{s^\infty}^\alpha).$$

Indeed, if there is no h_t such that $p_{\pi_\nu}(I(h_t)) > 0$ and $\nu(\cdot|h_t) = \nu_{s^\infty}^\alpha$, then $\alpha\left(\nu_{s^\infty}^\alpha\right)$ does not need to satisfy the one-shot-deviation property. Since $s^\infty \in G'$, then it follows by the definition of G', that $p_{\overline{\pi}}(I(\mathbf{h}_t^\alpha\left(s^{t-1}\right))) > 0$ for every finite t. By property (i) of pre self-confirming equilibria we have that $p_{\overline{\pi}}(I(\mathbf{h}_t^\alpha\left(s^{t-1}\right))) > 0$ implies $p_{\pi_\nu}(I(\mathbf{h}_t^\alpha\left(s^{t-1}\right))) > 0$. Hence,

$$\alpha(\nu(\cdot|\mathbf{h}_t^\alpha\left(s^{t-1}\right))) \in \arg\max_{a\in A} V(\alpha/a, \nu(\cdot|\mathbf{h}_t^\alpha\left(s^{t-1}\right))).$$

Now, let $a \notin A_\infty$, and fix $a^* \in A_\infty$. We have that

$$\begin{aligned}\lim_{t\to\infty} V(\alpha/a, \nu(\cdot|\mathbf{h}_t^\alpha\left(s^{t-1}\right))) &= V(\alpha/a, \nu_{s^\infty}^\alpha) \\ &< \max_{a'\in A} V(\alpha/a', \nu_{s^\infty}^\alpha) = V(\alpha/a^*, \nu_{s^\infty}^\alpha) \\ &= \lim_{t\to\infty} V(\alpha/a^*, \nu(\cdot|\mathbf{h}_t^\alpha\left(s^{t-1}\right))).\end{aligned}$$

Hence there exists$T_{s^\infty}^a$ such that $a \notin \alpha(\nu(\cdot|\mathbf{h}_t^\alpha\left(s^{t-1}\right)))$ for every $t \geq T_{s^\infty}^a$. Let $T_{s^\infty}^* = \max_{a\in A/A_\infty} T_{s^\infty}^a$. Then, from $T_{s^\infty}^*$ onward, the only actions played are in A_∞, that is, they satisfy the one-shot deviation property with respect to having the limit beliefs $\nu_{s^\infty}^\alpha$. Let $\widehat{T}_{s^\infty} = \max\{T, T_{s^\infty}^*\}$; we have that from $\widehat{T}_{s^\infty}$ onward the action prescribed by strategy α, $\mathbf{a}_t^\alpha(s^{t-1})$, satisfies the one-shot deviation property with respect to beliefs $\nu_{s^\infty}^\alpha$, and confirm them. By Proposition 5, this implies that $(\mathbf{a}_t^\alpha(s^{t-1}), \nu_{s^\infty}^\alpha, \overline{\pi})$ is a static SCE for every $t \geq \widehat{T}_{s^\infty}$. ■

Proof of Corollary 3 By hypothesis, we know that (α, ν, π) converges to a static SCE on s^∞. Therefore, there exists $\widehat{T}_{s^\infty}$ such that, for every $t \geq \widehat{T}_{s^\infty}$, the pair $(\mathbf{a}_t^\alpha(s^{t-1}), \nu_{s^\infty}^\alpha, \overline{\pi})$ is a static SCE, and so

$$\mathbf{a}_t^\alpha(s^{t-1}) \in \arg\max_{a\in A} \phi^{-1}\left(\int_{\Delta(S)} \phi\left(\sum_{s\in S} r(a,s)\pi(s)\right) \nu_{s^\infty}^\alpha(\mathrm{d}\pi)\right) = \left\{a_{s^\infty}^*\right\}.$$

Therefore, $(a^*_{s^\infty}, \nu^\alpha_{s^\infty}, \overline{\pi})$ is a static SCE. Now, let $a \neq a^*_{s^\infty}$. By continuity of the one-period value function with respect to beliefs, and by Proposition 3, it follows that

$$\begin{aligned}
& \lim_{t\to\infty} \phi^{-1}\left(\int_{\Delta(S)} \phi\left(\sum_{s\in S} r(a^*_{s^\infty}, s)\pi(s)\right) \nu(\mathrm{d}\pi|\mathbf{h}^\alpha_t\left(s^{t-1}\right))\right) \\
= \;& \phi^{-1}\left(\int_{\Delta(S)} \phi\left(\sum_{s\in S} r(a^*_{s^\infty}, s)\pi(s)\right) \nu^\alpha_{s^\infty}(\mathrm{d}\pi)\right) \\
> \;& \phi^{-1}\left(\int_{\Delta(S)} \phi\left(\sum_{s\in S} r(a, s)\pi(s)\right) \nu^\alpha_{s^\infty}(\mathrm{d}\pi)\right) \\
= \;& \lim_{t\to\infty} \phi^{-1}\left(\int_{\Delta(S)} \phi\left(\sum_{s\in S} r(a, s)\pi(s)\right) \nu(\mathrm{d}\pi|\mathbf{h}^\alpha_t\left(s^{t-1}\right))\right).
\end{aligned}$$

Therefore there exists $\bar{T}_{a,s^\infty} > \widehat{T}_{s^\infty}$ such that $t > \bar{T}_{a,s^\infty}$ implies

$$\begin{aligned}
& \phi^{-1}\left(\int_{\Delta(S)} \phi\left(\sum_{s\in S} r(a^*_{s^\infty}, s)\pi(s)\right) \nu(\mathrm{d}\pi|\mathbf{h}^\alpha_t\left(s^{t-1}\right))\right) \\
> \;& \phi^{-1}\left(\int_{\Delta(S)} \phi\left(\sum_{s\in S} r(a, s)\pi(s)\right) \nu(\mathrm{d}\pi|\mathbf{h}^\alpha_t\left(s^{t-1}\right))\right).
\end{aligned}$$

Let $\bar{T}_{s^\infty} = \max_{a\in A/a^*_{s^\infty}} \bar{T}_{a,s^\infty}$. We have that $t > \bar{T}_{s^\infty}$ implies

$$\begin{aligned}
& \phi^{-1}\left(\int_{\Delta(S)} \phi\left(\sum_{s\in S} r(a^*_{s^\infty}, s)\pi(s)\right) \nu(\mathrm{d}\pi|\mathbf{h}^\alpha_t\left(s^{t-1}\right))\right) \\
= \;& \max_{a\in A} \phi^{-1}\left(\int_{\Delta(S)} \phi\left(\sum_{s\in S} r(a, s)\pi(s)\right) \nu(\mathrm{d}\pi|\mathbf{h}^\alpha_t\left(s^{t-1}\right))\right).
\end{aligned}$$

And the thesis follows. ■

Proof of Proposition 7 Let

$$a_1(\nu) = \arg\max_{a\in A} \phi^{-1}\left(\int_{\Delta(S)} \phi\left(\sum_{s\in S} r(a, s)\pi(s)\right) \nu(\mathrm{d}\pi)\right)$$

be defined as the one-period best reply correspondence to the belief ν.

Since $(\alpha, \nu^*, \overline{\pi})$ is an SCE, we know by Proposition 5 that $\alpha(\nu^*) \in a_1(\nu^*)$. In what follows, let $\alpha(\nu)$ an arbitrary element of the image of ν under the correspondence a_1. By definition:

$$\begin{aligned}
V_1(\nu) \;&=\; \phi^{-1}\left(\int_{\Delta(S)} \phi\left(\sum_{s\in S} r(\alpha(\nu), s)\pi(s)\right) \nu(\mathrm{d}\pi))\right) \\
&\geq\; \phi^{-1}\left(\int_{\Delta(S)} \phi\left(\sum_{s\in S} r(\alpha(\nu^*), s)\pi(s)\right) \nu(\mathrm{d}\pi))\right).
\end{aligned}$$

From Remark 1 and by the observable payoffs property, we have that for every model π in $\operatorname{supp}\nu^*$,

$$\sum_{s\in S} r(\alpha(\nu^*),s)\pi(s) = \sum_{s\in S} r(\alpha(\nu^*),s)\overline{\pi}(s). \tag{21}$$

But since $\operatorname{supp}\nu \subseteq \operatorname{supp}\nu^*$, we have that (21) holds also for every model π in $\operatorname{supp}\nu^*$. Then,

$$\begin{aligned}
V_1(\nu) &= \phi^{-1}\left(\int_{\Delta(S)} \phi\left(\sum_{s\in S} r(\alpha(\nu),s)\pi(s)\right)\nu(\mathrm{d}\pi)\right)\\
&\geq \phi^{-1}\left(\int_{\Delta(S)} \phi\left(\sum_{s\in S} r(\alpha(\nu^*),s)\pi(s)\right)\nu(\mathrm{d}\pi)\right)\\
&= \phi^{-1}\left(\int_{\Delta(S)} \phi\left(\sum_{s\in S} r(\alpha(\nu^*),s)\overline{\pi}(s)\right)\nu(\mathrm{d}\pi)\right)\\
&= \sum_{s\in S} r(\alpha(\nu^*),s)\overline{\pi}(s)\\
&= \phi^{-1}\left(\int_{\Delta(S)} \phi\left(\sum_{s\in S} r(\alpha(\nu^*),s)\pi(s)\right)\nu^*(\mathrm{d}\pi)\right)\\
&= V_1(\nu^*).
\end{aligned}$$

■

References

[1] Arrow, K.J, and J.R. Green (1973): "Notes on expectations equilibria in Bayesian settings," *Institute for Mathematical Studies in the Social Sciences,* Working Paper 33.

[2] Battigalli, P., S. Cerreia-Vioglio, F. Maccheroni and M. Marinacci (2015): "Self-confirming equilibrium and model uncertainty," *American Economic Review*, 105, 646-677.

[3] Battigalli, P., S. Cerreia-Vioglio, F. Maccheroni and M. Marinacci (2016a): "Analysis of information feedback and selfconfirming equilibrium," *Journal of Mathematical Economics*, 66, 40-51.

[4] Battigalli, P., S. Cerreia-Vioglio, F. Maccheroni, M. Marinacci, and T. Sargent (2016b): "A framework for the analysis of self-confirming policies," *IGIER W.P. Series, 573.*

[5] Blanchard, O. (1985): "Debt, deficit, and finite horizon," *Journal of Political Economy*, 93, 223-247.

[6] Blackwell, D., and L. Dubins (1962): "Merging of opinions with increasing information," *Annals of Mathematical Statistics,* 882-886.

[7] Cerreia-Vioglio S., F. Maccheroni, M. Marinacci, and L. Montrucchio (2013): "Classical subjective expected utility," *Proceedings of the National Academy of Sciences*, 110, 6754-6759.

[8] Doob, J.L. (1949): "Application of the theory of martingales," *Colloq. Internat. du CNRS*, 23-27.

[9] Easley, D., and N. M Kiefer (1988): "Controlling a stochastic process with unknown parameters," *Econometrica*, 5, 1045-1064.

[10] Epstein, L.G., and M. Schneider (2007): "Learning under ambiguity," *The Review of Economic Studies*, 74, 1275-1303.

[11] Esponda, I., and D. Pouzo (2016): "Berk–Nash equilibrium: a framework for modeling agents with misspecified models," *Econometrica*, 84, 1093-1130.

[12] Fudenberg, D., and D.M. Kreps (1995): "Learning in extensive-form games I. Self-confirming equilibria," *Games and Economic Behavior,* 8, 20-55.

[13] Fudenberg, D., and D.K. Levine (1993): "Steady state learning and Nash equilibrium," *Econometrica*, 61, 547-573.

[14] Fudenberg, D., and D.K. Levine (1998): *The Theory of Learning in Games*, MIT Press, Cambridge MA.

[15] Gilboa, I., and D. Schmeidler (1989): "Maxmin expected utility with a non-unique prior," *Journal of Mathematical Economics*, 18, 141-153.

[16] Gittins, J.C. (1989): *Multi-armed Bandit Allocation Indices*, Wiley-Interscience Series in Systems and Optimization, Chichester: John Wiley & Sons, Ltd.

[17] Gray, R.M. (1989): *Probability, Random Processes, and Ergodic Properties*, 2$^{\text{nd}}$ ed., Springer, New York.

[18] Hanany, E., and P. Klibanoff (2009): "Updating ambiguity averse preferences," *The B.E. Journal of Theoretical Economics,* 9.

[19] Kabanov, Y.M., R. S. Liptser and A. N. Shiryaev (1977): “On the question of absolute continuity and singularity of probability measures,” *Math. USSR-Sb.,* 33, 203-221.

[20] Kalai, E., and E. Lehrer (1993): “Subjective equilibrium in repeated games,” *Econometrica,* 61.5, 1231-1240.

[21] Kalai, E., and E. Lehrer (1994): “Weak and strong merging of opinions,” *Journal of Mathematical Economics,* 23.1, 73-86.

[22] Kalai, E., and E. Lehrer (1995): “Subjective games and equilibria,” *Games and Economic Behavior,* 8, 123-163.

[23] Klibanoff, P., M. Marinacci and S. Mukerji (2005): “A smooth model of decision making under ambiguity,” *Econometrica*, 73, 1849-1892.

[24] Kreps, D.M. (2013): *Microeconomic Foundations I*, Princeton University Press, Princeton.

[25] Kuhn, H.W. (1953): extensive games and the problem of information, in *Contributions to the Theory of Games II*, ed. by H.W. Kuhn and A.W. Tucker. Princeton: Princeton University Press, 193-216.

[26] Marinacci, M. (2015): “Model uncertainty,” *Journal of the European Economic Association*, 13, 998-1076.

[27] Marschak, J., and R. Radner (1972): *Economic Theory of Teams*, Yale University Press, New Haven.

[28] Sargent, T.J. (1999): *The Conquest of American Inflation*, Princeton University Press, Princeton.

[29] Selten, R. (1975): “Re-examination of the perfectness concept for equilibrium points in extensive games,” *International Journal of Game Theory*, 4, 25-55.

[30] Siniscalchi, M. (2011): “Dynamic choice under ambiguity,” *Theoretical Economics*, 6, 379-421.

[31] Stroock, D.W. (1993): *Probability Theory: An Analytic View*, Cambridge University Press, Cambridge.

[32] Witsenhausen, H.S. (1971): “Separation of estimation and control for discrete time systems,” *Proceedings of the IEEE*, 59, 1557-1566.